Tracy-Widom Distribution for Heterogeneous Gram Matrices With Applications in Signal Detection

Xiucai Ding and Fan Yang

***Abstract*— Detection of the number of signals corrupted by high-dimensional noise is a fundamental problem in signal processing and statistics. This paper focuses on a general setting where the high-dimensional noise has an unknown complicated heterogeneous variance structure. We propose a sequential test which utilizes the edge singular values (i.e., the largest few singular values) of the data matrix. It also naturally leads to a consistent sequential testing estimate of the number of signals. We describe the asymptotic distribution of the test statistic in terms of the Tracy-Widom distribution. The test is shown to be accurate and have full power against the alternative, both theoretically and numerically. The theoretical analysis relies on establishing the Tracy-Widom law for a large class of Gram type random matrices with non-zero means and completely arbitrary variance profiles, which can be of independent interest.**

***Index Terms*— Signal-plus-noise model, Tracy-Widom law, random Gram matrices, Dyson Brownian motion.**

I. Introduction

DETECTION of unknown noisy signals is a fundamental task in many signal processing and wireless communication applications [4], [47], [61], [65]. Consider the following generic signal-plus-noise model

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{z}, \tag{I.1}$$

where $\boldsymbol{s}$ and $\boldsymbol{z}$ are independent p-dimensional centered signal and noise vectors, respectively. In many applications, $\boldsymbol{s}$ is usually generated from a low-dimensional MIMO filter such that $\boldsymbol{s} = \Gamma\boldsymbol{\nu}$ [47], where Γ is a $p \times r$ deterministic matrix, $\boldsymbol{\nu}$ is an r-dimensional centered random vector and r is some unknown fixed integer that does not depend on p. The value of r is one of the most important inputs for many computationally demanding parametric procedures such as direction of arrival estimation, blind source deconvolution, and so on. In the literature of statistical signal processing, the most common approaches to determine the value of r are perhaps the information theoretic criteria, including the minimum description length (MDL), Bayesian information criterion (BIC) and Akaike information criterion (AIC) and their variants. For a detailed review of this aspect, we refer the reader to [67]. All these methods assume that the dimension p is fixed and the sample size n, i.e., the number of observations, goes to infinity. Consequently, none of these estimators is applicable to large arrays where the number of sensors is comparable to or even larger than the sample size [49].

To address the issue of high dimensionality, many methods and statistics have been proposed to infer the value of r under various settings. Many methods have been proposed to test $\mathbf{H}_0 : r = 0$ against $\mathbf{H}_a : r \geq 1$, which is equivalent to testing the existence of the signals. When $\boldsymbol{z}$ is a white noise, a non-parametric method was proposed in [49], the generalized maximum likelihood test was studied [9] and a sample eigenvalue based method was proposed in [61]. When $\boldsymbol{z}$ is a colored noise, i.e., $\boldsymbol{z} = \Sigma^{1/2}\boldsymbol{x}$ for a positive definite covariance matrix Σ and a white noise $\boldsymbol{x}$, the same testing problem has been considered in [7], [15], [62], [72] under different moment assumptions on the entries of $\boldsymbol{x}$. However, all the aforementioned methods assume explicitly that the noise vectors $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n$ are generated independently from the same distribution. If the noise vectors are correlated or generated from possibly different distributions, none of these methods works or has been justified rigorously. One such example is the doubly heteroscedastic noise, whose matrix of noise vectors $(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n)$ take the form $A^{1/2}\mathcal{N}B^{1/2}$ [55], where $\mathcal{N}$ is a $p \times n$ white noise matrix, and A and B are two positive definite symmetric matrices representing the spatial and temporal covariances, respectively. Many previous works also depend crucially on the null hypothesis $r = 0$, and cannot be applied to the more general setting with null hypothesis $r = r_0$ for a fixed $r_0 \geq 0$.

A. Problem Setup and Test Statistics

In this paper, we present a more general setting for the statistical analysis of the detection of the number of signals. On the one hand, we propose some statistics to study the following hypothesis testing problem

$$\mathbf{H}_0 : r = r_0 \quad \text{vs} \quad \mathbf{H}_a : r > r_0, \tag{I.2}$$

where r_0 is some pre-given integer representing our belief of the true value of r. (I.2) generalizes the previous works, which mainly focus on the $r_0 = 0$ case, i.e., the testing

Manuscript received 12 April 2021; revised 13 February 2022; accepted 12 May 2022. Date of publication 20 May 2022; date of current version 15 September 2022. The work of Xiucai Ding was supported in part by the NSF under Grant DMS-2113489. The work of Fan Yang was supported in part by the Wharton Dean's Fund for Post-Doctoral Research. *(Corresponding author: Xiucai Ding.)*

Xiucai Ding is with the Department of Statistics, University of California at Davis, Davis, CA 95616 USA (e-mail: xcading@ucdavis.edu).

Fan Yang is with the Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: fyang75@wharton.upenn.edu).

Communicated by A. Maleki, Associate Editor for Signal Processing.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3176784.

Digital Object Identifier 10.1109/TIT.2022.3176784

of the existence of signals. On the other hand, we consider more general covariance structures of the noise, which include the doubly-heteroscedastic noise, sparse noise and noise with banded structures as special cases. We refer the readers to Examples 1 and 2 and the simulation settings in Section IV for more details. We emphasize that through (I.2), a natural consistent sequential testing estimate of r can be generated, that is,

$$\widehat{r} := \inf\{r_0 \geq 0 : \ \mathbf{H}_0 \text{ is accepted}\}. \tag{I.3}$$

We refer the readers to (III.10) and Corollary 3 for more rigorous arguments on this aspect.

In order to test (I.2), we propose some data-adaptive statistics utilizing the edge eigenvalues of the data matrix. Suppose we observe n data samples and stack them into the data matrix

$$Y = R + Z, \tag{I.4}$$

where $Y = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n) \in \mathbb{R}^{p\times n}$ collects the noisy observations, $R = (\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n)$ is the signal matrix of rank r, and $Z = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n)$ is the noise matrix. The matrix (I.4) is commonly referred to as the *signal-plus-noise matrix* in the literature, which is also closely related to the problem of low-rank matrix denoising [6], [17], [21], [60], [70], [73]. In the current paper, we consider the high-dimensional regime where p and n are comparably large so that

$$\tau \leq p/n \leq \tau^{-1},$$

for a small constant $0 < \tau < 1$. We assume that the entries of Y are independent random variables satisfying that

$$\mathbb{E}y_{ij} = r_{ij}, \quad \text{var}(y_{ij}) = s_{ij}. \tag{I.5}$$

Correspondingly, we will also call $R = (r_{ij})$ the *mean matrix*, while the *variance matrix* $S = (s_{ij})$ describes a heterogeneous variance profile for the noise. In this paper, we refer to $YY^\top$ as a *random Gram matrix*. We mention that the detection of the number of signals has been studied rigorously in the literature only when S is of sample covariance type, that is, $s_{ij} = a_i$ for some $a_i > 0$. Even for the doubly-heteroscedastic noise with $s_{ij} = a_i b_j$ for some $a_i, b_j > 0$, the aforementioned testing methods in the literature will lose their validity.

There exists a vast literature on conducting high-dimensional statistical inference using the largest eigenvalues of $YY^\top$ when S is of sample covariance type. For instance, they have been employed to test the existence and number of spikes for the spiked covariance matrix model [46], [65], test the number of factors in factor model [64], detect the signals in a signal-plus-noise model [4], [7], [9], [72], test the structure of covariance matrices [24], [40], and perform the multivariate analysis of variance (MANOVA) [37], [40]. In most of these applications, on the one hand, researchers aim to test between the null hypothesis of a non-spiked sample covariance matrix and the alternative of a spiked sample covariance matrix. Under the null hypothesis, the largest few eigenvalues have been proved to satisfy the Tracy-Widom law asymptotically under a proper scaling [7], [18], [24], [45], [48], [53], [63], [66]. More precisely, there exist parameters λ_+ and ϖ such that $\varpi p^{2/3}(\lambda_1 - \lambda_+)$ converges in law to the type-1 Tracy-Widom distribution [68], [69], where λ_1 is the largest eigenvalue of $YY^\top$. Then it is natural to choose $\varpi p^{2/3}(\lambda_1 - \lambda_+)$ as the test statistic. On the other hand, especially in the setting of factor models in economics, researchers are interested in inferring the number of factors. Under the null hypothesis that there are r large factors, the $(r+1)$-th eigenvalue λ_{r+1} obeys the Tracy-Widom distribution asymptotically [64].

Based on the above observations, if we can show that λ_{r+1} obeys the Tracy-Widom law in our setting (I.5), we can naturally choose $\varpi p^{2/3}(\lambda_{r+1} - \lambda_+)$ as the test statistic for the testing problem (I.2). However, in practice, the two parameters ϖ and λ_+ depend on the usually unknown variance matrix S. To resolve this issue, we can follow [64] to use the statistic

$$\mathbb{T} \equiv \mathbb{T}(r_0) := \max_{r_0 < i \leq r_*} \frac{\lambda_i - \lambda_{i+1}}{\lambda_{i+1} - \lambda_{i+2}}, \tag{I.6}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues of $YY^\top$ arranged in descending order, and r_* is a pre-chosen integer that is interpreted as the maximum possible number of signals the model can have. We will also see in Section III-B that (I.6) can be used to count the number of outlier eigenvalues that correspond to signals through a sequential testing procedure. Onatski [64] observed that in the setting of sample covariance matrices, $\mathbb{T}$ is independent of ϖ and λ_+ under the null hypothesis, and hence is asymptotically pivotal. Moreover, its asymptotic distribution is determined by the Tracy-Widom law of the edge eigenvalues. Consequently, we can approximate the distribution of $\mathbb{T}$ using Monte Carlo simulations of Wishart matrices.

We point out that in many literature and scientific applications [6], [44], [59], [60], [72], it is reasonable to assume that the signals are distinct. Under this assumption, we also propose the following statistic

$$\mathbb{T}_{r_0} := \frac{\lambda_{r_0+1} - \lambda_{r_0+2}}{\lambda_{r_*+1} - \lambda_{r_*+2}}. \tag{I.7}$$

Compared to (I.6), the statistic (I.7) relies on fewer (actually, only three or four) sample eigenvalues. Moreover, for commonly used alternatives with low-rank signals, we expect that the statistic (I.7) has better performance in terms of power (i.e., it is sensitive to a wider class of alternatives and has higher power for some fixed alternative). Our expectation, although without full theoretical justification, is partly due to the fact that $\mathbb{T}_{r_0}$ has smaller critical values compared to $\mathbb{T}$ as illustrated in Table I, which is reasonable because taking maximum over a sequence of random variables increases critical values. Empirically, our simulations in Section IV will show that (I.7) indeed has better finite-sample performance than (I.6) in terms of power. In fact, we believe that the statistic (I.7) will also work when the signals are degenerate, because the corresponding sample eigenvalues will be separated. We refer the reader to Remark 4 for more details.

The statistics (I.6) and (I.7) are applicable to statistical inference only if the Tracy-Widom law has been established for the associated random Gram matrix $YY^\top$. However, to the best of our knowledge, this has only been proved rigorously for sample covariance type random Gram matrices in the literature. Therefore, for hypothesis testing problems involving

random Gram matrices with general mean and variance profiles, we need to prove the Tracy-Widom fluctuation rigorously before validating the use of $\mathbb{T}$ and $\mathbb{T}_{r_0}$. This motivates us to study the limiting distributions of the edge eigenvalues in the general setup (I.5). Here the notion "edge eigenvalues" refers to the largest few eigenvalues near the right edge of the bulk eigenvalue spectrum, excluding the outliers of $YY^\top$ caused by the signals.

B. Tracy-Widom Distribution for Random Gram Matrices

The Tracy-Widom law for the edge eigenvalues of non-spiked sample covariance matrices has been proved in a series of papers. For Wishart matrices, it was first proved in [45] that the largest eigenvalue satisfies the Tracy-Widom law asymptotically. This result was later extended to more general sample covariance matrices with generally distributed entries (assuming only certain moment assumptions) and variance profiles $s_{ij} = a_i$ (assuming certain regularity conditions on the sequence $\{a_i : 1 \le i \le p\}$) in a series of papers under various settings; see e.g. [7], [18], [24], [48], [53], [63], [66]. However, when the mean and variance profiles of the random Gram matrix become more complicated, much less is known about the limiting distribution of the edge eigenvalues.

In this paper, motivated by the applications in signal detection as discussed in Section I-A, we establish the Tracy-Widom asymptotics for the edge eigenvalues of a general class of random Gram matrices. The informal statement is given in Theorem 1. Following the conventions in the random matrix theory literature, we shall rescale the matrix Y properly so that the limiting ESD of $YY^\top$ is compactly supported as $n \to \infty$. Moreover, recall that GOE (Gaussian orthogonal ensemble) refers to symmetric random matrices of the form $H := (X + X^\top)/\sqrt{2}$, where X is a $p \times p$ matrix with i.i.d. real Gaussian entries of mean zero and variance p^{-1}. In this paper, we will consistently denote the eigenvalues of H by

$$\mu_1^{\mathrm{GOE}} \ge \mu_2^{\mathrm{GOE}} \ge \cdots \ge \mu_p^{\mathrm{GOE}}. \tag{I.8}$$

Theorem 1 (Informal Statement of Theorem 3): For Y satisfying (I.5), we denote the eigenvalues of $\mathcal{Q} := YY^\top$ by $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$. Let λ_+ be the rightmost edge of the limiting bulk eigenvalue spectrum, and $a \in \mathbb{N}$ be the index of the largest edge eigenvalue. Then, there exists a deterministic sequence of numbers $\varpi \equiv \varpi(R, S)$ depending on R and S, such that for any fixed $k \in \mathbb{N}$, the first k rescaled edge eigenvalues, $\{\varpi p^{2/3}(\lambda_{a+i} - \lambda_+) : 0 \le i \le k-1\}$, have the same asymptotic joint distribution as the first k rescaled eigenvalues of GOE, $\{p^{2/3}(\mu_i^{\mathrm{GOE}} - 2) : 1 \le i \le k\}$, as $p \to \infty$.

It is well-known that $p^{2/3}(\mu_1^{\mathrm{GOE}} - 2)$ converges to the type-1 Tracy-Widom distribution [68], [69]. Furthermore, for any fixed $k \in \mathbb{N}$, the joint distribution of the largest k eigenvalues of GOE can be written in terms of the Airy kernel [38]. Hence Theorem 1 gives a complete description of the finite-dimensional correlation functions of the edge eigenvalues of $\mathcal{Q}$. Once Theorem 1 is established, we can determine the asymptotic distributions of the statistics (I.6) and (I.7), and apply them to the hypothesis testing problem (I.2).

Our proof of Theorem 1 is based on the following result on the edge eigenvalues of a general class of Gaussian divisible random Gram matrices.

Theorem 2 (Informal Statement of Theorem 4): For a parameter $t > 0$, we denote $\mathcal{Q}_t := (Y + \sqrt{t}X)(Y + \sqrt{t}X)^\top$, where X is a $p \times n$ random matrix independent of Y and has i.i.d. Gaussian entries of mean zero and variance n^{-1}. Denote the eigenvalues of $\mathcal{Q}_t$ by $\lambda_1(t) \ge \lambda_2(t) \ge \cdots \ge \lambda_p(t)$. Let $\eta_* > 0$ be a scale parameter depending on n. Suppose the empirical spectral distribution of $\mathcal{Q} = YY^\top$ has a regular square root behavior near the right edge λ_+ on any scale larger than η_* (in the sense of Definition 1 below). Let $a \in \mathbb{N}$ be the index of the largest edge eigenvalue. Then for any $t \gg \sqrt{\eta_*}$ and fixed $k \in \mathbb{N}$, there exist deterministic sequences of numbers ϖ_t and $\lambda_{+,t}$ such that the first k rescaled edge eigenvalues of $\mathcal{Q}_t$, $\{\varpi_t p^{2/3}(\lambda_{a+i}(t) - \lambda_{+,t}) : 0 \le i \le k-1\}$, have the same asymptotic joint distribution as the first k rescaled eigenvalues of GOE, $\{p^{2/3}(\mu_i^{\mathrm{GOE}} - 2) : 1 \le i \le k\}$, as $p \to \infty$.

On one hand, Theorem 2 covers more general matrices than the random Gram matrices proposed in (I.5), because it only requires a regular square root behavior of the ESD near the right edge without assuming any independence between matrix entries of Y. We remark that the square root behavior of the ESD is generally believed to be a necessary condition for the appearance of the Tracy-Widom law in the asymptotic limit. For example, if the ESD has a cubic root behavior, then the corresponding cusp universality is different from the Tracy-Widom law [16], [29]. On the other hand, Theorem 2 gives the Tracy-Widom law for the edge eigenvalues of a different matrix $\mathcal{Q}_t$ other than $\mathcal{Q}$. To obtain the Tracy-Widom law for the original matrix $\mathcal{Q}$, we still need to show that the edge eigenvalues of $\mathcal{Q}_t$ have the same joint distribution as those of $\mathcal{Q}$ asymptotically, which, however, is not always true. In fact, if t is too large, then the edge statistics of $\mathcal{Q}_t$ can be very different from those of $\mathcal{Q}$. For example, if Y is a rectangular matrix whose singular values are all the same, then $\mathcal{Q}$ trivially has a square root behavior on any scale larger than $\eta_* = 1$ in the sense of Definition 1. But in the setting of Theorem 2, for $t \gg 1$, the edge statistics of $\mathcal{Q}_t$ is dominated by a Wishart matrix $tXX^\top$.

From the above discussions, we see that in order to prove the Tracy-Widom law for the edge eigenvalues of $\mathcal{Q}$ using Theorem 2, we need to establish the following two results:

- the ESD of $\mathcal{Q}$ has a regular square root behavior near λ_+ on a sufficiently fine scale $\eta_* \ll 1$;
- for some $\sqrt{\eta_*} \ll t \ll 1$, the edge statistics of $\mathcal{Q}_t$ match those of $\mathcal{Q}$ asymptotically.

In random matrix theory, there is a general way to accomplish this by using some sharp estimates, called *local laws*, on the resolvent of $\mathcal{Q}$, defined as $G(z) := (\mathcal{Q} - z)^{-1}$ for $z \in \mathbb{C}$. Such local laws for the model (I.5) have been proved in [2], [3] under quite general conditions. Combining these local laws with Theorem 2, we can conclude Theorem 1 using some standard resolvent comparison arguments developed in e.g. [35], [48], [54], [74].

We remark that there exists another method in the literature [37], [52], [53], [75] to prove the Tracy-Widom law for sample

covariance type matrices, that is, a so-called *resolvent flow argument*. While we expect this method to be also applicable to our setting, the techniques seem to be much harder, and we do not pursue this direction in this paper.

The rest of this paper is organized as follows. In Section II, we give the precise assumptions on the signal matrix R and the variance matrix S. We also provide some concrete examples with complicated heterogeneous variance profiles S, which have not been studied rigorously in the literature. In Section III, we state our main results. The Tracy-Widom distribution for general random Gram matrices is presented in III-A, while the theoretical properties of the testing statistics (I.6) and (I.7) are analyzed in Section III-B. In Section IV, we conduct numerical simulations to verify the accuracy and power of the proposed statistics for the testing problem (I.2) under various noise settings that have not been considered in the literature. In Section V, we sketch the strategy for proving the Tracy-Widom distribution. The technical proofs are put into Appendices A–C.

II. The Model Assumptions and Examples

In this section, we impose some general assumptions on the signal matrix R and the variance matrix S. We also provide some important examples that have been used in the literature. Note that $YY^\top$ and $Y^\top Y$ have the same non-zero eigenvalues. Hence without loss of generality, we only need to consider the high-dimensional setting where the aspect ratio $c_n := p/n$ satisfies that

$$\tau \le c_n \le 1, \tag{II.1}$$

for a small constant $\tau > 0$. For the signal matrix R, we assume that

$$\operatorname{rank}(R) = r, \tag{II.2}$$

for a fixed $r \in \mathbb{N}$ that is independent of p and n. Note that when $r = 0$, Y is a centered random Gram matrix. Following [2], [3], we impose the following regularity assumptions on the heterogeneous variance profile.

Assumption 1: Suppose S satisfies the following regularity conditions.

(A1) The dimensions of S are comparable, that is, (II.1) holds.

(A2) The variances are bounded in the sense that there exist constants $s_*, \varepsilon_* > 0$ such that

$$\max_{i,j} s_{ij} \le \frac{s_*}{n}, \quad \min_{i,j} s_{ij} \ge \frac{n^{-1/3+\varepsilon_*}}{n}. \tag{II.3}$$

(A3) The matrices S and $S^\top$ are irreducible in the sense that there exist $L_1, L_2 \in \mathbb{N}$ and a small constant $\tau > 0$ such that

$$\min_{i,j}[(SS^\top)^{L_1}]_{ij} \ge \frac{\tau}{n}, \quad \min_{i,j}[(S^\top S)^{L_2}]_{ij} \ge \frac{\tau}{n}.$$

(A4) The rows and columns of S are sufficiently close to each other in the following sense. There is a continuous monotonically decreasing (n-independent) function $\Gamma : (0,1] \to (0,\infty)$ such that $\lim_{\epsilon \downarrow 0} \Gamma(\epsilon) = \infty$, and for all $\epsilon \in (0,1]$, we have

$$\Gamma(\epsilon) \le \min\Big\{ \inf_{1\le i\le p} \frac{1}{p}\sum_l \frac{1}{\epsilon + n\|S_i - S_l\|_2^2}, \inf_{1\le j\le n} \frac{1}{n}\sum_l \frac{1}{\epsilon + n\|(S^\top)_j - (S^\top)_l\|_2^2}\Big\}, \tag{II.4}$$

where S_i and $(S^\top)_j$ denote the i-th row of S and j-th row of $S^\top$, respectively.

Remark 1: The upper bound in (II.3) is chosen such that the limiting ESD of $(Y-R)(Y-R)^\top$ is compactly supported as $n \to \infty$. More precisely, we equip $\{1, \cdots, p\}$ and $\{1, \cdots, n\}$ with the ℓ^∞-norm and denote the induced operator norms by $\|\cdot\|_{\ell^\infty(n)\to\ell^\infty(p)}$ and $\|\cdot\|_{\ell^\infty(p)\to\ell^\infty(n)}$. Then, Proposition 2.1 of [2] shows that the rightmost edge λ_+ of the limiting ESD of $(Y-R)(Y-R)^\top$ satisfies

$$\lambda_+ \le 4\mathfrak{M}, \tag{II.5}$$

with

$$\mathfrak{M} := \max\left\{\|S\|_{\ell^\infty(n)\to\ell^\infty(p)}, \|S^\top\|_{\ell^\infty(p)\to\ell^\infty(n)}\right\}.$$

By (II.3), it is easy to see that $\mathfrak{M} \le s_*$.

Remark 2: As explained in (2.22) of [1], assumption (A4) aims to rule out possible spikes from S. In [2, Remark 2.4], an easier to check sufficient (but not necessary) condition for **(A4)** was also proposed.

(A4-s): There are two finite partitions $(I_\alpha)_{\alpha\in\mathcal{A}}$ and $(J_\beta)_{\beta\in\mathcal{B}}$ of $\{1, \cdots, p\}$ and $\{1, \cdots, n\}$, respectively, such that for any $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$, we have $|I_\alpha| \ge \tau p$, $|I_\beta| \ge \tau n$, and

$$\|S_{i_1} - S_{i_2}\|_2 \le \frac{|i_1 - i_2|^{1/2}}{\tau n} \text{ for } i_1, i_2 \in I_\alpha,$$
$$\|(S^\top)_{j_1} - (S^\top)_{j_2}\|_2 \le \frac{|j_1 - j_2|^{1/2}}{\tau n} \text{ for } j_1, j_2 \in J_\beta, \tag{II.6}$$

for a small constant $\tau > 0$. The condition (A4) follows easily from (A4-s) using an integral approximation.

In addition to (II.2), we introduce the following assumption on the signal strengths, i.e. the singular values of R.

Assumption 2: We assume that (II.2) holds. When $r \ge 1$, denote by $\sigma_r(R)$ the smallest non-trivial singular value of R. We assume that

$$\sigma_r(R) > (4+\tau)\sqrt{\mathfrak{M}}, \tag{II.7}$$

for a small constant $\tau > 0$, where $\mathfrak{M}$ is defined in (II.5).

Remark 3: (II.7) is commonly referred to as the supercritical condition, and has appeared in lots of literature in random matrix theory and statistics [6], [9], [60], [62]. It is a sufficient condition for the mean matrix R to give rise to r outliers of $YY^\top$ that are detached from the bulk spectrum. By Lemma 4 below, we have that the largest eigenvalue of $(Y-R)(Y-R)^\top$ is at most $\lambda_+ + \mathrm{o}(1)$ with high probability. Combining it with (II.7) and applying Weyl's inequality, it is easy to check that $YY^\top$ has r eigenvalues that are larger than $(2+\tau-\mathrm{o}(1))^2\mathfrak{M}$. On the other hand, by the Cauchy interlacing, the limiting bulk eigenvalue spectrum of $YY^\top$ is supported on $[0, \lambda_+]$. Hence, under (II.7), there are r outliers that are away from the spectrum edge λ_+.

However, we remark that $4\sqrt{\mathfrak{M}}$ is quite likely not the exact threshold for BBP transition [5]. To guarantee the Tracy-Widom law of the edge eigenvalues, it is necessary that all spikes of R are away from (i.e., either above or below) the BBP threshold. If there are critical spikes (i.e., spikes that are exactly equal to the BBP transition threshold), then the Tracy-Widom law of the edge eigenvalues can fail; see Theorem 1.1 in [5]. Here we have chosen (II.7) simply to ensure that all spikes are supercritical. To determine the exact BBP threshold and to include settings with subcritical spikes, we need to perform a more detailed study of spiked random Gram matrices. We postpone it to future works, since it is not the focus of the current paper.

In what follows, we give two concrete examples which satisfy the above assumptions and have not been studied rigorously in the literature.

Example 1 (Doubly-Heteroscedastic Noise, [55]): Consider the following doubly-heteroscedastic noise matrix

$$Y := A^{1/2}\mathcal{N}B^{1/2}, \tag{II.8}$$

where A and B are deterministic positive definite symmetric matrices, and $\mathcal{N} = (\mathcal{N}_{ij})$ is a $p \times n$ random matrix with i.i.d. entries of mean zero and variance n^{-1}. Suppose A and B are diagonal matrices

$$A = \mathrm{diag}(a_1, \ldots, a_p), \quad B = \mathrm{diag}(b_1, \ldots, b_n), \tag{II.9}$$

with $a_1 \geq a_2 \geq \cdots \geq a_p > 0$ and $b_1 \geq b_2 \geq \cdots \geq b_n > 0$. Then $\mathcal{Q} = YY^\top$ is a random Gram matrix as in Theorem 1 with variance matrix $S = ((a_i b_j)/n)$ and mean matrix $R = 0$. It is easy to see that (A2) and (A3) of Assumption 1 hold if a_i's and b_j's are all of order 1. Furthermore, assumption (II.4) is reduced to

$$\Gamma(\epsilon) \leq \min\Big\{ \inf_{1\leq i\leq p} \frac{1}{p}\sum_l \frac{1}{\epsilon + |a_i - a_l|^2}, \inf_{1\leq j\leq n} \frac{1}{n}\sum_l \frac{1}{\epsilon + |b_j - b_l|^2}\Big\}, \tag{II.10}$$

and condition (II.6) is reduced to

$$\begin{aligned} |a_{i_1} - a_{i_2}| &\leq \tau^{-1}\frac{|i_1 - i_2|^{1/2}}{n^{-1/2}} \quad \text{for } i_1, i_2 \in I_\alpha, \\ |b_{j_1} - b_{j_2}| &\leq \tau^{-1}\frac{|j_1 - j_2|^{1/2}}{n^{1/2}} \quad \text{for } j_1, j_2 \in J_\beta. \end{aligned} \tag{II.11}$$

In fact, if we have $a_i = f(i/p)$ and $b_j = g(j/n)$ for some piecewise $1/2$-Hölder continuous functions f and g, then (II.11) holds true. One special case is that f and g are piecewise constant functions, which happens when the eigenvalues of A and B take at most $\mathrm{O}(1)$ many different values. If (II.10) or (II.11) holds, as we will see in Section III-A, Theorem 1 applies to (II.8) with $r = 0$.

We remark that the diagonal assumption (II.9) is not necessary for the Tracy-Widom asymptotics. When the matrices A and B are non-diagonal, we get a model that extends the setting in (I.5) because the entries of $Y = A^{1/2}\mathcal{N}B^{1/2}$ can be correlated. Finally, we remark that (A4) of Assumption 1 can be violated by allowing for some large a_i's and b_j's. Then we get a spiked separable covariance matrix, which has been studied in detail in [20]. Our Theorem 1 also applies to this case.

Example 2 (Sparse Noise, [43], [57]): In this example, we consider the sparse noise matrix Z as proposed in [43]. The sparse random Gram matrices can be used as a natural model to study high-dimensional data with randomly missing observations. For instance, given a probability p, we set $z_{ij} = h_{ij}w_{ij}$, where w_{ij} are random variables independent of $\{h_{ij}\}$, and h_{ij} are i.i.d. (rescaled) Bernoulli random variables with $\mathbb{P}(h_{ij} = (n\mathsf{p})^{-1/2}) = \mathsf{p}$ and $\mathbb{P}(h_{ij} = 0) = 1 - \mathsf{p}$. More generally, we say that $\mathcal{Q} = YY^\top$ is a sparse random Gram matrix if Y satisfies the following properties: the entries y_{ij}, $1 \leq i \leq p, 1 \leq j \leq n$, are independent random variables satisfying

$$\mathbb{E}\left|\frac{y_{ij} - \mathbb{E}y_{ij}}{\sqrt{ns_{ij}}}\right|^k \leq \frac{(Ck)^{Ck}}{nq^{k-2}}, \quad k \geq 3, \tag{II.12}$$

for a large constant $C > 0$ and sparsity parameter q with $1 \ll q \leq \sqrt{n}$. In the above setting with randomly missing observations, we have that $q = \sqrt{n\mathsf{p}}$.

III. Main Results

In this section, we state the main results of this paper. The Tracy-Widom distribution of the edge eigenvalues for a general class of random Gram matrices, i.e., the formal statement of Theorem 1, will be presented in Section III-A. The theoretical properties of the test statistics (I.6)–(I.7) and the associated sequential estimator (I.3) will be given in Section III-B.

A. Tracy-Widom Distribution for Random Gram Matrices

In this subsection, we provide the formal statement for Theorem 1. Before stating our main result, we first introduce the necessary notations. Let $\mathbb{C}_+ := \{z \in \mathbb{C} : Imz > 0\}$ be the upper half complex plane. If (II.3) holds, then there exists a unique vector of holomorphic functions

$$\mathbf{m}(z) = (m_1(z), \cdots, m_p(z)) : \mathbb{C}_+ \to \mathbb{C}^p,$$

satisfying the so-called *vector Dyson equation*

$$\frac{1}{\mathbf{m}} = -z\mathbf{1} + S\frac{1}{\mathbf{1} + S^\top\mathbf{m}}, \tag{III.1}$$

such that $Imm_k(z) > 0$, $k = 1, \cdots, p$, for any $z \in \mathbb{C}_+$ [2], [3], [39]. In the above equation, $\mathbf{1}$ denotes the vector whose entries are all equal to 1, and both $1/\mathbf{m}$ and $1/(\mathbf{1} + S^\top\mathbf{m})$ mean the entrywise reciprocals. Moreover, for each $k = 1, \cdots, p$, there exists a unique probability measure ν_k that has support contained in $[0, 4\mathfrak{M}]$ and is absolutely continuous with respect to the Lebesgue measure, such that m_k is the Stieltjes transform of ν_k:

$$m_k(z) = \int \frac{\nu_k(\mathrm{d}x)}{x - z}, \quad z \in \mathbb{C}_+.$$

(If we consider the case $p > n$, then ν_k will also have a point mass at zero, but we do not have to worry about this issue under (II.1).) Let ρ_k be the density function associated with ν_k. Then the asymptotic ESD of $(Y - R)(Y - R)^\top$ is given

by $\nu := p^{-1}\sum_k \nu_k$, with the following density ρ and Stieltjes transform m,

$$\rho := \frac{1}{p}\sum_k \rho_k, \quad m(z) := \frac{1}{p}\sum_k m_k(z). \tag{III.2}$$

We summarize the basic properties of the density functions ρ and ρ_k, $1 \le k \le p$.

Lemma 1 (Theorem 2.3 of [2]): Under Assumption 1, for any $1 \le k \le p$, there exists a sequence of positive numbers $a_1 > a_2 > \cdots > a_{2\mathfrak{q}} \ge 0$ such that

$$\operatorname{supp}\rho_k = \operatorname{supp}\rho = \bigcup_{i=1}^{\mathfrak{q}} [a_{2i}, a_{2i-1}],$$

where $\mathfrak{q} \in \mathbb{N}$ depends only on S. Moreover, ρ has the following square root behavior near a_1:

$$\rho(a_1 - x) = \pi^{-1}\varpi\sqrt{x} + \mathrm{O}(x), \quad x \downarrow 0, \tag{III.3}$$

where $\varpi \equiv \varpi(S)$ is an order 1 positive value determined by S.

In what follows, we shall call a_k the spectral edges. In particular, we will focus on the right-most edge a_1 and denote it by $\lambda_+ \equiv a_1$ following the convention in the random matrix theory literature. We remark that as discussed in [2], it is possible that the density ρ has some cusp singularities when two edges are close to each other or when ρ touches zero. In the current paper, since we are mainly interested in the edge eigenvalue statistics around a_1, we only need assumptions to ensure (III.3). However, to show the Tracy-Widom law at other edges, we need extra edge regularity and edge separation conditions to avoid cusp singularities as in [37], [48]. We will pursue this direction in future works. Now, we are ready to state the Tracy-Widom law of the largest edge eigenvalues for a general class of random Gram matrices with variance and mean matrices satisfying Assumptions 1 and 2.

Theorem 3: Let $Y = (y_{ij})$ be a $p \times n$ random matrix such that $\widetilde{y}_{ij} := (y_{ij} - r_{ij})/\sqrt{s_{ij}}$ are real i.i.d. random variables. Suppose $\widetilde{y}_{11}$ follows a probability distribution that does not depend on n, and satisfies $\mathbb{E}\widetilde{y}_{11} = 0$, $\mathbb{E}\widetilde{y}_{11}^2 = 1$ and

$$\lim_{x\to\infty} x^4 \mathbb{P}\left(|\widetilde{y}_{11}| \ge x\right) = 0. \tag{III.4}$$

Suppose the variance matrix $S = (s_{ij})$ satisfies Assumption 1 and the mean matrix $R = (r_{ij})$ satisfies Assumption 2. Denote the eigenvalues of $\mathcal{Q} = YY^\top$ by $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$. Then we have that for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\varpi^{2/3}p^{2/3}(\lambda_{r+1} - \lambda_+) \le x\right) = F_1(x), \tag{III.5}$$

where ϖ is the value defined in (III.3), and F_1 is the type-1 Tracy-Widom cumulative distribution function. More generally, for any fixed $k \in \mathbb{N}$, we have that

$$\lim_{p\to\infty} \mathbb{P}\left[\left(\varpi^{2/3}p^{2/3}(\lambda_{i+r} - \lambda_+) \le x_i\right)_{1\le i\le k}\right] = \lim_{p\to\infty} \mathbb{P}\left[\left(p^{2/3}(\mu_i^{\mathrm{GOE}} - 2) \le x_i\right)_{1\le i\le k}\right], \tag{III.6}$$

for all $(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, where we recall that μ_i^{GOE} are the eigenvalues of GOE as given by (I.8).

Furthermore, the condition (III.4) is necessary in the following sense: if $r = 0$ and (III.4) does not hold, then we have that for any fixed $x > \lambda_+$,

$$\limsup_{n\to\infty} \mathbb{P}(\lambda_1 > x) > 0. \tag{III.7}$$

Hence $\varpi^{2/3}p^{2/3}(\lambda_1 - \lambda_+)$ does not converge to F_1 in distribution.

For the reader's convenience, we state the Tracy-Widom distributions for the models in Examples 1 and 2 as corollaries of Theorem 3.

Corollary 1: Assume that (II.1) holds. Consider the doubly-heteroscedastic matrix in (II.8), where $\mathcal{N}$ is a $p \times n$ random matrix with $\mathcal{N}_{ij} = n^{-1/2}\widetilde{y}_{ij}$ for a sequence of i.i.d. random variables $\widetilde{y}_{ij}$. Suppose $\widetilde{y}_{11}$ follows a probability distribution that does not depend on n, and satisfies $\mathbb{E}\widetilde{y}_{11} = 0$, $\mathbb{E}\widetilde{y}_{11}^2 = 1$ and (III.4). In addition, assume that

$$\mathbb{E}\left(\widetilde{y}_{11}^3\right) = 0. \tag{III.8}$$

Let A and B be $p \times p$ and $n \times n$ deterministic positive definite symmetric matrices, whose eigenvalues satisfy that

$$\tau \le a_p \le a_1 \le \tau^{-1}, \quad \tau \le b_n \le b_1 \le \tau^{-1}, \tag{III.9}$$

for a small constant $\tau > 0$, and satisfy the condition (II.10) for a continuous monotonically decreasing function $\Gamma : (0, 1] \to (0, \infty)$ such that $\lim_{\epsilon\downarrow 0}\Gamma(\epsilon) = \infty$. Then, for any fixed $k \in \mathbb{N}$, we have that

$$\lim_{n\to\infty} \mathbb{P}\left[\left(\varpi^{2/3}p^{2/3}(\lambda_i - \lambda_+) \le x_i\right)_{1\le i\le k}\right] = \lim_{n\to\infty} \mathbb{P}\left[\left(p^{2/3}(\mu_i^{\mathrm{GOE}} - 2) \le x_i\right)_{1\le i\le k}\right],$$

for all $(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, where λ_+ and ϖ are defined for the variance matrix $S = ((a_i b_j)/n)$. Finally, the condition (III.8) is not necessary if either A or B is diagonal.

Corollary 2: Suppose $\mathcal{Q} = YY^\top$ is a sparse random Gram matrix, where the entries of Y satisfy (II.12) with $q \ge n^{1/3+c_\phi}$ for a small constant $c_\phi > 0$. Suppose the variance matrix $S = (s_{ij})$ satisfies Assumption 1 and the mean matrix $R = (r_{ij})$ satisfies Assumption 2. Then for any fixed $k \in \mathbb{N}$, we have that

$$\lim_{n\to\infty} \mathbb{P}\left[\left(\varpi^{2/3}p^{2/3}(\lambda_{i+r} - \lambda_+) \le x_i\right)_{1\le i\le k}\right] = \lim_{n\to\infty} \mathbb{P}\left[\left(p^{2/3}(\mu_i^{\mathrm{GOE}} - 2) \le x_i\right)_{1\le i\le k}\right],$$

for all $(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$.

The proofs of Theorem 3, Corollary 1 and Corollary 2 will be given in Appendix A. We remark that the settings of Corollaries 1 and 2 are actually beyond the one in Theorem 3: in Corollary 1, the entries of Y can be correlated because we did not assume that A and B are diagonal, while in Corollary 2, the distribution of $\widetilde{y}_{11} = (y_{11} - r_{11})/\sqrt{s_{11}}$ may depend on n under the condition (II.12). Hence, they are not trivial corollaries of Theorem 3. But in the proof, we can reduce their settings to ones that are compatible with Theorem 3. For example, for doubly-heteroscedastic matrices, under the

setting of Corollary 1, [74] has proved the edge universality—the limiting distribution of the edge eigenvalues is the same as that in the Gaussian case with i.i.d. Gaussian $\widetilde{y}_{ij}$. On the other hand, by the rotational invariance of Gaussian $\mathcal{N}$, we can reduce the model to one with diagonal A and B so that Theorem 3 applies. Combining these two results finishes the proof of Corollary 1.

We also mention that the condition (III.8) in Corollary 1 and the condition $q \geq n^{1/3+c_\phi}$ in Corollary 2 are mainly technical. The edge universality in [74] was proved under the vanishing third moment condition. Hence, we have kept (III.8) in Corollary 1, but we believe it can be removed with further theoretical development. We also believe that $q \geq n^{1/3+c_\phi}$ can be weakened to $q \geq n^{1/6+c_\phi}$, while Corollary 2 may fail when $q \leq n^{1/6}$. Since these problems are not the main focus of this paper, we will pursue them in future works. We also refer the readers to Remark 8 for more details.

B. Theoretical Properties of the Test Statistics

With Theorem 3, we can readily obtain the asymptotic distributions of the statistics $\mathbb{T}(r_0)$ in (I.6) and $\mathbb{T}_{r_0}$ in (I.7) under the null hypothesis in (I.2), and analyze the statistical power of them under the alternatives. Corresponding to $\mathbb{T}(r_0)$ and $\mathbb{T}_{r_0}$, we define the following two sequential testing estimators

$$\begin{aligned} \widehat{r}_1 &:= \inf\{r_0 \geq 0 : \mathbb{T}(r_0) < \delta_n^{(1)}\}, \\ \widehat{r}_2 &:= \inf\{r_0 \geq 0 : \mathbb{T}_{r_0} < \delta_n^{(2)}\}. \end{aligned} \tag{III.10}$$

We will show that $\widehat{r}_1$ and $\widehat{r}_2$ are consistent estimators of r as long as we choose the critical values $\delta_n^{(1)}$ and $\delta_n^{(2)}$ properly. Let $\mathcal{W} \sim W_p(I_p, n)$ be a standard Wishart matrix. We define the following statistics $\mathbb{G}_1$ and $\mathbb{G}_2$ in terms of the eigenvalues of $\mathcal{W}$,

$$\begin{aligned} \mathbb{G}_1 &:= \max_{1\leq i\leq r_*-r_0} \frac{\lambda_i(\mathcal{W}) - \lambda_{i+1}(\mathcal{W})}{\lambda_{i+1}(\mathcal{W}) - \lambda_{i+2}(\mathcal{W})}, \\ \mathbb{G}_2 &:= \frac{\lambda_1(\mathcal{W}) - \lambda_2(\mathcal{W})}{\lambda_{r_*-r_0+1}(\mathcal{W}) - \lambda_{r_*-r_0+2}(\mathcal{W})}. \end{aligned}$$

Corollary 3: Suppose the assumptions of Theorem 3 hold and $r_* > r$. Under the null hypothesis $\mathbf{H}_0$ in (I.2), we have that

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(\mathbb{T} \leq x) &= \lim_{n\to\infty} \mathbb{P}(\mathbb{G}_1 \leq x), \\ \lim_{n\to\infty} \mathbb{P}(\mathbb{T}_{r_0} \leq x) &= \lim_{n\to\infty} \mathbb{P}(\mathbb{G}_2 \leq x), \end{aligned} \tag{III.11}$$

for all $x \in \mathbb{R}$. On the other hand, if $\delta_n^{(1)} p^{-2/3} \to 0$, then

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{T} > \delta_n^{(1)}) = 1, \quad \text{under} \quad \mathbf{H}_a; \tag{III.12}$$

if $\delta_n^{(2)} p^{-2/3} / (\lambda_{r_0+1} - \lambda_{r_0+2}) \to 0$, then

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{T}_{r_0} > \delta_n^{(2)}) = 1, \quad \text{under} \quad \mathbf{H}_a. \tag{III.13}$$

Consequently, if $\delta_n^{(1)} \to \infty$ and $\delta_n^{(1)} p^{-2/3} \to 0$, then

$$\lim_{n\to\infty} \mathbb{P}(\widehat{r}_1 = r) = 1; \tag{III.14}$$

if $\delta_n^{(2)} \to \infty$ and $\delta_n^{(2)} p^{-2/3} / (\lambda_{r_0+1} - \lambda_{r_0+2}) \to 0$, then

$$\lim_{n\to\infty} \mathbb{P}(\widehat{r}_2 = r) = 1. \tag{III.15}$$

Proof: (III.11) follows directly from (III.6). On the other hand, under $\mathbf{H}_a$ and the assumption $r_* > r$, we have that

$$\mathbb{T} \geq \frac{\lambda_r - \lambda_{r+1}}{\lambda_{r+1} - \lambda_{r+2}}.$$

By Theorem 3, we have that

$$\begin{aligned} &\lambda_{r+1} - \lambda_+ = \mathrm{O}(p^{-2/3}), \quad \lambda_{r+1} - \lambda_{r+2} = \mathrm{O}(p^{-2/3}), \\ &\lambda_{r_*+1} - \lambda_{r_*+2} = \mathrm{O}(p^{-2/3}), \end{aligned} \tag{III.16}$$

with probability $1 - \mathrm{o}(1)$. Furthermore, under Assumption 2, as discussed in Remark 3 we have that $|\lambda_r - \lambda_+| \geq c_\tau$ for a small constant $c_\tau > 0$. Hence we get that with probability $1 - \mathrm{o}(1)$,

$$\mathbb{T} \geq \frac{\lambda_r - \lambda_{r+1}}{\lambda_{r+1} - \lambda_{r+2}} \gtrsim p^{2/3},$$

which concludes (III.12) and (III.14). Finally, using (III.16), we immediately conclude (III.13) and (III.15). □

Remark 4: We make a few remarks here. First, the conditions $\delta_n^{(1)} \to \infty$ and $\delta_n^{(2)} \to \infty$ are necessary and sufficient to guarantee that $\mathbb{T}$ and $\mathbb{T}_{r_0}$ have asymptotic zero type I errors. For any fixed $r_* - r_0$, the joint distribution of $\{\lambda_i(\mathcal{W})\}_{1\leq i\leq r_*-r_0+2}$ can be expressed in terms of the Airy kernel [38]. Although it is hard to get explicit expressions of the limiting distributions of $\mathbb{G}_1$ and $\mathbb{G}_2$, it is easy to check that both the distributions are supported on the whole positive real line. Consequently, it is necessary to let $\delta_n^{(1)}$ and $\delta_n^{(2)}$ diverge. Second, in order to choose a non-trivial $\delta_n^{(2)}$ satisfying $\delta_n^{(2)} \to \infty$ and $\delta_n^{(2)} p^{-2/3} / (\lambda_{r_0+1} - \lambda_{r_0+2}) \to 0$, we need the following estimate:

$$p^{2/3} (\lambda_{r_0+1} - \lambda_{r_0+2}) \to \infty \quad \text{in probability.} \tag{III.17}$$

The condition (III.17) can be guaranteed if $\mathbf{H}_a$ holds and the (r_0+1)-th and (r_0+2)-th singular values of R are non-degenerate. However, we believe that even in the degenerate case, the condition (III.17) still holds. In fact, following [5], [11], we conjecture that the degenerate (r_0+1)-th and (r_0+2)-th spikes of R will give rise to outliers satisfying that $\lambda_{r_0+1} - \lambda_{r_0+2} \gtrsim p^{-1/2}$ with probability $1 - \mathrm{o}(1)$. To prove this fact, we need to establish the limiting distributions of the outliers of spiked random Gram matrices, and we postpone the study to a future work.

In Table I, we report some simulated finite sample critical values of $\mathbb{G}_1$ and $\mathbb{G}_2$ corresponding to type I error rate $\alpha = 0.1$ for different choices of $r_* - r_0$, $n \in \{200, 500\}$ and $c_n = p/n \in \{0.5, 1, 2\}$ based on $5,000$ Monte Carlo simulations. All the simulations in Section IV will be based on these critical values.

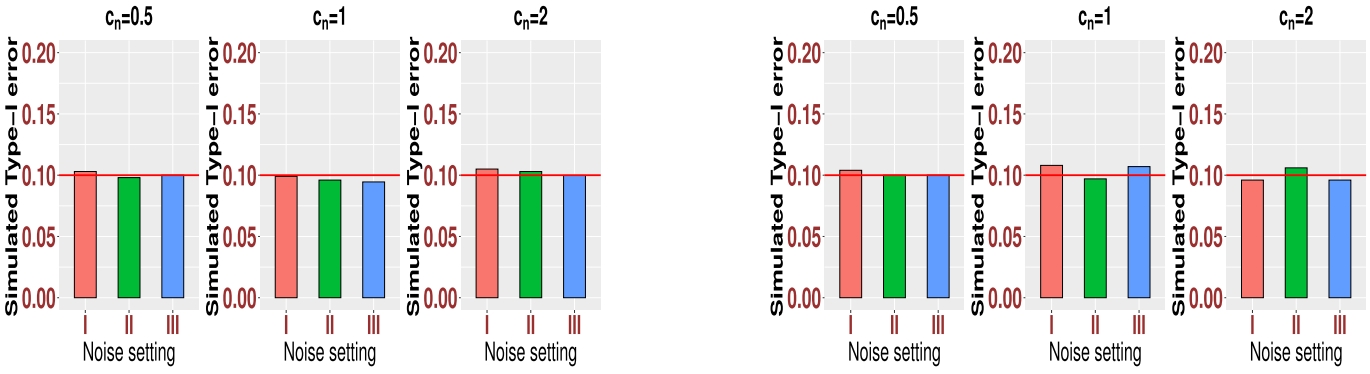
IV. Numerical Simulations

In this section, we design Monte-Carlo simulations to demonstrate the accuracy and power of our proposed statistics for the hypothesis testing problem (I.2) under some general noise structures. By Corollary 3, we will use the statistics $\mathbb{T}$ and $\mathbb{T}_{r_0}$ and reject the null hypothesis $\mathbf{H}_0$ of (I.2) if they are larger than the critical values in Table I. For the simulations, we always consider the following scenario: R is

TABLE I

CRITICAL VALUES FOR $\mathbb{G}_1$ AND $\mathbb{G}_2$ (INSIDE THE PARENTHESES) FOR DIFFERENT COMBINATIONS OF p, n AND $r_* - r_0$ UNDER THE NOMINAL SIGNIFICANCE LEVEL 0.1. WHEN $r_* - r_0 = 1$, WE HAVE $\mathbb{G}_1 = \mathbb{G}_2$, SO THEY SHARE THE SAME CRITICAL VALUES. NOTE THAT $\mathbb{G}_2$ ALWAYS HAS SMALLER CRITICAL VALUES THAN $\mathbb{G}_1$

$r_* - r_0/(p, n)$	$(100, 200)$	$(250, 500)$	$(200, 200)$	$(500, 500)$	$(400, 200)$	$(1000, 500)$
1	4.77	4.68	4.71	4.53	4.51	4.51
2	5.68 (4.98)	5.6 (4.86)	5.68 (5.02)	5.62 (4.96)	5.59 (4.95)	5.62 (4.87)
3	6.37 (5.15)	6.42 (4.95)	6.51 (5.23)	6.41 (5.48)	6.63 (5.23)	6.38 (5.19)
4	6.94 (5.41)	7.12 (5.28)	6.98 (5.63)	6.96 (5.52)	7.07 (5.34)	7.93 (5.48)
5	7.86 (5.94)	8.12 (5.87)	8.23 (6.03)	7.89 (5.94)	7.91 (5.82)	7.78 (5.79)



(a) (b)

Fig. 1. Simulated type I error rates under the nominal level 0.1 for $\mathbb{T}$ and $\mathbb{T}_{r_0}$. We take $n = 200$ and report the results based on 2,000 Monte-Carlo simulations and the critical values from Table I.

of rank $r \leq 5$, and all the singular values of R are non-degenerate. In the above scenario, we consider the following three noise structures, whose impact on the signal detection is still unknown rigorously in the literature.

(I) Z is a doubly-heteroscedastic noise matrix. Specifically, we take $Z = A^{1/2}\mathcal{N}B^{1/2}$, where $\mathcal{N}$ is a $p \times n$ white noise matrix with i.i.d. entries of mean zero and variance n^{-1}, and A and B are two positive definite matrices generated as follows: A and B have spectral decompositions $A = U_A \Sigma_A U_A^\top$ and $B = U_B \Sigma_B U_B^\top$, where

$$\Sigma_A = \mathrm{diag}(\underbrace{1, \cdots, 1}_{p/2 \text{ times}}, \underbrace{2, \cdots, 2}_{p/2 \text{ times}}),$$
$$\Sigma_B = \mathrm{diag}(\underbrace{3, \cdots, 3}_{p/4 \text{ times}}, \underbrace{4, \cdots, 4}_{p/4 \text{ times}}, \underbrace{5, \cdots, 5}_{p/2 \text{ times}}),$$

and U_A and U_B are two orthogonal matrices generated from the `R` package `pracma`.

(II) Z is a sparse noise matrix. Specifically, we take $z_{ij} = h_{ij} w_{ij}$, where h_{ij} are i.i.d. (rescaled) Bernoulli random variables satisfying $\mathbb{P}(h_{ij} = (n\mathsf{p})^{-1/2}) = \mathsf{p}$ and $\mathbb{P}(h_{ij} = 0) = 1 - \mathsf{p}$, and w_{ij} are independent $\mathcal{N}(0, s_{ij})$ random variables. In the simulations, we take $\mathsf{p} = n^{-1/4}$ and $s_{ij} = \alpha_i \beta_j$ with α_i being i.i.d. random variables uniformly distributed on $[1, 2]$ and β_j being i.i.d. random variables uniformly distributed on $[3, 4]$.

(III) $Z = (z_{ij})$ is a noise matrix whose variance matrix S has a banded latent structure. Specifically, we assume that $z_{ij} \sim \mathcal{N}(0, s_{ij})$ with

$$s_{ij} = \left(1 + \nu_{ij} \mathbf{1}_{|i-j| \leq 5}\right)/n,$$

where ν_{ij} are i.i.d. random variables uniformly distributed on $[1, 2]$.

In the simulations, we always take $r_* = 5$ and $c_n \in \{0.5, 1, 2\}$.

First, under the null hypothesis $\mathbf{H}_0$ in (I.2), we check the accuracy of the statistics under the nominal significance level 0.1. We consider the above settings (I)–(III) under the null hypothesis $r_0 = 3$, with signal matrix $R = 18\mathbf{e}_{1p}\mathbf{e}_{1n}^\top + 16\mathbf{e}_{2p}\mathbf{e}_{2n}^\top + 14\mathbf{e}_{3p}\mathbf{e}_{3n}^\top$. Here, $\mathbf{e}_{ip}$ and $\mathbf{e}_{in}$ denote the unit vectors along the i-th coordinate axis in $\mathbb{R}^p$ and $\mathbb{R}^n$, respectively. In Figure 1, we report the simulated type I error rates for both the statistics (I.6) and (I.7) in the settings (I)–(III) for the noise matrices. We find that both statistics combined with the critical values in Table I can attain reasonable accuracy even when $n = 200$.

Second, we examine the power of the statistics under the nominal level 0.1 when $r_0 = 0$ in (I.2). We set the alternative as

$$\mathbf{H}_a: \ R = d\mathbf{e}_{1p}\mathbf{e}_{1n}^\top, \quad \text{for some fixed value } d > 0. \tag{IV.1}$$

In Figure 2, we report the simulated power for both the statistics (I.6) and (I.7) as d increases, where we take $c_n = 2$ and the settings (I)–(III) for the noise matrices. We see that both statistics have high power even for a not so large n, $n = 200$, as long as d is above some threshold. Furthermore, when d is in a certain range, we find that the statistic $\mathbb{T}_{r_0}$ in

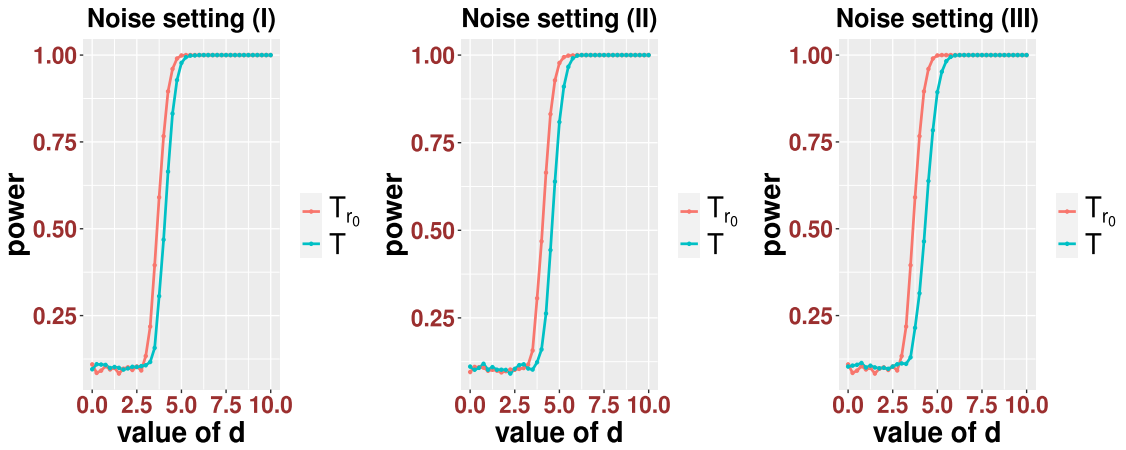


Fig. 2. Simulated power of the statistics $\mathbb{T}$ and $\mathbb{T}_{r_0}$ for the alternative (IV.1) under the nominal level 0.1. We take $n = 200$ and $c_n = 2$. We report the results based on 2,000 Monte-Carlo simulations and the critical values from Table I.

(I.7) has better performance in terms of power than the statistic $\mathbb{T}$ in (I.6). Finally, the statistic $\mathbb{T}_{r_0}$ starts to have non-zero power for smaller values of d compared to $\mathbb{T}$. This enables us to study a wider range of alternatives in terms of the d value. We expect that this is due to the fact that the statistic $\mathbb{T}$ needs a larger critical value to reject $\mathbf{H}_0$ as illustrated in Table I.

V. Proof Strategies

In this section, we describe the main strategy for the proof of Theorem 3. All the technical details can be found in the appendix. From the theoretical point of view, our proof of Theorem 3 employs the following three step strategy.

Step 1: Proving a local law on the Stieltjes transform of the random Gram matrix $\mathcal{Q}$, $m_{\mathcal{Q}}(z) := p^{-1}\mathrm{tr}(\mathcal{Q} - z)^{-1}$. This is needed in order to check the square root behavior of the ESD of $\mathcal{Q}$ around the right edge.

Step 2: Establishing the asymptotic Tracy-Widom law for the edge eigenvalues of the Gaussian divisible random Gram matrix $\mathcal{Q}_t$ in Theorem 2 for a small $t > 0$.

Step 3: Showing that $\mathcal{Q}$ has the same edge eigenvalue statistics as $\mathcal{Q}_t$ asymptotically.

This three step strategy has been widely used in the proof of bulk universality of random matrices [30]–[32], [34]. For a more extensive review, we refer the reader to [33] and references therein. However, it has been rarely (if any) used in the study of the edge eigenvalues of random Gram matrices. One of the main reasons is that the above Step 2 for Gram type random matrices—the core of the strategy—was not well-understood previously.

Regarding the proof of Theorem 3, even though the results of Step 1 have been established in [2], [3], Steps 2 and 3 are still missing. For Step 3, we can employ some standard resolvent comparison arguments developed in e.g. [7], [18], [35], [48], [54], [66], [74]. In this paper, we mainly focus on Step 2, which is completed by Theorem 2. We will provide the formal statement of Theorem 2 in Theorem 4. For this purpose, we first need to introduce some new notations.

Let Y be a $p \times n$ data matrix, and X be an independent $p \times n$ random matrix whose entries are i.i.d. centered Gaussian random variables with variance n^{-1}. Since the multivariate Gaussian distribution is rotationally invariant under orthogonal transforms, for any $t > 0$ we have that

$$Y + \sqrt{t}X \stackrel{d}{=} U_1 \left(W + \sqrt{t}X\right) U_2^\top,$$

where $Y = U_1 W U_2^\top$ is a singular value decomposition of Y with W being a $p \times n$ rectangular diagonal matrix,

$$W := \begin{pmatrix} D & 0 \end{pmatrix}, \quad D = \mathrm{diag}(\sqrt{d_1}, \cdots, \sqrt{d_p}).$$

Here, $\sqrt{d_1} \geq \sqrt{d_2} \geq \cdots \geq \sqrt{d_p} > 0$ are the singular values of Y arranged in descending order. Thus, to study the singular values of $Y + \sqrt{t}X$, it suffices to assume that the initial data matrix is W. We assume that the ESD of $V := WW^\top$ has a regular square root behavior near the spectral edge, which is generally believed to be a necessary condition for the appearance of the Tracy-Widom law. Following [51], we state the regularity conditions in terms of the Stieltjes transform of V,

$$m_V(z) := \frac{1}{p}\mathrm{tr}\,(V - z)^{-1} = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{d_i - z}, \quad z \in \mathbb{C}_+.$$

Definition 1 (η_*-regular): Let η_* be a deterministic parameter satisfying $\eta_* := n^{-\phi_*}$ for some constant $0 < \phi_* \leq 2/3$. We say V is η_*-regular around the right-edge $\lambda_+ := d_{i_0}$ for a fixed $i_0 \in \mathbb{N}$, if the following properties hold for some constants $c_V, C_V > 0$.

(i) For $z = E + \mathrm{i}\eta$ with $\lambda_+ - c_V \leq E \leq \lambda_+$ and $\eta_* + \sqrt{\eta_*|\lambda_+ - E|} \leq \eta \leq 10$, we have

$$\frac{1}{C_V}\sqrt{|\lambda_+ - E| + \eta} \leq Im m_V(z) \leq C_V\sqrt{|\lambda_+ - E| + \eta}, \tag{V.1}$$

and for $z = E + \mathrm{i}\eta$ with $\lambda_+ \leq E \leq \lambda_+ + c_V$ and $\eta_* \leq \eta \leq 10$, we have

$$\frac{1}{C_V}\frac{\eta}{\sqrt{|\lambda_+ - E| + \eta}} \leq Im m_V(z) \leq C_V\frac{\eta}{\sqrt{|\lambda_+ - E| + \eta}}. \tag{V.2}$$

(ii) There are no eigenvalues d_i of V insider the interval $[\lambda_+ + \eta_*, \lambda_+ + c_V]$.
(iii) We have $2c_V \le \lambda_+ \le C_V/2$ and $\|V\| \le N^{C_V}$.

Remark 5: For our setting in Theorem 3, the index i_0 is equal to $r+1$, which labels the first non-outlier eigenvalue of V. The motivation for (i) is as follows: if $m(z)$ is the Stieltjes transform of a density ρ with square root behavior around λ_+, i.e.,

$$\rho(x) \sim \sqrt{(\lambda_+ - x)_+}, \tag{V.3}$$

then (V.1) and (V.2) hold for $Imm(z)$ with $\eta_* = 0$. For a general $\eta_* > 0$, (V.1) and (V.2) essentially mean that the empirical spectral density of V behaves like a square root function near λ_+ on any scale larger than η_*. The condition $\eta \le 10$ in the definition is purely for definiteness of presentation—we can replace 10 with any constant of order 1.

Regarding t as a time parameter, we are interested in the dynamics of the edge eigenvalues of $\mathcal{Q}_t := (W + \sqrt{t}X)(W + \sqrt{t}X)^\top$ with respect to t for $0 < t \ll 1$. Let $\rho_{w,t}$ be the asymptotic spectral density of $\mathcal{Q}_t$, and $m_{w,t}$ be the corresponding Stieltjes transform. It is known that for any $t > 0$, $m_{w,t}$ is the unique solution to

$$m_{w,t} = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{\frac{d_i}{1+c_n t m_{w,t}} - (1+c_n t m_{w,t})z + t(1-c_n)}, \tag{V.4}$$

such that $Imm_{w,t} > 0$ for $z \in \mathbb{C}_+$ [22], [23], [71]. Adopting the notations from free probability theory, we shall call $\rho_{w,t}$ the *rectangular free convolution* (RFC) of $\rho_{w,0}$ with Marchenko-Pastur (MP) law at time t. Let $\lambda_{+,t}$ be the rightmost edge of the bulk component of $\rho_{w,t}$. By Lemma 11, we know that $\rho_{w,t}$ has a square root behavior near $\lambda_{+,t}$.

We introduce the notation

$$\zeta_t(z) := [1 + c_n t m_{w,t}(z)]^2 z - (1-c_n)t[1 + c_n t m_{w,t}(z)], \tag{V.5}$$

which is the so-called subordination function for the RFC. Then, we define the function

$$\Phi_t(\zeta) = [1 - c_n t m_{w,0}(\zeta)]^2 \zeta + (1-c_n)t[1 - c_n t m_{w,0}(\zeta)], \tag{V.6}$$

and the parameter

$$\gamma_n \equiv \gamma_n(t) := \left(\frac{1}{2}\left[4\lambda_{+,t}\zeta_{+,t} + (1-c_n)^2 t^2\right] c_n^2 t^2 \Phi_t''(\zeta_{+,t})\right)^{-1/3}, \tag{V.7}$$

where we have abbreviated that $\zeta_{+,t} \equiv \zeta_t(\lambda_{+,t})$. Here we used the short-hand notation $\zeta_t(\lambda_{+,t}) \equiv \lim_{\eta \downarrow 0} \zeta_t(\lambda_{+,t} + \mathrm{i}\eta)$. Now we are ready to give the formal statement of Theorem 2.

Theorem 4: Suppose W is η_*-regular in the sense of Definition 1 with $\eta_* = n^{-\phi_*}$. Suppose t satisfies $n^\epsilon \eta_* \le t^2 \le n^{-\epsilon}$ for a small constant $\epsilon > 0$. Fix any $k \in \mathbb{N}$, and let $f : \mathbb{R}^k \to \mathbb{R}$ be a test function such that

$$\|f\|_\infty \le C, \quad \|\nabla f\|_\infty \le C,$$

for a constant $C > 0$. Denote the eigenvalues of $\mathcal{Q}_t$ by $\lambda_1(t) \ge \lambda_2(t) \ge \cdots \ge \lambda_p(t)$. Then, we have that

$$\lim_{n\to\infty} \mathbb{E}\Big[f\Big(\gamma_n p^{2/3}(\lambda_{i_0}(t) - \lambda_{+,t}), \cdots, \gamma_n p^{2/3}(\lambda_{i_0+k-1}(t) - \lambda_{+,t})\Big)\Big] = \lim_{n\to\infty} \mathbb{E} f\left(p^{2/3}(\mu_1^{\mathrm{GOE}} - 2), \cdots, p^{2/3}(\mu_k^{\mathrm{GOE}} - 2)\right), \tag{V.8}$$

where we recall that μ_i^{GOE} are the eigenvalues of GOE as given by (I.8).

Since the edge eigenvalues of GOE at ± 2 obey the type-1 TW fluctuation [68], [69], by Theorem 4 and the Portmanteau lemma we immediately obtain that

$$\lim_{n\to\infty} \mathbb{P}(\gamma_n p^{2/3}(\lambda_1(t) - \lambda_{+,t}) \le x) = F_1(x) \text{ for all } x \in \mathbb{R},$$

where recall that F_1 is the type-1 TW distribution function.

Following the literature, we shall call the evolution of $\mathcal{Q}_t$ with respect to t the *rectangular matrix Dyson Brownian motion*, while we call the evolution of the *eigenvalues* of $\mathcal{Q}_t$ with respect to t the *rectangular Dyson Brownian motion*. We remark that the edge statistics of the symmetric Dyson Brownian motion (DBM) have been studied in [51] for Wigner type matrix ensembles. The above Theorem 4 extends the result there to Gram type matrix ensembles.

Before the end of this section, we summarize the basic ideas for the proof of Theorem 4 and provide some (possibly helpful) heuristic discussions. The proof utilizes the matching and coupling strategy in [13], [51]. First, in order to see the Tracy-Widom limit, we need to show that: (i) the rectangular free convolution (RFC) has a square root behavior near the right edge in the sense of (V.3), and (ii) the edge eigenvalues of $\mathcal{Q}_t$ distribute according to the RFC on scales $\ge n^{-2/3}$. However, at $t = 0$, the conditions (V.1) and (V.2) are not strong enough for both of these purposes. We need to run the dynamics for an amount of time t_0 to regularize both the RFC and the rectangular DBM. To show (i), we need a detailed analysis of the RFC, which has been done in another paper [19]. In particular, the analysis shows that under the η_*-regular assumption, we are able to obtain the square root behavior of RFC once $t_0 \gg \sqrt{\eta_*}$. We summarize some key properties of the RFC in Appendix B-A. To show (ii), we need to prove some sharp local laws on the resolvent $(\mathcal{Q}_{t_0} - z)^{-1}$ for $z = E + \mathrm{i}\eta$ with E around the right edge and $\eta \ge n^{-2/3-\epsilon}$. These local laws are also proved in [19] and summarized in Section B-B.

Next, we consider the rectangular DBM starting with the regular initial data $\mathcal{Q}_{t_0}$ (i.e., the evolution of the eigenvalues of $\mathcal{Q}_{t_0+t}$). It is known from the literature that the rectangular DBM satisfies a system of SDEs in equation (C.2), which is the main tool for our proof. We couple it with the system of SDEs for another rectangular DBM of a properly chosen sample covariance matrix, whose Tracy-Widom law is known from the literature and whose asymptotic ESD matches that of $\mathcal{Q}_{t_0}$ around the right edge. Under this coupling, we will show that after shifted by respective right edges, the differences between the edge eigenvalues of the two rectangular DBMs are much smaller than $n^{-2/3}$ if we run them for an amount

of time t_1 so that $n^{-1/3} \ll t_1 \ll t_0$. This key result is summarized in Theorem 7. Here, $t_1 \ll t_0$ is required so that the RFC does not change much from t_0 to $t_0 + t_1$. In particular, the right edge $\lambda_{+,t}$ and the scaling factor $\gamma_n(t)$ remain approximately constant throughout the evolution. On the other hand, the condition $t_1 \gg n^{-1/3}$ is essential because the "relaxation time to equilibrium" of the coupled DBM is of order $n^{-1/3}$ at the right edge, which we will explain below.

To prove Theorem 7, it suffices to study the differences between the two coupled rectangular DBMs, denoted by $\{\lambda_i(t)\}$ and $\{\mu_i(t)\}$, respectively. For this purpose, we consider an interpolating process $z_i(t,\alpha)$ for $0 \le \alpha \le 1$ (cf. equation (C.6)), which is a rectangular DBM with initial data $z_i(0,\alpha) = \alpha\lambda_i(0) + (1-\alpha)\mu_i(0)$. Note that $z_i(t,0) = \mu_i(t)$ and $z_i(t,1) = \lambda_i(t)$, so we only need to control $\partial_\alpha z_i(t,\alpha)$ for $0 \le \alpha \le 1$. In the proof, we find that it is more convenient to work with the singular values $y_i(t,\alpha) := \sqrt{z_i(t,\alpha)}$ and its shifted (by the right edge) version $\widetilde{y}_i(t,\alpha)$. Then, it suffices to control $\partial_\alpha \widetilde{y}_i(t,\alpha)$ by analyzing a system of SDEs given by equation (C.35). However, for the analysis, we have to cut off the effect of bulk eigenvalues away from the edge, because the η_*-regular condition only describes the edge behavior of the initial data. Hence, similar to [14], [51], we localize the analysis by introducing to the SDEs of $\widetilde{y}_i(t,\alpha)$ a short-range approximation (cf. equations (C.37)–(C.39)), whose solutions are denoted by $\widehat{y}_i(t,\alpha)$. Through a careful analysis, we find that the bulk eigenvalues indeed have negligible effect and the differences $|\widehat{y}_i(t,\alpha) - \widetilde{y}_i(t,\alpha)|$ are much smaller than $n^{-2/3}$ (cf. Lemma 23).

Now, armed with the above preparation, it remains to control $\partial_\alpha \widehat{y}_i(t,\alpha)$, which turns out to satisfy a deterministic parabolic PDE in (C.60). Using the local laws for $(\mathcal{Q}_{t_0} - z)^{-1}$, we can show that the eigenvalues of $\mathcal{Q}_{t_0}$ satisfy a rigidity estimate (see Lemma 15), which implies that the initial data $\{\widehat{y}_i(t,\alpha)\}$ has an ℓ^q norm bounded by $n^{-2/3+\epsilon}$ for any $q \ge 4$ and small constant $\epsilon > 0$. The last piece is then to prove an energy estimate for this PDE, which is summarized in Proposition 2. Roughly speaking, Proposition 2 shows that the ℓ^∞ norm of the solution at time t is smaller than the ℓ^q norm of the initial data by a factor of order $n^{-1/3}t^{-1}$. Consequently, as long as $t_1 \ge n^{-1/3+\delta}$ for a constant $\delta > 0$ and ϵ is chosen small enough, the ℓ^∞ norm of the solution at time t_1 is much smaller than $n^{-2/3}$.

Combining all the above pieces shows that the eigenvalues of $\mathcal{Q}_t$ satisfy (V.8) for $t = t_0 + t_1$. We can see from the above arguments that there are two conditions that lead to a lower bound for t: $t > t_0 \gg \sqrt{\eta_*}$ to ensure a regular square root behavior of the RFC and sharp local laws for $\mathcal{Q}_{t_0}$; $t > t_1 \gg n^{-1/3}$ to ensure the "closeness" of the two coupled rectangular DBMs. Since we have assumed $0 < \phi_* \le 2/3$ in Definition 1, we only need to take $t \gg \sqrt{\eta_*}$. In fact, in the application to the proof of Theorem 3, we will take $\phi_* = 2/3$ so that we run the rectangular DBM for an amount of time $t \gg n^{-1/3}$.

Finally, we discuss the comparison argument for Step 3 of the proof of Theorem 3. First, it requires a moment matching condition, as is well-known in the random matrix theory literature. More precisely, we will construct another random Gram matrix, say $Y' = (y'_{ij})$, with independent entries that have the same mean r_{ij} but different variances $\mathrm{var}(y'_{ij}) = s_{ij} - t/n$. Then, the rectangular matrix DBM $Y' + \sqrt{t}X$ has the same mean matrix R and variance matrix S as Y. Now, applying Theorem 4 shows that the edge eigenvalues (denoted by $\lambda'_{i,t}$) of $Y' + \sqrt{t}X$ satisfy the Tracy-Widom law around the right edge (denoted by $\lambda'_{+,t}$) of the corresponding RFC. It remains to show that the limiting law of the (shifted and rescaled) edge eigenvalues $p^{2/3}(\lambda_i - \lambda_+)$, $1 \le i \le k$, of Y match that of $p^{2/3}(\lambda'_{i,t} - \lambda'_{+,t})$, $1 \le i \le k$,. This uses a standard resolvent comparison argument in the literature, and the key technical input is the local law for the resolvent of $(Y' + \sqrt{t}X)(Y' + \sqrt{t}X)^\top$, which is given in Appendix B-B. While the resolvent comparison argument is almost the same as the ones in e.g., [18], [54], it only gives that $p^{2/3}(\lambda_i - \lambda'_{+,t})$ satisfy the Tracy-Widom law. We still need to show that the difference between the right edges λ_+ and $\lambda'_{+,t}$ is much smaller than the Tracy-Widom fluctuation scale $n^{-2/3}$. By analyzing the Stieltjes transform of the RFC, we will see (cf. equation (A.39)) that for any small constant $\epsilon > 0$, with high probability

$$|\lambda_{+,t} - \lambda_+| \le n^{-2/3-\epsilon} + n^{-2/3+\epsilon}t + n^{-1+\epsilon}t^{-1}, \qquad \text{(V.9)}$$

Since we need to control the second and third terms on the right-hand side, we have to take $n^{-1/3+\delta} \le t \le n^{-\delta}$ for a constant $\delta > 0$. To summarize, for the above argument to work, we need that $n^{-1/3+\delta} \le t \le n^{-\delta} \wedge \min s_{ij}$. In particular, taking a smaller t means relaxing the lower bound on s_{ij}, so that we can handle a more general class of random Gram matrices. On the other hand, we have seen a lower bound $t \gg \sqrt{\eta_*} \vee n^{-1/3}$ for Step 2. Therefore, in the proof of Theorem 3, we will take (almost) optimal parameters: $\eta_* = n^{-2/3}$ and $t = n^{-1/3+\delta}$. This also leads to the lower bound on s_{ij} in (II.3).

APPENDIX A
PROOFS OF THEOREM 3, COROLLARY 1 AND COROLLARY 2

We will use the following notion of stochastic domination, which was first introduced in [25] and subsequently used in many works on random matrix theory. It simplifies the presentation of the results and their proofs by systematizing statements of the form "ξ is bounded by ζ with high probability up to a small power of n".

Definition 2 (Stochastic domination and high probability event): (i) Let

$$\xi = \left(\xi^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right),$$
$$\zeta = \left(\zeta^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right),$$

be two families of nonnegative random variables, where $U^{(n)}$ is a possibly n-dependent parameter set. We say ξ is stochastically dominated by ζ, uniformly in u, if for any fixed (small) $\epsilon > 0$ and (large) $D > 0$,

$$\sup_{u \in U^{(n)}} \mathbb{P}\left(\xi^{(n)}(u) > n^\epsilon \zeta^{(n)}(u)\right) \le n^{-D}$$

for large enough $n \ge n_0(\epsilon, D)$, and we will use the notation $\xi \prec \zeta$ to denote it. Throughout this paper, the stochastic

domination will always be uniform in all parameters that are not explicitly fixed, such as the matrix indices and the spectral parameter z. If for some complex family ξ we have $|\xi| \prec \zeta$, then we will also write $\xi \prec \zeta$ or $\xi = \mathrm{O}_{\prec}(\zeta)$.

(ii) We say an event Ξ holds with high probability if for any constant $D > 0$, $\mathbb{P}(\Xi) \geq 1 - n^{-D}$ for large enough n.

The following lemma collects basic properties of stochastic domination, which will be used tacitly in the following proof.

Lemma 2 (Lemma 3.2 of [10]): Let ξ and ζ be two families of nonnegative random variables, U and V be two parameter sets and $C > 0$ be a large constant.

(i) Suppose that $\xi(u,v) \prec \zeta(u,v)$ uniformly in $u \in U$ and $v \in V$. If $|V| \leq n^C$, then $\sum_{v\in V} \xi(u,v) \prec \sum_{v\in V} \zeta(u,v)$ uniformly in $u \in U$.

(ii) If $\xi_1(u) \prec \zeta_1(u)$ and $\xi_2(u) \prec \zeta_2(u)$ uniformly in $u \in U$, then $\xi_1(u)\xi_2(u) \prec \zeta_1(u)\zeta_2(u)$ uniformly in $u \in U$.

(iii) Suppose that $\Psi(u) \geq n^{-C}$ is deterministic and $\xi(u)$ satisfies $\mathbb{E}\xi(u)^2 \leq n^C$ for all $u \in U$. Then if $\xi(u) \prec \Psi(u)$ uniformly in $u \in U$, we have that $\mathbb{E}\xi(u) \prec \Psi(u)$ uniformly in $u \in U$.

We introduce the following bounded support condition, which has been used in a sequence of papers to improve the moment assumption, see e.g. [18], [20], [54], [74].

Definition 3 (Bounded Support Condition): We say a random matrix Y satisfies the bounded support condition with ϕ_n if

$$\max_{i,j} |y_{ij} - \mathbb{E}y_{ij}| \prec \phi_n, \tag{A.1}$$

where ϕ_n is a deterministic parameter satisfying that $n^{-1/2} \leq \phi_n \leq n^{-c_\phi}$ for some small constant $c_\phi > 0$. Whenever (A.1) holds, we say that Y has support ϕ_n.

We introduce the following $(p+n) \times (p+n)$ symmetric block matrix

$$H \equiv H(Y) := \begin{pmatrix} 0 & Y \\ Y^\top & 0 \end{pmatrix}, \tag{A.2}$$

and its resolvent

$$G(z) \equiv G(Y,z) := (z^{1/2}H - z)^{-1}, \quad z \in \mathbb{C}_+. \tag{A.3}$$

Moreover, for $\mathcal{Q}_1 := YY^\top$ and $\mathcal{Q}_2 := Y^\top Y$, we define their resolvents as

$$\mathcal{G}_1(z) := (\mathcal{Q}_1 - z)^{-1}, \quad \mathcal{G}_2(z) := (\mathcal{Q}_2 - z)^{-1}. \tag{A.4}$$

Using the Schur complement formula, it is easy to check that

$$\begin{aligned} G &= \begin{pmatrix} \mathcal{G}_1 & z^{-1/2}\mathcal{G}_1 Y \\ z^{-1/2}Y^\top \mathcal{G}_1 & \mathcal{G}_2 \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{G}_1 & z^{-1/2} Y \mathcal{G}_2 \\ z^{-1/2}\mathcal{G}_2 Y^\top & \mathcal{G}_2 \end{pmatrix}. \end{aligned} \tag{A.5}$$

Thus, a control of G yields directly a control of the resolvents $\mathcal{G}_1$ and $\mathcal{G}_2$. We denote the empirical spectral density ρ_1 of $\mathcal{Q}_1$ and its Stieltjes transform by

$$\rho_1 = \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\mathcal{Q}_1)}, \quad g_1(z) = \int \frac{\rho_1(\mathrm{d}x)}{x - z} = \frac{\mathrm{Tr}\,\mathcal{G}_1(z)}{p}. \tag{A.6}$$

In [3], it has been shown that if Y is centered, i.e. $R = 0$, then the diagonal entries $(\mathcal{G}_1)_{ii}$ and $(\mathcal{G}_2)_{jj}$ can be approximated by $M_{1,i}$ and $M_{2,j}$, respectively, where $M_1 = (M_{1,1}, \cdots, M_{1,p}) : \mathbb{C}_+ \to \mathbb{C}^p$ and $M_2 = (M_{2,1}, \cdots, M_{2,n}) : \mathbb{C}_+ \to \mathbb{C}^n$ are the unique solution of

$$\frac{1}{M_1} = -z - zSM_2, \quad \frac{1}{M_2} = -z - zS^\top M_1, \tag{A.7}$$

such that $ImM_{1,i}(z) > 0$, $i = 1, 2, \cdots, p$, and $ImM_{2,j}(z) > 0$, $j = 1, 2, \cdots, n$, for all $z \in \mathbb{C}_+$. Here both $1/M_1$ and $1/M_2$ denote the entrywise reciprocals. Notice that if we plug the second equation of (A.7) into the first equation, then M_1 satisfies equation (III.1), which shows that $M_1(z) = \mathbf{m}(z)$. Then we define the asymptotic matrix limit of G as

$$\begin{aligned} \Pi(z) := \mathrm{diag}\,&(M_{1,1}(z), \cdots M_{1,p}(z), \\ &M_{2,1}(z), \cdots M_{2,n}(z))\,. \end{aligned} \tag{A.8}$$

We define the following spectral domains: for some small constants $c_0, \vartheta > 0$,

$$\begin{aligned} \mathcal{D}(c_0,\vartheta) := \Big\{ &z = E + \mathrm{i}\eta : \lambda_+ - c_0 \leq E \leq \lambda_+ + c_0, \\ &n^{-1+\vartheta} \leq \eta \leq c_0^{-1} \Big\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{D}_{out}(c_0,\vartheta) :=& \mathcal{D}(c_0,\vartheta) \\ &\cap \{ z = E + \mathrm{i}\eta : E \geq \lambda_+, n\eta\sqrt{\kappa+\eta} \geq n^\vartheta \}. \end{aligned}$$

Finally, we define the distance to the rightmost edge as

$$\kappa \equiv \kappa(z) := |E - \lambda_+| \quad \text{for} \quad z = E + \mathrm{i}\eta. \tag{A.9}$$

Then, the following local law has been proved in [2].

Lemma 3 (Theorem 2.6 of [2]): Assume that Y is a $p \times n$ random matrix with real independent entries satisfying (I.5) and that for any fixed $k \in \mathbb{N}$,

$$\mathbb{E}|y_{ij} - \mathbb{E}y_{ij}|^k \leq C_k s_{ij}^{k/2}, \tag{A.10}$$

for some constant $C_k > 0$. Moreover, suppose that the variance matrix S satisfies Assumption 1, and the mean matrix is $R = 0$. Then there exists a constant $c_0 > 0$ such that the following *averaged local laws* hold for any (small) constant $\vartheta > 0$. For any $z \in \mathcal{D}(c_0,\vartheta)$, we have that

$$|g_1(z) - m(z)| \prec (n\eta)^{-1}, \tag{A.11}$$

where $m(z)$ is defined in (III.2) and $g_1(z)$ is defined in (A.6), and for any $z \in \mathcal{D}_{out}(c_0,\vartheta)$, we have a stronger estimate

$$|g_1(z) - m(z)| \prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}. \tag{A.12}$$

Both of the above estimates are uniform in the spectral parameter z.

Remark 6: Strictly speaking, the estimate (A.12) was not proved in [2]. However, its proof is standard by combining the results in [2] with a separate argument for $z \in \mathcal{D}_{out}(c_0,\vartheta)$; see e.g. the proof of (2.20) in [27].

As a consequence of (A.11) and (A.12), we obtain the following rigidity estimate in Lemma 4 for the eigenvalues

of $\mathcal{Q}_1$ near the right edge λ_+. We define the classical location γ_j of the j-th eigenvalue as

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} \rho(x)\mathrm{d}x > \frac{j-1}{p} \right\}, \tag{A.13}$$

where ρ was defined in (III.2). In other words, γ_j's are the quantiles of the asymptotic spectral density ρ of $\mathcal{Q}_1$. Note that under the above definition, we have $\gamma_1 = \lambda_+$.

Lemma 4: Under the assumptions of Lemma 3, for any j such that $\lambda_+ - c_0/2 < \gamma_j \le \lambda_+$, we have

$$|\lambda_j(\mathcal{Q}_1) - \gamma_j| \prec j^{-1/3} n^{-2/3}. \tag{A.14}$$

Proof: The estimate (A.14) follows from (A.11) and (A.12) combined with a standard argument using Helffer-Sjöstrand calculus. The details are already given in [28], [35], [66]. □

Combining Lemma 4 with the Cauchy interlacing theorem, we immediately obtain the following result when R is non-zero and satisfies Assumption 2.

Lemma 5: Assume that Y is a $p \times n$ random matrix with real independent entries satisfying (I.5) and (A.10). Suppose that the variance matrix S satisfies Assumption 1 and the mean matrix R satisfies Assumption 2. Denote the eigenvalues of $YY^\top$ by $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$. Then there exists a constant $c_0 > 0$ such that the following statements hold for any small constant $\vartheta > 0$.

(1) *Outliers*: The first r eigenvalues satisfy

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r \ge \lambda_+ + 2c_0. \tag{A.15}$$

(2) *Eigenvalues rigidity*: For any j such that $\lambda_+ - c_0/4 < \gamma_j \le \lambda_+$, we have that

$$|\lambda_{j+r} - \gamma_j| \prec j^{-1/3} n^{-2/3}. \tag{A.16}$$

(3) *Averaged local law*: (A.11) holds uniformly for all $z \in \mathcal{D}(c_0, \vartheta)$, and (A.12) holds uniformly for all $z \in \mathcal{D}_{out}(c_0, \vartheta)$.

Proof: For simplicity, we denote $\widetilde{Y} := Y - \mathbb{E}Y$, and the eigenvalues of $\widetilde{\mathcal{Q}}_1 := \widetilde{Y}\widetilde{Y}^\top$ by $\widetilde{\lambda}_1 \ge \widetilde{\lambda}_2 \ge \cdots \ge \widetilde{\lambda}_p$. As in (A.6), we define the Stieltjes transform of the ESD of $\widetilde{\mathcal{Q}}_1$ as

$$\widetilde{g}_1(z) := \frac{1}{p}\sum_{i=1}^{p} \frac{1}{\widetilde{\lambda}_i - z} = \frac{1}{p}\mathrm{Tr}\frac{1}{\widetilde{\mathcal{Q}}_1 - z}.$$

By Lemma 4, the eigenvalues $\widetilde{\lambda}_i$ satisfy that

$$|\widetilde{\lambda}_j - \gamma_j| \prec j^{-1/3} n^{-2/3}, \tag{A.17}$$

for any j satisfying $\lambda_+ - c_0/2 < \gamma_j \le \lambda_+$. By the Cauchy interlacing theorem, we have that

$$\widetilde{\lambda}_{j+r} \le \lambda_j \le \widetilde{\lambda}_{j-r}, \quad 1 \le j \le p, \tag{A.18}$$

where we adopt the conventions $\widetilde{\lambda}_j = \infty$ if $j \le 0$, and $\widetilde{\lambda}_j = 0$ if $j \ge p+1$. By the square root behavior of $\rho(x)$ around λ_+ as shown in (III.3), it is easy to get that

$$|\gamma_j - \gamma_{j+2r}| \lesssim j^{-1/3} n^{-2/3} \tag{A.19}$$

for any j satisfying $\lambda_+ - c_0 < \gamma_j \le \lambda_+$ as long as c_0 is sufficiently small. Combining (A.17), (A.18) and (A.19), we obtain (A.16).

Now suppose the mean matrix R has SVD

$$R = \sum_{i=1}^{r} \sigma_i(R) u_i v_i^\top,$$

with $\sigma_1(R) \ge \sigma_2(R) \ge \cdots \ge \sigma_r(R) \ge (4+\tau)\sqrt{\mathfrak{M}}$ by (II.7). Using Weyl's inequality for singular values, we obtain that

$$\begin{aligned}
\lambda_r &\ge \left[\sigma_r(R) - \widetilde{\lambda}_1^{1/2}\right]^2 \\
&\ge \left[(4+\tau)\sqrt{\mathfrak{M}} - \left(\lambda_+ + \mathrm{O}_\prec(n^{-2/3})\right)^{1/2}\right]^2 \\
&\ge \left[(4+\tau)\sqrt{\mathfrak{M}} - 2\sqrt{\mathfrak{M}} + \mathrm{O}_\prec(n^{-2/3})\right]^2 \\
&\ge \left[4 + 4\tau + \mathrm{O}_\prec(n^{-2/3})\right]\mathfrak{M} \\
&\ge \left[1 + \tau + \mathrm{O}_\prec(n^{-2/3})\right]\lambda_+,
\end{aligned}$$

where we used (A.17) for $\widetilde{\lambda}_1$ in the second step, and (II.5) in the third and last steps. This gives (A.15).

Finally, using (A.17) and the interlacing result (A.18), we can show that for $z \in \mathcal{D}(c_0, \vartheta)$,

$$|\widetilde{g}_1(z) - g_1(z)| \prec (n\eta)^{-1},$$

and for $z \in \mathcal{D}_{out}(c_0, \vartheta)$,

$$|\widetilde{g}_1(z) - g_1(z)| \prec [n(\kappa+\eta)]^{-1}.$$

We omit the details because it is a standard argument, which involves bounding the real and imaginary parts of $\widetilde{g}_1(z) - g_1(z)$ using (A.18). Combining the above two estimates with Lemma 3 for $\widetilde{g}_1(z)$, we conclude part (3) of Lemma 5. □

From (A.10) and Markov's inequality, we get that the matrix Y in Lemma 3 has support $\max_{i,j} s_{ij}^{1/2}$. Now combining the analysis of the vector Dyson equation (III.1) in [2] with the arguments for local law in [18], we can relax the moment condition (A.10) to a weaker bounded support condition.

Lemma 6: Assume that Y is a $p \times n$ random matrix with real independent entries satisfying (I.5). Suppose that the variance matrix S satisfies Assumption 1 and the mean matrix R satisfies Assumption 2. Moreover, assume that Y satisfies the bounded support condition (A.1) with $\phi_n \le n^{-c_\phi}$ for a small constant $c_\phi > 0$. Then there exists a constant $c_0 > 0$ such that the following estimates hold for any small constant $\vartheta > 0$.

(1) *Averaged Local Law*: For any $z \in \mathcal{D}(c_0, \vartheta)$, we have that

$$|g_1(z) - m(z)| \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{1}{n\eta}, \tag{A.20}$$

and for $z \in \mathcal{D}_{out}(c_0, \vartheta)$, we have a stronger estimate

$$\begin{aligned}
|g_1(z) - m(z)| \prec &\min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{1}{n(\kappa+\eta)} \\
&+ \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}.
\end{aligned} \tag{A.21}$$

(2) *Entrywise Local Law*: For any $z \in \mathcal{D}(c_0, \vartheta)$, we have that

$$\max_{1\le i,j\le p+n} |G_{ij}(z) - \Pi_{ij}(z)| \prec \phi_n + \sqrt{\frac{Imm(z)}{n\eta}} + \frac{1}{n\eta}, \tag{A.22}$$

where Π is defined in (A.8).

All of the above estimates are uniform in the spectral parameter z.

Proof: With the stability analysis of equation (III.1) in [2, Section 3], we can repeat the same proofs for Lemma 3.11 of [18] and Theorem 3.6 of [74] to conclude (A.20)–(A.22). We omit the details. □

Now we are ready to give the proof of Theorem 3.

Proof of Theorem 3: Using the estimates in Lemma 6, we can repeat the proof for [18, Theorem 2.7] almost verbatim to conclude (III.7) and the following universality result as $n \to \infty$:

$$\mathbb{P}\left[\left(p^{2/3}(\lambda_{i+r} - \lambda_+) \le x_i\right)_{1\le i\le k}\right] - \mathbb{P}^G\left[\left(p^{2/3}(\lambda_{i+r} - \lambda_+) \le x_i\right)_{1\le i\le k}\right] \to 0 \tag{A.23}$$

for any $(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, where $\mathbb{P}^G$ denotes the law for $Y = (y_{ij})$ with independent Gaussian entries satisfying (I.5). To conclude (III.5) and (III.6), it remains to show that $(\varpi^{2/3}p^{2/3}(\lambda_{i+r} - \lambda_+))_{1\le i\le k}$ has the same asymptotic distribution as $(p^{2/3}(\mu_i^{\mathrm{GOE}} - 2))_{1\le i\le k}$ in the Gaussian case. For simplicity of notations, we only write down details of the proof for the $r = 0$ case, which is based on Theorem 4, Lemma 3 and Lemma 4. The argument for the $r > 0$ case is similar and will be discussed at the end of the proof.

Let $t_0 = n^{-1/3+\varepsilon_0}$ for a small constant $\varepsilon_0 < \varepsilon_*$, where recall that ε_* is the constant in (II.3). Then, we pick the initial data matrix W to be a $p \times n$ random matrix with independent Gaussian entries satisfying

$$\mathbb{E}w_{ij} = 0, \quad \mathbb{E}w_{ij}^2 = s_{ij} - t_0/n.$$

Let X be an independent $p \times n$ matrix with i.i.d. Gaussian entries of mean zero and variance n^{-1}. Then, we have that

$$Y \overset{d}{=} W + \sqrt{t_0}X.$$

We regard $W + \sqrt{t}X$ as a rectangular matrix DBM starting at W, and at time t_0 it has the same distribution as Y.

We now fix the notations for the proof. First, in light of (A.23), we denote the eigenvalues of $\mathcal{Q} := (W + \sqrt{t_0}X)(W + \sqrt{t_0}X)^\top$ by $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$. We define its asymptotic spectral density ρ and the corresponding Stieltjes transform $m(z)$ as in (III.2). Moreover, let λ_+ be the right edge of ρ, and γ_j be the quantiles of ρ defined as in (A.13). We denote the variance matrix of W by $S_w = (s_{ij} - t_0/n : 1 \le i \le p, 1 \le j \le n)$, and let $\mathbf{M}_w(z) = (M_{w,1}(z), \cdots, M_{w,p}(z)) : \mathbb{C}_+ \to \mathbb{C}^p$ be the unique solution to the vector Dyson equation

$$\frac{1}{\mathbf{M}_w} = -z\mathbf{1} + S_w\frac{1}{\mathbf{1} + S_w^\top \mathbf{M}_w}, \tag{A.24}$$

such that $ImM_{w,k}(z) > 0$, $k = 1, 2, \cdots, p$, for any $z \in \mathbb{C}_+$. Then, we define $M_w(z) := p^{-1}\sum_k M_{w,k}(z)$, which is the Stieltjes transform of the asymptotic spectral density of $WW^\top$, denoted by ρ_w. We denote the right edge of ρ_w by $\lambda_{+,w}$, and define the quantiles of ρ_w as

$$\gamma_{j,w} := \sup_x\left\{\int_x^{+\infty} \rho_w(x)dx > \frac{j-1}{p}\right\}, \tag{A.25}$$

for $1 \le j \le p$. Finally, following the notations in Section V, we denote

$$m_V(z) \equiv m_{w,0}(z) := p^{-1}\mathrm{tr}(WW^\top - z)^{-1},$$

and the eigenvalues of $WW^\top$ by $d_1 \ge d_2 \ge \cdots \ge d_p$. Then, we define $m_{w,t}$ as in (V.4), and let $\lambda_{+,t}$ be the rightmost edge of the rectangular free convolution $\rho_{w,t}$.

We take $\eta_* = n^{-2/3+\varepsilon_1}$ for a small enough constant $0 < \varepsilon_1 < \varepsilon_0$. We first verify that m_V is η_*-regular in the sense of Definition 1. Notice that W is also a random Gram matrix satisfying the assumptions of Lemma 3. Denoting $z = E + \mathrm{i}\eta$ and $\kappa = |E - \lambda_{+,w}|$, by (A.11) and (A.12) we have that for $\lambda_{+,w} - c_0 \le E \le \lambda_{+,w}$ and $n^{-2/3+\vartheta} \le \eta \le 10$,

$$|m_{w,0}(z) - M_w(z)| \prec (n\eta)^{-1}, \tag{A.26}$$

and for $\lambda_{+,w} \le E \le \lambda_{+,w} + c_0$ and $n^{-2/3+\vartheta} \le \eta \le 10$,

$$|m_{w,0}(z) - M_w(z)| \prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}. \tag{A.27}$$

Moreover, as a consequence of the square root behavior of $\rho_{+,w}$ around $\lambda_{+,w}$ as given by (III.3), it is easy to show that

$$|M_w(z)| \sim 1, \quad ImM_w(z) \sim \begin{cases}\eta/\sqrt{\kappa+\eta}, & \text{if } E \ge \lambda_{+,w}\\ \sqrt{\kappa+\eta}, & \text{if } E \le \lambda_{+,w},\end{cases} \tag{A.28}$$

for any $z = E + \mathrm{i}\eta$ satisfying that $\lambda_{+,w} - c_0 \le E \le \lambda_{+,w} + c_0$ and $0 \le \eta \le c_0^{-1}$ for a small enough constant $c_0 > 0$. In this paper, given two sequences of positive values a_n and b_n, we use $a_n \sim b_n$ to mean that there exists a constant $C > 0$ so that $C^{-1}a_n \le b_n \le Ca_n$. Finally, using (A.14) we get that

$$|d_j - \gamma_{j,w}| \prec j^{-1/3}n^{-2/3}, \tag{A.29}$$

for any j such that $\lambda_{+,w} - c_0/2 < \gamma_{j,w} \le \lambda_{+,w}$. Combining the above estimates (A.26)–(A.29), we obtain that for some constants $0 < c_V < c_0/2$ and $C_V > 0$, the following estimates hold on a high probability event Ξ: for $d_1 - c_V \le E \le d_1$ and $\eta_* \le \eta \le 10$,

$$\frac{1}{C_V}\sqrt{|d_1 - E| + \eta} \le Imm_{w,0}(E + \mathrm{i}\eta) \le C_V\sqrt{|d_1 - E| + \eta}\,;$$

for $d_1 \le E \le d_1 + c_V$ and $\eta_* \le \eta \le 10$,

$$\frac{1}{C_V}\frac{\eta}{|d_1 - E| + \eta} \le Imm_{w,0}(E + \mathrm{i}\eta) \le C_V\frac{\eta}{|d_1 - E| + \eta}\,.$$

Thus, on event Ξ, m_V is η_*-regular. Then, applying Theorem 4 to $\mathcal{Q} = (W + \sqrt{t_0}X)(W + \sqrt{t_0}X)^\top$, we conclude that there

exists a parameter $\gamma_n \sim 1$ such that for any fixed $k \in \mathbb{N}$,

$$\begin{aligned}&\left(\gamma_n p^{2/3}(\lambda_i - \lambda_{+,t_0})\right)_{1\le i\le k}\\ &\overset{d}{\sim} \left(p^{2/3}(\mu_i^{\mathrm{GOE}} - 2)\right)_{1\le i\le k},\end{aligned} \tag{A.30}$$

where $\overset{d}{\sim}$ means that the two random vectors have the same asymptotic distribution. Now, to conclude the proof, it remains to show that

$$p^{2/3}|\lambda_{+,t_0} - \lambda_+| \to 0 \quad \text{in probability.} \tag{A.31}$$

We recall that λ_+ is the right edge of the asymptotic density ρ, which by definition is also the rectangular free convolution of ρ_w with MP law at time t_0. On the other hand, for a given W, λ_{+,t_0} is the right edge of $\rho_{w,t}$, which is the rectangular free convolution of $\rho_{w,0} := p^{-1}\sum_{i=1}^p \delta_{d_i}$ with MP law at time t_0. Hence λ_{+,t_0} and λ_+ are different quantities, but we can control their difference using (A.26), (A.27) and (A.29).

Recalling the notation in (V.5), we denote

$$\begin{aligned}\zeta_{+,t_0} :=&[1 + c_n t_0 m_{w,t_0}(\lambda_{+,t_0})]^2\lambda_{+,t_0}\\ &- (1 - c_n)t_0[1 + c_n t_0 m_{w,t_0}(\lambda_{+,t_0})],\end{aligned}$$

and

$$\begin{aligned}\zeta_{+} :=&[1 + c_n t_0 m(\lambda_{+})]^2\lambda_{+}\\ &- (1 - c_n)t_0[1 + c_n t_0 m(\lambda_{+})].\end{aligned}$$

Using (B.11) below and (A.29), we can obtain that

$$|\zeta_+ - \lambda_{+,w}| \sim |\zeta_{+,t_0} - \lambda_{+,w}| \sim t_0^2. \tag{A.32}$$

Then, repeating the proof of Lemma 13 (which is given in [19, Lemma A.2]), we can obtain that

$$\begin{aligned}|\lambda_{+,t_0} - \lambda_+| \lesssim& |\zeta_{+,t_0} - \zeta_+| + t_0|M_w(\zeta_{+,t_0}) - M_w(\zeta_+)|\\ &+ t_0\left|m_{w,0}(\zeta_{+,t_0}) - M_w(\zeta_{+,t_0})\right|,\end{aligned} \tag{A.33}$$

and

$$|\zeta_{+,t_0} - \zeta_+| \lesssim t_0^3\left|m'_{w,0}(\zeta_{+,t_0}) - M'_w(\zeta_{+,t_0})\right|. \tag{A.34}$$

Using the definition of $\gamma_{j,w}$, we can get that

$$\begin{aligned}&\left|m'_{w,0}(\zeta_{+,t_0}) - M'_w(\zeta_{+,t_0})\right|\\ &= \Big|\frac{1}{p}\sum_j \frac{1}{(d_j - \zeta_{+,t_0})^2} - \int_0^{\lambda_{+,w}} \frac{\rho_w(x)}{(x - \zeta_{+,t_0})^2}\mathrm{d}x\Big|\\ &\le \sum_{j:\gamma_{j,w} > \lambda_{+,w} - c_0/2}\int_{\gamma_{j+1,w}}^{\gamma_{j,w}} \left|\frac{\rho_w(x)}{(d_j - \zeta_{+,t_0})^2} - \frac{\rho_w(x)}{(x - \zeta_{+,t_0})^2}\right|\mathrm{d}x\\ &\quad + \mathrm{O}(1)\\ &\prec \sum_{j:\gamma_{j,w} > \lambda_{+,w} - c_0/2}\int_{\gamma_{j+1,w}}^{\gamma_{j,w}} \frac{j^{-1/3}n^{-2/3}(|\lambda_{+,w} - x| + t_0^2)\rho_w(x)}{|x - \zeta_{+,t_0}|^4}\mathrm{d}x\\ &\quad + \mathrm{O}(1)\\ &\lesssim n^{-1}\int_{\lambda_{+,w} - c_0/2}^{\lambda_{+,w}} \frac{(\lambda_{+,w} - x) + t_0^2}{|(\lambda_{+,w} - x) + t_0^2|^4}\mathrm{d}x + \mathrm{O}(1)\\ &\lesssim \frac{1}{nt_0^4}.\end{aligned} \tag{A.35}$$

Here in the third step we used that for $\gamma_{j+1,w} \le x \le \gamma_{j,w}$,

$$|(x - \zeta_{+,t_0})^2 - (d_j - \zeta_{+,t_0})^2| \prec j^{-1/3}n^{-2/3}(|\lambda_{+,w} - x| + t_0^2),$$

by (A.29), (A.32) and $\lambda_{+,w} - \gamma_{j+1,w} \sim j^{2/3}n^{-2/3}$. In the fourth step, we used that $\rho_w(x) \sim \sqrt{\lambda_{+,w} - x}$ and $j^{-1/3} \sim n^{-1/3}(\lambda_{+,w} - x)^{-1/2}$. Plugging (A.35) into (A.34), we obtain that

$$|\zeta_{+,t_0} - \zeta_+| \prec n^{-1}t_0^{-1}. \tag{A.36}$$

Moreover, as a consequence of the square root behavior of ρ_w around λ_+, it is easy to check that

$$\begin{aligned}&t_0|M_w(\zeta_{+,t_0}) - M_w(\zeta_+)|\\ &\lesssim t_0\frac{|\zeta_{+,t_0} - \zeta_+|}{\min\{|\zeta_+ - \lambda_{+,w}|^{1/2}, |\zeta_{+,t_0} - \lambda_{+,w}|^{1/2}\}}\\ &\prec n^{-1}t_0^{-1},\end{aligned} \tag{A.37}$$

where we used (A.32) and (A.36) in the last step. Finally, we need to bound $|m_{w,0}(\zeta_{+,t_0}) - M_w(\zeta_{+,t_0})|$. Denote $z_0 := \zeta_{+,t_0} + \mathrm{i}\eta_0$ with $\eta_0 := n^{-2/3+\vartheta}$ for some small constant $\vartheta > 0$. We now decompose $m_{w,0}(\zeta_{+,t_0}) - M_w(\zeta_{+,t_0})$ as

$$\begin{aligned}&m_{w,0}(\zeta_{+,t_0}) - M_w(\zeta_{+,t_0})\\ &= m_{w,0}(z_0) - M_w(z_0) + \mathcal{K}_1 + \mathcal{K}_2,\end{aligned}$$

where

$$\begin{aligned}&\mathcal{K}_1\\ &:= \sum_{j:\gamma_{j,w} > \lambda_{+,w} - c_0/2}\left(\frac{1}{p}\frac{1}{d_j - \zeta_{+,t_0}} - \int_{\gamma_{j+1,w}}^{\gamma_{j,w}} \frac{\rho_w(x)}{x - \zeta_{+,t_0}}\mathrm{d}x\right)\\ &- \sum_{j:\gamma_{j,w} > \lambda_{+,w} - c_0/2}\left(\frac{1}{p}\frac{1}{d_j - z_0} - \int_{\gamma_{j+1,w}}^{\gamma_{j,w}} \frac{\rho_w(x)}{x - z_0}\mathrm{d}x\right),\\ &\mathcal{K}_2:\\ &= \sum_{j:\gamma_{j,w} \le \lambda_{+,w} - c_0/2}\left(\frac{1}{p}\frac{1}{d_j - \zeta_{+,t_0}} - \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \frac{\rho_w(x)}{x - \zeta_{+,t_0}}\mathrm{d}x\right)\\ &- \sum_{j:\gamma_{j,w} \le \lambda_{+,w} - c_0/2}\left(\frac{1}{p}\frac{1}{d_j - z_0} - \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \frac{\rho_w(x)}{x - z_0}\mathrm{d}x\right).\end{aligned}$$

By (A.27), we have

$$|m_{w,0}(z_0) - M_w(z_0)| \prec \frac{1}{nt_0^2} + \frac{1}{(n\eta_0)^2 t_0}.$$

Using (A.29), it is easy to bound $\mathcal{K}_2 \lesssim \eta_0$ with high probability. Then using a similar argument as for (A.35), we can bound

$$\begin{aligned}\mathcal{K}_1 &\prec \sum_{j:\gamma_{j,w} > \lambda_{+,w} - c_0/2}\int_{\gamma_{j+1,w}}^{\gamma_{j,w}} \frac{(|\lambda_{+,w} - x| + t_0^2)\rho_w(x)}{j^{1/3}n^{2/3}|x - \lambda_{+,t_0}|^2}\mathrm{d}x\\ &\lesssim n^{-1}\int_{\lambda_{+,w} - c_0/2}^{\lambda_{+,w}} \frac{\mathrm{d}x}{(\lambda_{+,w} - x) + t_0^2} \lesssim n^{-1}\log n.\end{aligned}$$

Combining the above three estimates, we get that

$$|m_{w,0}(\lambda_{+,t_0}) - M_w(\lambda_{+,t_0})| \prec \eta_0 + \frac{1}{nt_0^2} + \frac{1}{(n\eta_0)^2 t_0}. \tag{A.38}$$

Now, with (A.33), (A.36), (A.37) and (A.38), we can bound that

$$|\lambda_{+,t_0} - \lambda_+| \prec t_0\eta_0 + \frac{1}{nt_0} + \frac{1}{(n\eta_0)^2}. \tag{A.39}$$

Plugging into $t_0 = n^{-1/3+\varepsilon_0}$ and $\eta_0 = n^{-2/3+\vartheta}$, we conclude (A.31) as long as ε_0 and ϑ are chosen such that $\varepsilon_0 + \vartheta < 1/3$.

Combining (A.30) and (A.31), we obtain that $(\gamma_n p^{2/3}(\lambda_i - \lambda_+))_{1\le i\le p}$ converges weakly to the Tracy-Widom law. Furthermore, matching the gap between the quantiles γ_1 and γ_2 (recall (A.13)) of the density ρ in (III.3) and the one for the semicircle law $\rho_{sc}(2-x) = \pi^{-1}\sqrt{x} + \mathrm{O}(x)$ around the right edge at 2, we see that γ_n must be $\varpi^{2/3}$. This concludes the proof of (III.5) and (III.6).

Finally, we briefly discuss the proof for the $r > 0$ case. In fact, its proof uses the same argument as above, except that we need to replace Lemma 4 with Lemma 5 and apply Theorem 4 with $i_0 = r+1$. For example, the equation (A.30) above should be replaced by

$$\left(\gamma_n p^{2/3}(\lambda_{i+r} - \lambda_{+,t_0})\right)_{1\le i\le k} \overset{d}{\sim} \left(p^{2/3}(\mu_i^{\mathrm{GOE}} - 2)\right)_{1\le i\le k}.$$

We omit the details. □

Finally, we complete the proofs of Corollaries 1 and 2 using Theorem 3.

Proof of Corollary 1: In [74], the following edge universality result was proved under the assumptions of this corollary:

$$\lim_{n\to\infty}\left\{\mathbb{P}\left[\left(p^{2/3}(\lambda_i - \lambda_+) \le x_i\right)_{1\le i\le k}\right] - \mathbb{P}^G\left[\left(p^{2/3}(\lambda_i - \lambda_+) \le x_i\right)_{1\le i\le k}\right]\right\} = 0, \tag{A.40}$$

for all $(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, where $\mathbb{P}^G$ denotes the law for $\mathcal{N}$ with i.i.d. Gaussian entries of mean zero and variance n^{-1}. In particular, the condition (III.8) is not necessary if A or B is diagonal. Note that if $\mathcal{N}$ is Gaussian, then using the rotational invariance of multivariate Gaussian distribution, we can reduce $Q = YY^\top$ to a random Gram matrix satisfying (I.5) with $R = 0$ and variance matrix $S = ((a_ib_j)/n)$. Furthermore, notice that (III.9) is stronger than (II.3) and equivalent to (A3) of Assumption 1. Hence $YY^\top$ satisfies the assumptions of Theorem 3 with $r = 0$, which immediately concludes the proof. □

Remark 7: Regarding Example 1, suppose there are some spikes in the eigenvalue spectrum of A and B such that $a_1 \ge \cdots \ge a_r \ge a_{r+1} + \tau$ and $b_1 \ge \cdots \ge b_s \ge b_{s+1} + \tau$ for some $r, s \in \mathbb{N}$ and a small constant $\tau > 0$. Then it is easy to check that

$$\min\left\{\inf_{1\le i\le p}\frac{1}{p}\sum_l \frac{1}{\epsilon + |a_i - a_l|^2}, \inf_{1\le j\le n}\frac{1}{n}\sum_l \frac{1}{\epsilon + |b_j - b_l|^2}\right\} \lesssim \frac{1}{n\varepsilon} + 1,$$

and the condition (II.10) cannot hold for all n. Hence the condition (II.10) rules out the existence of outliers.

But the condition (II.10) sometimes is too strong because it does not allow for any spikes or isolated eigenvalues in the eigenvalue spectrum of A and B. (Here by an isolated eigenvalue of A, we mean an a_i such that $a_{i+1} + \tau \le a_i \le a_{i-1} - \tau$ for some $1 \le i \le p$ and a small constant $\tau > 0$. For the isolated eigenvalues of B, we have a similar definition.) On the other hand, in [20] we have found that a spike of A or B gives rise to an outlier only when it is above the BBP transition threshold. In fact, the following weaker regularity condition was used in [20], [74]. For $\mathbf{m}(z)$ in (III.1), we define another two holomorphic functions

$$m_{1c}(z) := \frac{1}{n}\sum_{i=1}^p a_i m_i(z),$$
$$m_{2c}(z) := \frac{1}{n}\sum_{j=1}^n \frac{b_j}{-z(1 + b_j m_{1c}(z))}.$$

Then, we say that the spectral edge λ_+ is *regular* if for some constant $\tau > 0$,

$$1 + m_{1c}(\lambda_+)b_1 \ge \tau, \quad 1 + m_{2c}(\lambda_+)a_1 \ge \tau. \tag{A.41}$$

This condition not only allows for isolated eigenvalues of A and B, but also allows for zero a_i's or b_j's, that is, the lower bounds in (III.9) can be relaxed to some extent. Compared with conditions (II.10) and (II.11), the condition (A.41) is less explicit and harder to check, but it appears more often in the random matrix theory literature.

Proof of Corollary 2: Combining (II.12) with Markov's inequality, we see that Y satisfies the bounded support condition (A.1) with $\phi_n = q^{-1} \le n^{-1/3-c_\phi}$. Then Lemma 6 holds, and in [18, Lemma 3.11] we have shown that (A.20) and (A.21) imply the following weaker rigidity estimate than (A.14):

$$|\lambda_j - \gamma_j| \prec j^{-1/3}n^{-2/3} + \phi_n^2. \tag{A.42}$$

With (A.20), (A.21) and (A.42) as the main inputs, using the same argument as for [26, Theorem 2.7], we can show that the edge statistics of $\mathcal{Q}$ match those of the Gaussian case in the sense of (A.23) as long as $\phi_n \le n^{-1/3-c_\phi}$. Then we immediately conclude the proof using Theorem 3. □

Remark 8: We make a few remarks on the technical assumptions (III.8) and $q \ge n^{1/3+c_\phi}$ in Corollaries 1 and 2, respectively. First, as mentioned in the proof of Corollary 1, we need to use the edge universality result (A.40) from [74], where the vanishing third moment condition (III.8) is needed (see the discussion below Theorem 3.6 in [74]). More precisely, a continuous self-consistent comparison argument is used in [74] to show that the non-Gaussian case is close to the Gaussian case in the sense of limiting distributions of edge eigenvalues. For the comparison argument to work, we need to match the third moment of $\widetilde{y}_{ij}$ with that of a standard Gaussian random variable, which leads to the condition (III.8). However, we believe that (III.8) is not necessary and can be removed with further theoretical development.

Second, we believe that the condition $q \ge n^{1/3+c_\phi}$ in Corollary 2 can be weakened to $q \ge n^{1/6+c_\phi}$. In fact, following the arguments in [43], we expect that (A.42) can be sharpened to

$$|\lambda_j - \gamma_j - \delta(q)| \prec j^{-1/3}n^{-2/3} + q^{-4},$$

for some deterministic shift $\delta(q) = \mathrm{O}(q^{-2})$. As long as $q \geq n^{1/6+c}$, the term q^{-4} will be much smaller than the Tracy-Widom scale $n^{-2/3}$, and the Tracy-Widom law around $\lambda_+ + \delta(q)$ can be established. However, when $q \ll n^{1/6}$, the limiting distribution of the second largest eigenvalue (i.e., the largest edge eigenvalue) of the Erdős-Rényi graph will become Gaussian [41], [42]. We conjecture that a similar phenomenon also occurs for the model in Corollary 2.

Since the above directions are not the focus of this paper, we will pursue them in future works.

APPENDIX B
RECTANGULAR FREE CONVOLUTION AND LOCAL LAWS

In this section, we collect some basic estimates on the rectangular free convolution $\rho_{w,t}$ and its Stieltjes transform $m_{w,t}$ for an η_*-regular $V = WW^\top$ as in Definition 1. Furthermore, we will state an (almost) sharp local law on the resolvent of $\mathcal{Q}_t = (W + \sqrt{t}X)(W + \sqrt{t}X)^\top$, and a rigidity estimate on the rectangular DBM $\{\lambda_i(t) : 1 \leq i \leq p\}$. These estimates will serve as important inputs for the detailed analysis of the rectangular DBM in Section C below. Most of the results in this section were proved in [19] under more general assumptions on X, and we will provide the exact reference for each of them. Without loss of generality, throughout this section, we assume that $i_0 = 1$. The general case with $i_0 > 1$ will be discussed in Remark 9 below.

A. Properties of Rectangular Free Convolution

For simplicity, we denote $b_t(z) := 1 + c_n t m_{w,t}(z)$. It is easy to see from (V.4) that b_t satisfies the following equation

$$b_t = 1 + \frac{c_n t}{p} \sum_{i=1}^{p} \frac{1}{b_t^{-1} d_i - b_t z + t(1 - c_n)}. \tag{B.1}$$

Recalling ζ_t defined in (V.5), the equation (B.1) can be also rewritten as

$$\frac{1}{c_n t}\left(1 - \frac{1}{b_t}\right) = m_{w,0}(\zeta_t). \tag{B.2}$$

Recall that $\rho_{w,t}$ is the asymptotic probability density associated with $m_{w,t}$, and let $\mu_{w,t}$ be the corresponding probability measure. Moreover, we denote the support of $\mu_{w,t}$ by $S_{w,t}$, with a right-most edge at $\lambda_{+,t}$. We first summarize some basic properties of these quantities, which have been proved in previous works [22], [23], [71].

Lemma 7 (Existence and Uniqueness of Asymptotic Density): The following properties hold for any $t > 0$.

(i) There exists a unique solution $m_{w,t}$ to equation (V.4) satisfying that $Im m_{w,t}(z) > 0$ and $Im z m_{w,t}(z) > 0$ for $z \in \mathbb{C}_+$.

(ii) For all $x \in \mathbb{R} \setminus \{0\}$, $\lim_{\eta \downarrow 0} m_{w,t}(x + \mathrm{i}\eta)$ exists, and we denote it by $m_{w,t}(x)$. The function $m_{w,t}(x)$ is continuous on $\mathbb{R}\setminus\{0\}$, and the measure $\mu_{w,t}$ has a continuous density $\rho_{w,t}$ given by $\rho_{w,t}(x) = \pi^{-1} Im m_{w,t}(x)$ on $\mathbb{R} \setminus \{0\}$. Finally, $m_{w,t}(x)$ is a solution to (V.4) for $z = x$.

(iii) For all $x \in \mathbb{R} \setminus \{0\}$, $\lim_{\eta \downarrow 0} \zeta_t(x + \mathrm{i}\eta)$ exists, and we denote it by $\zeta_t(x)$. Moreover, we have $Im \zeta_t(z) > 0$ for $z \in \mathbb{C}_+$.

(iv) For any $z \in \mathbb{C}_+$, we have $Re b_t(z) > 0$ and $|m_{w,t}(z)| \leq (c_n t |z|)^{-1/2}$.

(v) The interior $\mathtt{Int}(S_{w,t})$ of $S_{w,t}$ is given by

$$\begin{aligned}\mathtt{Int}(S_{w,t}) &= \{x > 0 : Im m_{w,t}(x) > 0\} \\ &= \{x > 0 : Im \zeta_t(x) > 0\},\end{aligned}$$

which is a subset of $\mathbb{R}_+ := \{x \in \mathbb{R} : x > 0\}$. Moreover, $\zeta_t(x) \notin \{d_1, \cdots, d_p\}$ if $x \notin \partial S_{w,t}$.

Proof: (i) follows from [23, Theorem 4.1], (ii) and (iii) follow from [22, Lemma 2.1] and [71, Proposition 1], (iv) follows from [22, Lemma 2.1], and (v) follows from [71, Propositions 1 and 2]. □

The following lemma characterizes the right-most edge of $S_{w,t}$. Using ζ_t in (V.5) and the definition of b_t, we can rewrite the equation (B.2) as

$$\Phi_t(\zeta_t(z)) = z, \tag{B.3}$$

where Φ_t is defined in (V.6). We recall that by definition

$$\begin{aligned}m_{w,0}(\zeta) &= p^{-1} \operatorname{Tr}[(WW^\top - \zeta)^{-1}] \\ &= \frac{1}{p} \int \frac{1}{x - \zeta} \mathrm{d}\mu_{w,0}(x).\end{aligned} \tag{B.4}$$

In [71], the authors characterize the support of $\mu_{\omega,t}$ and its edges using the local extrema of Φ_t on $\mathbb{R}$.

Lemma 8: Fix any $t > 0$. The function $\Phi_t(x)$ on $\mathbb{R} \setminus \{0\}$ admits $2q$ positive local extrema counting multiplicities for some $q \in \mathbb{N}$. The preminages of these extrema are denoted by $\zeta_{1,-}(t) < 0 < \zeta_{1,+}(t) \leq \zeta_{2,-}(t) \leq \zeta_{2,+}(t) \leq \cdots \leq \zeta_{q,-}(t) \leq \zeta_{q,+}(t)$, and they all belong to the set $\{\zeta \in \mathbb{R} : 1 - c_n t m_{w,0}(\zeta) > 0\}$. Moreover, the rightmost edge of $\operatorname{supp}(\mu_{w,t})$ is given by $\lambda_+(t) = \Phi_t(\zeta_{q,+}(t))$, and Φ_t is strictly increasing on the intervals $(-\infty, \zeta_{1,-}(t)]$, $[\zeta_{1,+}(t), \zeta_{2,-}(t)]$, $\cdots$, $[\zeta_{q-1,+}(t), \zeta_{q,-}(t)]$ and $[\zeta_{q,+}(t), \infty)$. Finally, for $k = 1, 2, \cdots, q$, each interval $(\zeta_{k,-}(t), \zeta_{k,+}(t))$ contains at least one of the elements in $\{d_1, \cdots, d_p, 0\}$, and in particular, $d_1 \in (\zeta_{q,-}(t), \zeta_{q,+}(t))$.

Proof: See [71, Proposition 3] and the discussion below [71, Theorem 2], or see [56, Lemma 1]. □

Now, we rewrite (B.2) into another equation in terms of ζ and z. We focus on $z \in \mathbb{C}_+$ with $Re z > 0$. Then, we can solve from (V.5) that

$$b_t = \frac{t(1 - c_n) + \sqrt{t^2(1 - c_n)^2 + 4\zeta z}}{2z}, \tag{B.5}$$

where we have chosen the branch of the solution such that Lemma 7 (iv) holds. Plugging (B.5) into (B.2), we find that (z, b_t) is a solution to (B.2) if and only if (z, ζ_t) is a solution to

$$F_t(z, \zeta) = 0, \tag{B.6}$$

where

$$\begin{aligned}F_t(z, \zeta) :=& 1 + \frac{t(1 - c_n) - \sqrt{t^2(1 - c_n)^2 + 4\zeta z}}{2\zeta} \\ &- c_n t m_{w,0}(\zeta).\end{aligned}$$

Since the two equations $\Phi_t(\zeta_t(x)) = x$ and $F_t(x, \zeta_t(x)) = 0$ are equivalent, from Lemma 8 we can obtain the following characterization of the edges of $S_{w,t}$.

Lemma 9: Denote $a_{k,\pm}(t) := \Phi_t(\zeta_{k,\pm}(t))$, $1 \le k \le q$. Then, $(a_{k,\pm}(t), \zeta_{k,\pm}(t))$ are real solutions to

$$F_t(z,\zeta) = 0, \quad \text{and} \quad \frac{\partial F_t}{\partial \zeta}(z,\zeta) = 0. \tag{B.7}$$

Proof: By chain rule, if we regard z as a function of ζ, then we have

$$0 = \frac{\mathrm{d}F_t}{\mathrm{d}\zeta} = \frac{\partial F_t}{\partial \zeta} + \frac{\partial F_t}{\partial z} z'(\zeta). \tag{B.8}$$

By Lemma 8, we have $\Phi_t'(\zeta_{k,\pm}) = 0$ since $\zeta_{k,\pm}$ are local extrema of Φ_t. Then, from equation (B.3), we can derive that

$$z'(\zeta_{k,\pm}) = \Phi_t'(\zeta_{k,\pm}) = 0,$$

with $z(\zeta_{k,\pm}) = a_{k,\pm}$. Plugging this equation into (B.8), we get

$$\frac{\partial F_t}{\partial \zeta}(a_{k,\pm}, \zeta_{k,\pm}) = 0,$$

which concludes the proof. □

Now we use Lemma 9 to derive an expression for the derivative $\partial_t \lambda_{+,t}$, which will be used in the analysis of the rectangular DBM in Section C. Taking derivative of (B.6) with respect to t and using (B.7), we get that for $z = \lambda_{+,t}$ and $\zeta_{+,t} := \zeta_t(\lambda_{+,t})$,

$$\frac{\partial F(t, \lambda_{+,t}, \zeta_{+,t})}{\partial t} + \frac{\partial F(t, \lambda_{+,t}, \zeta_{+,t})}{\partial z} \frac{\mathrm{d}\lambda_{+,t}}{\mathrm{d}t} = 0,$$

where we denoted $F(t,z,\zeta) \equiv F_t(z,\zeta)$. From this equation, we can solve that

$$\begin{aligned} \frac{\mathrm{d}\lambda_{+,t}}{\mathrm{d}t} &= \left[\frac{1-c_n}{2\zeta_{+,t}} - c_n m_{w,0}(\zeta_{+,t})\right] \\ &\quad \times \sqrt{t^2(1-c_n)^2 + 4\zeta_{+,t}\lambda_{+,t}} - \frac{(1-c_n)^2 t}{2\zeta_{+,t}} \\ &= \left[\frac{1-c_n}{2\zeta_{+,t}} - \frac{c_n m_{w,t}(\lambda_{+,t})}{b(\lambda_{+,t})}\right] \\ &\quad \times \sqrt{t^2(1-c_n)^2 + 4\zeta_{+,t}\lambda_{+,t}} - \frac{(1-c_n)^2 t}{2\zeta_{+,t}}, \end{aligned} \tag{B.9}$$

where we used (B.2) in the second step.

Next we describe some more precise properties of $\rho_{w,t}$ and $m_{w,t}$ for an η_*-regular V as in Definition 1. For the following results, we always assume that

$$t := n^{-1/3+\omega}, \text{ with } \frac{1}{3} - \frac{\phi_*}{2} + \frac{\varepsilon}{2} \le \omega \le \frac{1}{3} - \frac{\varepsilon}{2}, \tag{B.10}$$

for some constant $\varepsilon > 0$. Note that under this condition, we have $n^{\varepsilon}\eta_* \le t^2 \le n^{-\varepsilon}$.

Lemma 10 (Lemma 3.7 of [19]): Suppose $V = WW^\top$ is η_*-regular and t satisfies (B.10). Then, we have $\zeta_{+,t} \ge \lambda_+$ and

$$\zeta_{+,t} - \lambda_+ \sim t^2. \tag{B.11}$$

The following lemma describes the square root behavior of the asymptotic density $\rho_{w,t}$.

Lemma 11 (Lemmas 3.18 and 3.19 of [19]): Suppose $V = WW^\top$ is η_*-regular and t satisfies (B.10). If $\kappa := |E - \lambda_+| \le 3c_V/4$, then the asymptotic density satisfies that

$$\rho_{w,t}(E) \sim \sqrt{(\lambda_{+,t} - E)_+}. \tag{B.12}$$

Moreover, if $-\tau t^2 \le E - \lambda_{+,t} \le 0$ for a sufficiently small constant $\tau > 0$, then we have that

$$\begin{aligned} \rho_{w,t}(E) =& \frac{1}{\pi} \sqrt{\frac{2(\lambda_{+,t} - E)}{[4\lambda_{+,t}\zeta_{+,t} + (1-c_n)^2 t^2] c_n^2 t^2 \Phi_t''(\zeta_{+,t})}} \\ &\times \left[1 + \mathrm{O}\left(\frac{|E - \lambda_{+,t}|}{t^2}\right)\right], \end{aligned} \tag{B.13}$$

where $t^2 \Phi_t''(\zeta_+(t)) \sim 1$. Finally, as a consequence of (B.12), the following estimates hold:

$$|m_{w,t}| \lesssim 1, \quad Imm_{w,t}(z) \sim \begin{cases} \sqrt{\kappa+\eta}, & E \le \lambda_{+,t} \\ \frac{\eta}{\sqrt{\kappa+\eta}}, & E \ge \lambda_{+,t} \end{cases}, \tag{B.14}$$

for any $z = E + \mathrm{i}\eta$ satisfying $|E - \lambda_+| \le 3c_V/4$ and $0 \le \eta \le 10$.

We also need to control the derivative $\partial_z m_{w,t}(z)$. First, note that with the definition of $m_{w,t}$, we can get the trivial estimate

$$\begin{aligned} |\partial_z m_{w,t}(z)| &= \left| \int \frac{\mathrm{d}\mu_{w,t}(x)}{(x-z)^2} \right| \\ &\le \int \frac{\mathrm{d}\mu_{w,t}(x)}{|x-z|^2} = \frac{Imm_{w,t}}{\eta}. \end{aligned} \tag{B.15}$$

Moreover, we claim the following estimates.

Lemma 12 (Lemma 3.20 of [19]): Suppose $V = WW^\top$ is η_*-regular and t satisfies (B.10). Consider any $z = E + \mathrm{i}\eta$ with $\kappa := |E - \lambda_+| \le 3c_V/4$ and $0 \le \eta \le 10$. If $\kappa + \eta \le t^2$, then we have that

$$|\partial_z m_{w,t}(z)| \lesssim (\kappa+\eta)^{-1/2}. \tag{B.16}$$

If $\kappa + \eta \ge t^2$, we have that for $E \ge \lambda_{+,t}$,

$$|\partial_z m_{w,t}(z)| \lesssim (\kappa+\eta)^{-1/2}, \tag{B.17}$$

and for $E \le \lambda_{+,t}$,

$$|\partial_z m_{w,t}(z)| \lesssim \frac{\sqrt{\kappa+\eta}}{t\sqrt{\kappa+\eta} + \eta}. \tag{B.18}$$

Finally, in Section C, we will need to compare the edge behaviors of two free rectangular convolutions satisfying certain matching properties. Specifically, let $t_0 = N^{-1/3+\omega_0}$ for some constant $0 < \omega_0 < 1/3$. We consider two probability measures ρ_1 and ρ_2 having densities on the interval $[0, 2\psi]$ with $\psi \sim 1$ being a positive constant, such that for some constant $c_\psi > 0$ the following properties hold:

$$\rho_1(\psi - x) = \rho_2(\psi - x)\left[1 + \mathrm{O}\left(\frac{|x|}{t_0^2}\right)\right], \tag{B.19}$$

for $0 \le x \le c_\psi t_0^2$;

$$\begin{aligned} &\rho_1(x) = \rho_2(x) = 0 \quad \text{for } x \in [\psi, 2\psi], \\ &\rho_1(x) \sim \rho_2(x) \sim \sqrt{\psi - x} \quad \text{for } \ x \in [\psi - c_\psi, \psi]. \end{aligned} \tag{B.20}$$

Let $\rho_{1,t}$ and $\rho_{2,t}$ be the free rectangular convolutions of the MP law with ρ_1 and ρ_2, respectively. Moreover, the Stieltjes transform of $\rho_{i,t}$, denoted by $m_{i,t}$, satisfies a similar equation as in (B.2):

$$\frac{1}{c_n t}\left(1 - \frac{1}{b_{i,t}}\right) = \int \frac{\rho_i(x)}{x - \zeta_{i,t}} \mathrm{d}x, \quad i = 1, 2,$$

where

$$b_{i,t}(z) := 1 + c_n t m_{i,t}(z),$$
$$\zeta_{i,t}(z) := z b_{i,t}^2 - (1 - c_n) t b_{i,t}. \tag{B.21}$$

For $i = 1, 2$, let $\lambda_{+,i}(t)$ be the right edge of $\rho_{i,t}$, and denote $\zeta_{+,i}(t) := \zeta_{i,t}(\lambda_{+,i}(t))$. Due to the matching condition (B.19), we can show that $\zeta_{+,1}(t)$ and $\zeta_{+,2}(t)$ are close to each other with a distance of order $\mathrm{o}(t^2)$ for $t \ll t_0$.

Lemma 13 (Lemma A.2 of [19]): Suppose (B.19) and (B.20) hold. Then there exists a constant $C > 0$ such that for any $0 \le t \le t_0$,

$$|\zeta_{+,1}(t) - \zeta_{+,2}(t)| \le \frac{Ct^3}{t_0}, \tag{B.22}$$

and

$$|\lambda_{+,1}(t) - \psi| + |\lambda_{+,2}(t) - \psi| \le Ct. \tag{B.23}$$

The following matching estimates will play an important role in constructing the short-range approximation of the rectangular DBM in Section C-B.

Lemma 14 (Lemmas A.4 and A.5 of [19]): Suppose (B.19) and (B.20) hold, and $0 < t \le t_0 n^{-\epsilon_0}$ for a constant $\epsilon_0 > 0$. If $0 \le x \le \tau n^{-2\epsilon} t_0^2$ for some small enough constants $\tau, \epsilon > 0$, then for any (large) constant $D > 0$ we have that

$$\rho_{1,t}(\lambda_{+,1} - x) = \rho_{2,t}(\lambda_{+,2} - x)\left[1 + \mathrm{O}\left(\frac{n^\epsilon t}{t_0} + n^{-D}\right)\right], \tag{B.24}$$

and

$$\begin{aligned}&\Big|Re[m_{1,t}(\lambda_{+,1} - x) - m_{1,t}(\lambda_{+,1})] \\ &- Re[m_{2,t}(\lambda_{+,2} - x) - m_{2,t}(\lambda_{+,2})]\Big| \\ &\lesssim \left(\frac{n^\epsilon}{t_0} + \frac{n^{-D}}{t}\right) x.\end{aligned} \tag{B.25}$$

If $0 \le x \le \tau n^{-2\epsilon} t_0 t$, then for any (large) constant $D > 0$ we have that

$$\begin{aligned}&\Big|Re[m_{1,t}(\lambda_{+,1} + x) - m_{1,t}(\lambda_{+,1})] \\ &- Re[m_{2,t}(\lambda_{+,2} + x) - m_{2,t}(\lambda_{+,2})]\Big| \\ &\lesssim \left(n^\epsilon \frac{t^{1/2}}{t_0^{1/2}} + n^{-D}\frac{t_0^{1/2}}{t^{1/2}}\right) x^{1/2}.\end{aligned} \tag{B.26}$$

B. Local Laws

In this section, we state the local laws and rigidity estimates for the rectangular DBM considered in this paper. We first consider t satisfying (B.10). Define the following $(p+n) \times (p+n)$ symmetric block matrix

$$H_t := \begin{bmatrix} 0 & W + \sqrt{t}X \\ (W + \sqrt{t}X)^\top & 0 \end{bmatrix}.$$

Definition 4 (Resolvents): We define the resolvent of H_t as

$$G(z) \equiv G_t(X, W, z) := (z^{1/2} H_t - z)^{-1}, \quad z \in \mathbb{C}_+. \tag{B.27}$$

For $\mathcal{Q}_{1,t} := (W + \sqrt{t}X)(W + \sqrt{t}X)^\top$ and $\mathcal{Q}_{2,t} := (W + \sqrt{t}X)^\top (W + \sqrt{t}X)$, we define the resolvents

$$\mathcal{G}_1(z) \equiv \mathcal{G}_{1,t}(X, W, z) := (\mathcal{Q}_{1,t} - z)^{-1},$$
$$\mathcal{G}_2(z) \equiv \mathcal{G}_{2,t}(X, W, z) := (\mathcal{Q}_{2,t} - z)^{-1}. \tag{B.28}$$

We denote the empirical spectral density $\rho_{1,t}$ of $\mathcal{Q}_{1,t}$ and its transform by

$$\rho_1 \equiv \rho_{1,t}(X, W, z) := \frac{1}{p}\sum_{i=1}^p \delta_{\lambda_i(\mathcal{Q}_{1,t})},$$
$$m_1(z) \equiv m_{1,t}(X, W, z) := \int \frac{\rho_1(\mathrm{d}x)}{x - z} = \frac{\mathrm{Tr}\,\mathcal{G}_1(z)}{p}.$$

For any constant $\vartheta > 0$, we define the spectral domain

$$\begin{aligned}\mathcal{D}_\vartheta := &\left\{z = E + \mathrm{i}\eta : \lambda_{+,t} - \frac{3}{4}c_V \le E \le \lambda_{+,t}, \right. \\ &\left. \frac{n^\vartheta}{n\eta} \le \sqrt{\kappa + \eta} \le 10\right\} \\ &\bigcup \left\{z = E + \mathrm{i}\eta : \lambda_{+,t} \le E \le \lambda_{+,t} + \frac{3}{4}c_V, \right. \\ &\left. n^{-2/3+\vartheta} \le \eta \le 10\right\},\end{aligned} \tag{B.29}$$

where recall that $\lambda_{+,t}$ is the right-edge of $\rho_{w,t}$. The following theorem gives the local laws on the domain $\mathcal{D}_\vartheta$.

Theorem 5 (Theorem 2.7 of [19]): Suppose $V = WW^\top$ is η_*-regular, and t satisfies (B.10). For any constant $\vartheta > 0$, the following estimates hold uniformly in $z \in \mathcal{D}_\vartheta$:

(i) for $E \le \lambda_{+,t}$, we have

$$|m_{1,t}(z) - m_{w,t}(z)| \prec \frac{1}{n\eta}; \tag{B.30}$$

(ii) for $E \ge \lambda_{+,t}$, we have

$$|m_{1,t}(z) - m_{w,t}(z)| \prec \frac{1}{n(\kappa + \eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa + \eta}}. \tag{B.31}$$

As a consequence of this theorem, we can obtain the following rigidity estimate for the eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$ of $\mathcal{Q}_{1,t}$ near the right edge $\lambda_{+,t}$. We define the quantiles of $\rho_{w,t}$ as in (A.13): for $1 \le j \le p$,

$$\gamma_j := \sup_x \left\{\int_x^{+\infty} \rho_{w,t}(x)\mathrm{d}x > \frac{j-1}{p}\right\}. \tag{B.32}$$

Lemma 15: Suppose the local laws (B.30) and (B.31) hold. Then, for any j such that $\lambda_{+,t} - c_V/2 < \gamma_j \le \lambda_{+,t}$, we have

$$|\lambda_j - \gamma_j| \prec j^{-1/3} n^{-2/3}. \tag{B.33}$$

Proof: The estimate (B.33) follows from the local laws (B.30) and (B.31) combined with a standard argument using Helffer-Sjöstrand calculus. The details are already given in [28], [35], [66]. □

Then, we present the local laws for the case where W already satisfies a local law.

Assumption 3: Suppose $m_V(z) \equiv m_{w,0}(z)$ satisfies the following estimates for any constant $\vartheta > 0$:

$$|m_{w,0}(z) - m_c(z)| \prec \frac{1}{n\eta},$$

for $\lambda_+ - c_V \le E \le \lambda_+$ and $n^{\vartheta}(n\eta)^{-1} \le \sqrt{|E-\lambda_+|+\eta} \le 10$;

$$|m_{w,0}(z) - m_c(z)| \prec \frac{1}{n(|E-\lambda_+|+\eta)} + \frac{1}{(n\eta)^2\sqrt{|E-\lambda_+|+\eta}},$$

for $\lambda_+ \le E \le \lambda_+ + c_V$ and $n^{-2/3+\vartheta} \le \eta \le 10$. Here $m_c(z)$ is the Stieltjes transform of a deterministic probability density $\rho_c(x)$ that is compactly supported on $[0, \lambda_+]$, and satisfies $\rho_c(x) \sim \sqrt{x}$ for $\lambda_+ - c_V \le x \le \lambda_+$.

We denote the rectangular free convolution of ρ_c with MP law at time t by $\rho_{c,t}$, and its Stieltjes transform by $m_{c,t}$. We also denote the right edge of $\rho_{c,t}$ by $\lambda_{c,t}$ and define $\kappa_c := |E - \lambda_{c,t}|$. Then we define the following spectral domain

$$\begin{aligned}\mathcal{D}_{\vartheta,c} :=& \Big\{ z = E + \mathrm{i}\eta : \lambda_{c,t} - \frac{3}{4}c_V \le E \le \lambda_{c,t}, \\ & \frac{n^{\vartheta}}{n\eta} \le \sqrt{\kappa_c + \eta} \le 10 \Big\} \\ & \bigcup \Big\{ z = E + \mathrm{i}\eta : \lambda_{c,t} \le E \le \lambda_{c,t} + \frac{3}{4}c_V, \\ & n^{-2/3+\vartheta} \le \eta \le 10 \Big\}. \end{aligned} \tag{B.34}$$

Then, we have the following local law on the domain $\mathcal{D}_{\vartheta,c}$.

Theorem 6 (Theorem 2.10 of [19]): Suppose Assumption 3 holds. For any fixed constants $\vartheta, \delta > 0$, the following estimates hold uniformly in $z \in \mathcal{D}_{\vartheta,c}$ and $0 \le t \le n^{-\delta}$:

(i) for $E \le \lambda_{c,t}$, we have

$$|m_{1,t}(z) - m_{c,t}(z)| \prec \frac{1}{n\eta}; \tag{B.35}$$

(ii) for $E \ge \lambda_{c,t}$, we have

$$|m_{1,t}(z) - m_{c,t}(z)| \prec \frac{1}{n(\kappa_c+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa_c+\eta}}. \tag{B.36}$$

Again using Theorem 6, we can prove the following rigidity estimate for the eigenvalues of $\mathcal{Q}_{1,t}$ near the right edge $\lambda_{c,t}$. We define the quantiles γ_j^c as in (B.32) but with $\rho_{w,t}$ replaced by $\rho_{c,t}$.

Lemma 16: Suppose the local laws (B.35) and (B.36) hold. Then, for any j such that $\lambda_{c,t} - c_V/2 < \gamma_j^c \le \lambda_{c,t}$, we have

$$|\lambda_j - \gamma_j^c| \prec j^{-1/3}n^{-2/3}. \tag{B.37}$$

Proof: The estimate (B.37) follows from the local laws (B.35) and (B.36) combined with a standard argument using Helffer-Sjöstrand calculus. The details are already given in [28], [35], [66]. □

Remark 9: We now briefly discuss how to handle the general case with $i_0 > 1$. When $i_0 > 1$, the $i_0 - 1$ outliers will give rise to several small peaks of $\rho_{w,t}$ around the spikes d_i, $1 \le i \le i_0 - 1$. We can exclude them and only consider the bulk component of $\rho_{w,t}$ with a right edge $\lambda_{+,t}$ that is close to d_{i_0}. Then, all the results in this section still hold for the $i_0 > 1$ case with $\lambda_+ := d_{i_0}$ except that c_V needs to be chosen sufficiently small so that the spectral domains $\mathcal{D}_{\vartheta}$ and $\mathcal{D}_{\vartheta,c}$ are away from the spikes d_i, $1 \le i \le i_0 - 1$, by a distance of order 1 and j will be restricted to $j \ge i_0$ in Lemmas 15 and 16.

Appendix C
Proof of Theorem 4

This section is devoted to the proof of Theorem 4. For simplicity of presentation, we only provide the detailed proof for the $i_0 = 1$ case without outliers. The general case with $i_0 > 1$ will be discussed in Remark 10 below.

In the proof, we fix two time scales

$$t_0 = n^{\omega_0}/n^{1/3}, \quad t_1 = n^{\omega_1}/n^{1/3}, \tag{C.1}$$

for some constants ω_0 and ω_1 satisfying $1/3 - \phi_*/2 + \varepsilon/2 \le \omega_0 \le 1/3 - \varepsilon/2$ and $0 < \omega_1 < \omega_0/100$. The reason for choosing these two scales is the same as the one in [51]. That is, we first run the DBM for t_0 amount of time to regularize the global eigenvalue density, and then for the DBM from t_0 to $t_0 + t_1$, we will show that the local statistics of the edge eigenvalues converge to the Tracy-Widom law. Since $t_1 \ll t_0$, for the time period $t_0 \le t \le t_0 + t_1$ the locations of the quantiles defined in (B.32) remain approximately constant.

The eigenvalue dynamics of $\mathcal{Q}_t = (W + \sqrt{t}X)(W + \sqrt{t}X)^\top$ with respect to t is described by the *rectangular Dyson Brownian motion* defined as follows. Let $B_i(t)$, $i = 1, \cdots, p$, be independent standard Brownian motions. For $t \ge 0$, we define the process $\{\lambda_i(t) : 1 \le i \le p\}$ as the unique strong solution to the following system of SDEs [14, Appendix C]:

$$d\lambda_i = 2\sqrt{\lambda_i}\frac{dB_i}{\sqrt{n}} + \left(\frac{1}{n}\sum_{j\ne i}\frac{\lambda_i + \lambda_j}{\lambda_i - \lambda_j} + 1\right)dt, \tag{C.2}$$

for $1 \le i \le p$, with initial data

$$\lambda_i(0) := \lambda_i(\gamma_w \mathcal{Q}_{t_0}),$$
$$\gamma_w := \left(\frac{1}{2}[4\lambda_{+,t_0}\zeta_{+,t_0} + (1-c_n)^2 t_0^2]c_n^2 t_0^2 \Phi''_{t_0}(\zeta_{+,t_0})\right)^{-1/3}.$$

In other words, the initial data is chosen as the eigenvalues of the regularized matrix $\mathcal{Q}_{t_0}$, and γ_w is chosen to match the edge eigenvalue gaps of $\mathcal{Q}_{t_0}$ with those of the Wigner matrices. Here we recall that the asymptotic density $\rho_{w,t}$ is given by (B.13), while the Wigner semicircle law has density $\pi^{-1}\sqrt{(2-x)_+} + \mathrm{O}((2-x)_+)$ around 2. The system of SDEs (C.2) for the rectangular DBM is defined in a way such that for any time $t > 0$, the process $\{\lambda_i(t)\}$ has the same joint distribution as the eigenvalues of the matrix

$$\gamma_w \mathcal{Q}_{t_0 + t/\gamma_w} = (\sqrt{\gamma_w}W + \sqrt{\gamma_w t_0 + t}X)(\sqrt{\gamma_w}W + \sqrt{\gamma_w t_0 + t}X)^\top.$$

We shall denote the rectangular free convolution of the empirical spectral density of $\sqrt{\gamma_w}V$ with MP law at time $\gamma_w t_0 + t$ by $\rho_{\lambda,t}$, which gives the asymptotic ESD for $\gamma_w \mathcal{Q}_{t_0+t/\gamma_w}$. Moreover, we use $m_{\lambda,t}$ to denote the Stieltjes transform of $\rho_{\lambda,t}$. It is easy to see that the right edge of $\rho_{\lambda,t}$ is given by

$$E_\lambda(t) := \gamma_w \lambda_{+,t_0+t/\gamma_w},$$

where recall that $\lambda_{+,t}$ denotes the right edge of $\rho_{w,t}$ at time t. Note that the scaling factor γ_w is fixed throughout the evolution, but the right edge evolves in time.

We would like to compare the edge eigenvalue statistics of the rectangular DBM $\{\lambda_i(t)\}$ with those of a carefully

chosen deformed Wishart matrix. We define a $p \times p$ sample covariance matrix $\mathcal{U}\mathcal{U}^\top$, where $\mathcal{U}$ is a random matrix of the form $\mathcal{U} := \Sigma^{1/2}\mathcal{X}$. Here $\mathcal{X}$ is a $p \times n$ random matrix with i.i.d. Gaussian entries of mean zero and variance n^{-1}, and $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_p)$ is a diagonal population covariance matrix. Recall that the asymptotic ESD of $\mathcal{U}\mathcal{U}^\top$, denoted as $\rho_{\mu,0}$, is given by the multiplicative free convolution of the Marchenko-Pastur law and the ESD of Σ, which is also referred to as the deformed Marchenko-Pastur law [58]. We choose Σ such that $\rho_{\mu,0}$ matches $\rho_{\lambda,0}$ near the right edge $E_\lambda(0)$, that is, $\rho_{\mu,0}(x)$ satisfies that

$$\rho_{\mu,0}(x) = \pi^{-1}\sqrt{(E_\lambda(0) - x)_+} + \mathrm{O}((E_\lambda(0) - x)_+), \quad \text{(C.3)}$$

for x around $E_\lambda(0)$. Note that there are only two parameters to match, i.e. the right spectral edge and the curvature of the spectral density at the right edge, but there are a lot of degrees of freedom in Σ for tuning to ensure that (C.3) holds. Now we define a rectangular DBM with initial data $\{\mu_i\}$ being the eigenvalues of $\mathcal{U}\mathcal{U}^\top$. More precisely, for $t \geq 0$ we define the process $\{\mu_i(t) : 1 \leq i \leq p\}$ as the unique strong solution to the following system of SDEs:

$$d\mu_i = 2\sqrt{\mu_i}\frac{dB_i}{\sqrt{n}} + \left(\frac{1}{n}\sum_{j\neq i}\frac{\mu_i + \mu_j}{\mu_i - \mu_j} + 1\right)dt, \quad \text{(C.4)}$$

for $1 \leq i \leq p$, with initial data $\mu_i(0) := \mu_i(\mathcal{U}\mathcal{U}^\top)$. For any $t > 0$, the process $\{\mu_i(t)\}$ has the same joint distribution as the eigenvalues of the matrix $(\mathcal{U} + \sqrt{t}X)(\mathcal{U} + \sqrt{t}X)^\top$, which is still a sample covariance matrix with population covariance $\Sigma + tI$. In particular, by [53] we know that the edge eigenvalues of $\{\mu_i(t)\}$ obey the Tracy-Widom distribution asymptotically. We will denote the rectangular free convolution of $\rho_{\mu,0}$ with MP law at time t by $\rho_{\mu,t}$, which gives the asymptotic ESD for $(\mathcal{U} + \sqrt{t}X)(\mathcal{U} + \sqrt{t}X)^\top$. Furthermore, we denote the Stieltjes transform of $\rho_{\mu,t}$ by $m_{\mu,t}$, and the right edge of $\rho_{\mu,t}$ by $E_\mu(t)$. Note that we have $E_\mu(0) = E_\lambda(0)$ by (C.3).

The main result of this section is the following comparison theorem.

Theorem 7: Fix any integer $k \in \mathbb{N}$. Under the assumptions of Theorem 4, there exists a constant $\varepsilon > 0$ such that with high probability

$$\max_{1\leq i\leq k} |[\lambda_i(t_1) - E_\lambda(t_1)] - [\mu_i(t_1) - E_\mu(t_1)]| \leq n^{-2/3-\varepsilon}. \quad \text{(C.5)}$$

With Theorem 7, we can conclude Theorem 4.

Proof of Theorem 4: We take $t_0 = t - t_1$ for a small enough constant ω_1. Then, together with the fact that $\mu_i(t_1) - E_\mu(t_1)$ satisfies the Tracy-Widom fluctuation by [7], [18], [24], [53], [63], the estimate(C.5) implies (V.8). □

Remark 10: We now make some remarks about the general case with $i_0 > 1$. Its proof is almost the same as that for the $i_0 = 1$ case, except that we need to apply some standard arguments in the study of DBM regarding the reindexing of the eigenvalues and the padding with dummy particles. More precisely, in equation (C.5), we should control

$$|[\lambda_{i+i_0-1}(t_1) - E_\lambda(t_1)] - [\mu_i(t_1) - E_\mu(t_1)]| \leq n^{-2/3-\varepsilon}.$$

Then, in defining the two rectangular DBMs, we add to the initial data of the SDEs some dummy particles, which are away from the edge eigenvalues by a distance of order N^C for a large constant $C > 0$. These dummy particles have a negligible effect on the evolution of edge eigenvalues, and hence are irrelevant to our final results. But they allow us to take the difference $\lambda_{i+i_0-1} - \mu_i$ for all $1 \leq i \leq p$. We refer the reader to equations (3.10)-(3.12) of [51] for more details.

A. Interpolating Processes

To estimate the difference $\lambda_i(t) - \mu_i(t)$, we study the following interpolating processes for $0 \leq \alpha \leq 1$ and $1 \leq i \leq p$:

$$dz_i(t,\alpha) = 2\sqrt{z_i(t,\alpha)}\frac{\mathrm{d}B_i}{\sqrt{n}} + \left(\frac{1}{n}\sum_{j\neq i}\frac{z_i(t,\alpha) + z_j(t,\alpha)}{z_i(t,\alpha) - z_j(t,\alpha)} + 1\right)\mathrm{d}t, \quad \text{(C.6)}$$

with the interpolated initial data $z_i(0,\alpha) := \alpha\lambda_i(0) + (1-\alpha)\mu_i(0)$. Correspondingly, we denote the Stieltjes transform of the ESD of $\{z_i(t,\alpha)\}$ by

$$\widetilde{m}_t(z,\alpha) := \frac{1}{p}\sum_{i=1}^{p}\frac{1}{z_i(t,\alpha) - z}. \quad \text{(C.7)}$$

Note that by Lemma 11, due to the choice of γ_w and (C.3), we have that

$$\rho_{\lambda,0}(E_\lambda(0) - E) = \rho_{\mu,0}(E_\mu(0) - E)\left[1 + \mathrm{O}\left(\frac{|E|}{t_0^2}\right)\right], \quad \text{(C.8)}$$

for $0 \leq E \leq \tau t_0^2$, where $\tau > 0$ is a sufficiently small constant. Let $\gamma_{\mu,i}(t)$ and $\gamma_{\lambda,i}(t)$ be the quantiles of $\rho_{\mu,t}$ and $\rho_{\lambda,t}$ defined as

$$\gamma_{\mu,i}(t) := \sup_x\left\{\int_x^{+\infty}\rho_{\mu,t}(x)\mathrm{d}x > \frac{i-1}{p}\right\},$$
$$\gamma_{\lambda,i}(t) := \sup_x\left\{\int_x^{+\infty}\rho_{\lambda,t}(x)\mathrm{d}x > \frac{i-1}{p}\right\}. \quad \text{(C.9)}$$

By Theorem 5 and [7, Theorem 3.2], both $|\widetilde{m}_0(z,0) - m_{\mu,0}(z)|$ and $|\widetilde{m}_0(z,1) - m_{\lambda,0}(z)|$ satisfy local laws as in (B.30) and (B.31). Hence, by Lemma 15, there exists a small enough constant $c_0 > 0$ depending on c_V such that for $1 \leq i \leq k_0$ with $k_0 := \lceil c_0 n\rceil$,

$$\sup_{0\leq t\leq 10t_1}(|z_i(t,0) - \gamma_{\mu,i}(t)| + |z_i(t,1) - \gamma_{\lambda,i}(t)|) \prec i^{-1/3}n^{-2/3}. \quad \text{(C.10)}$$

Here, to get (C.10), we used a standard stochastic continuity argument to pass from fixed times t to all times. Roughly speaking, taking a sequence of fixed times $t_k = 10t_1 \cdot k/n^C$ for a large constant $C > 0$, by Lemma 15 and a simple union bound we get that

$$\sup_{0\leq k\leq n^C}(|z_i(t_k,0) - \gamma_{\mu,i}(t_k)| + |z_i(t_k,1) - \gamma_{\lambda,i}(t_k)|) \prec i^{-1/3}n^{-2/3}. \quad \text{(C.11)}$$

Then, we can show that with high probability, the difference $|z_i(t,0)-z_i(t_k,0)|+|z_i(t,1)-z_i(t_k,1)|$ is small enough for all $t_k \le t \le t_{k+1}$ using a simple continuity estimate. We refer the reader to Appendix B of [50] for more details.

Combining (C.8) and (C.9), we can get the following simple control on the quantiles near the edge.

Lemma 17: For $i = \mathrm{O}(n^{6\omega_0/5})$, we have that

$$|\gamma_{\mu,i}(0)-\gamma_{\lambda,i}(0))| \lesssim \frac{i^{4/3}}{n^{2\omega_0}n^{2/3}}. \tag{C.12}$$

Proof: For simplicity, we denote $x := E_\mu(0)-\gamma_{\mu,i}(0)$ and $y := E_\lambda(0)-\gamma_{\lambda,i}(0)$. Without loss of generality, we assume that $x \le y$. Note that by the square root behaviors of $\rho_{\mu,0}$ and $\rho_{\lambda,0}$ near the right edges, it is easy to get that $x \sim y \sim i^{2/3}n^{-2/3}$ for $i \ge 2$. Now using (C.8) and (C.9), we obtain that

$$\int_0^x \left[\rho_{\mu,0}(E_\mu(0)-E)-\rho_{\lambda,0}(E_\lambda(0)-E)\right]\mathrm{d}E = \int_x^y \rho_{\lambda,0}(E_\lambda(0)-E)\mathrm{d}E,$$

which gives $|y^{3/2}-x^{3/2}| \lesssim x^{5/2}/t_0^2$. From this estimate, we get that $|y-x| \lesssim x^2/t_0^2$, which concludes the proof together with the facts $x \sim i^{2/3}n^{-2/3}$ and $E_\lambda(0) = E_\mu(0)$. □

Next, we will construct a collection of measures that match the asymptotic densities of the interpolating ensembles and have well-behaved square root densities near the right edge. Our main goal is that for each $0 \le \alpha \le 1$, we have a density which matches the distribution of $\{z_i(0,\alpha)\}$ approximately, and with which we can take a rectangular free convolution for any $0 \le t \le t_1$.

At $t = 0$, define the eigenvalue counting functions near the edge $E_\mu(0) = E_\lambda(0)$ as

$$n_\mu(E) = \int_E^{E_\mu(0)} \rho_{\mu,0}(y)\mathrm{d}y, \quad n_\lambda(E) = \int_E^{E_\lambda(0)} \rho_{\lambda,0}(y)\mathrm{d}y.$$

Since $\rho_{\mu,0}(y) > 0$ for $E_\mu(0)-\tau \le y < E_\mu(0)$ and $\rho_{\lambda,0}(y) > 0$ for $E_\lambda(0)-\tau \le y < E_\lambda(0)$ for a small enough constant $\tau > 0$, the functions n_μ and n_λ are strictly increasing near the right edges. Hence, we can define the inverse functions (i.e. the continuous versions of quantiles) $\varphi_\mu(s)$ and $\varphi_\lambda(s)$ by the equations

$$n_\mu(\varphi_\mu(s)) = s, \quad n_\lambda(\varphi_\lambda(s)) = s, \quad 0 \le s \le c_*,$$

where $c_* \equiv c_*(n) := \lceil c_0 n\rceil/n$ for a small enough constant $c_0 > 0$. Then, for $\alpha \in [0,1]$, we define

$$\varphi(s,\alpha) := \alpha\varphi_\mu(s)+(1-\alpha)\varphi_\lambda(s),$$

which maps $[0,c_*]$ onto

$$\mathtt{D}_\alpha := [\alpha\varphi_\mu(c_*)+(1-\alpha)\varphi_\lambda(c_*), E_\lambda(0)]. \tag{C.13}$$

Now, for any $\alpha \in [0,1]$, we define the inverse function $n(E,\alpha) : \mathtt{D}_\alpha \to [0,c_*]$ of $\varphi(s,\alpha)$ by the equation

$$n(\varphi(s,\alpha),\alpha) = s.$$

Then, we define the asymptotic density as

$$\rho(E,\alpha) := \frac{\partial n(E,\alpha)}{\partial E}.$$

By inverse function theorem, we can calculate that

$$\rho(E,\alpha) = \left[\frac{\alpha}{\rho_{\mu,0}(\varphi_\mu(n(E,\alpha)))}+\frac{1-\alpha}{\rho_{\lambda,0}(\varphi_\lambda(n(E,\alpha)))}\right]^{-1}.$$

Combining it with (C.8), we immediately find that for $0 \le E \le \tau t_0^2$

$$\rho(E_+(0,\alpha)-E,\alpha) = \rho_{\mu,0}(E_\mu(0)-E)\left[1+\mathrm{O}\left(\frac{|E|}{t_0^2}\right)\right], \tag{C.14}$$

for a sufficiently small constant $\tau > 0$, where $E_+(0,\alpha)$ is the right edge of $\rho(E,\alpha)$. We now construct a (random) measure $\mu(E,\alpha)$ as

$$\mathrm{d}\mu(E,\alpha) = \rho(E,\alpha)\mathbf{1}_{\{E\in\mathtt{D}_\alpha\}}\mathrm{d}E + p^{-1}\sum_{i>c_*n}\delta_{z_i(0,\alpha)}(\mathrm{d}E).$$

This measure is defined in a way such that its Stieltjes transform is close to $\widetilde{m}_0(z,\alpha)$ in (C.7). Moreover, the motivation behind this definition is as follows. We need a deterministic density that behaves well around the right edge in order to use the results in Section B. But we do not have any estimate on the density far away from the edge. Hence for the remaining eigenvalues that are away from the right edge by a distance of order 1, we just take δ functions. Although the sum of delta measures is random, its effect on deterministic quantities that we are interested in is negligible.

Let $\rho_t(E,\alpha)$ be the rectangular free convolution of $\mathrm{d}\mu(E,\alpha)$ with the MP law at time t. Moreover, we denote its Stieltjes transform by $m_t(z,\alpha)$ and its right edge by $E_+(t,\alpha)$. Some key properties of $\rho_t(E,\alpha)$ and $m_t(z,\alpha)$ have been given in Section B. In particular, we know that $\rho_t(E,\alpha)$ has a square root behavior near $E_+(t,\alpha)$. Although $\rho_t(E,\alpha)$ is random, with the results in Section B we can provide a deterministic control on it.

Lemma 18: Let $\varepsilon,\tau > 0$ be sufficiently small constants. For $0 \le E \le \tau n^{-2\epsilon}t_0^2$, we have that for any constant $D > 0$,

$$\rho_t(E_+(t,\alpha)-E,\alpha) = \rho_{\mu,t}(E_\mu(t)-E)\left[1+\mathrm{O}(n^\epsilon t/t_0+n^{-D})\right]. \tag{C.15}$$

Moreover, for a small constant $c_\tau > 0$, we have that

$$\max_{1\le i\le c_\tau n^{1-3\varepsilon}t_0^3}|\widetilde{\gamma}_i(t,\alpha)-\widetilde{\gamma}_i(t,0)| \le \left(n^\epsilon\frac{t}{t_0}+n^{-D}\right)\frac{i^{2/3}}{n^{2/3}}, \tag{C.16}$$

where we introduced the short-hand notation $\widetilde{\gamma}_i(t,\alpha) := E_+(t,\alpha)-\gamma_i(t,\alpha)$.

Proof: The estimates (C.15) follows directly from (B.24). The estimate (C.16) follows from (C.15) using the same argument as in the proof of (C.12). □

With the eigenvalues rigidity (C.10) and the construction of $\mathrm{d}\mu(E,\alpha)$, we can verify that $|m_0(z,\alpha)-\widetilde{m}_0(z,\alpha)|$ satisfies Assumption 3. Then, by Lemma 16, we have the following rigidity estimate of $\{z_i(t,\alpha)\}$. As before, we define the quantiles $\gamma_i(t,\alpha)$ by

$$\gamma_i(t,\alpha) := \sup_x\left\{\int_x^{+\infty}\rho_t(E,\alpha)\mathrm{d}E > \frac{i-1}{p}\right\}.$$

Lemma 19: There exists a constant $c_* > 0$ so that for $1 \le i \le c_* n$

$$\sup_{0\le\alpha\le 1}\sup_{0\le t\le 10t_1}|z_i(t,\alpha)-\gamma_i(t,\alpha)| \prec i^{-1/3}n^{-2/3}. \quad \text{(C.17)}$$

Proof: This estimate follows from Lemma 16 combined with a standard stochastic continuity argument in t. □

Using (B.9) and (B.14), we can calculate that

$$\begin{aligned}
&\frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t}\\
&=\left[\frac{1-c_n}{2\zeta_t(E_+(t,\alpha),\alpha)}-\frac{c_n m_t(E_+(t,\alpha),\alpha)}{b_t(E_+(t,\alpha),\alpha)}\right]\\
&\quad\times\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}\\
&\quad-\frac{(1-c_n)^2 t}{4\zeta_t(E_+(t,\alpha),\alpha)\sqrt{E_+(t,\alpha)}}+\mathrm{O}(t^2)\\
&=\frac{1-c_n}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}}\\
&\quad-c_n m_t(E_+(t,\alpha),\alpha)\sqrt{E_+(t,\alpha)-t(1-c_n)}\\
&\quad-\frac{(1-c_n)^2 t}{4[E_+(t,\alpha)]^{3/2}}+\mathrm{O}(t^2), \quad \text{(C.18)}
\end{aligned}$$

where we used the notations

$$\begin{aligned}
&b_t(z,\alpha):=1+c_n t m_t(z,\alpha),\\
&\zeta_t(z,\alpha):=z b_t^2(z,\alpha)-t(1-c_n)b_t(z,\alpha).
\end{aligned}$$

In the proof, we will also need to use the following function defined for $E\in[-\tau,\tau]$ for a small enough constant $\tau>0$:

$$\begin{aligned}
&\Psi_t(E,\alpha)\\
&:=\frac{1-c_n}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}}-\frac{1-c_n}{2\sqrt{E_+(t,\alpha)-E}}\\
&\quad-\frac{(1-c_n)^2 t}{4[E_+(t,\alpha)]^{3/2}}\\
&\quad-Re\Big[c_n m_t(E_+(t,\alpha),\alpha)\sqrt{E_+(t,\alpha)-t(1-c_n)}\\
&\quad-c_n m_t(E_+(t,\alpha)-E,\alpha)\sqrt{E_+(t,\alpha)-E}\Big]. \quad \text{(C.19)}
\end{aligned}$$

Next, we prove some matching estimates for the function $\Psi_t(E,\alpha)$ in Lemma 20. The proof of this lemma explores a rather delicate cancellation in $\Psi_t(E,\alpha)$.

Lemma 20: Let $\varepsilon,\tau>0$ be sufficiently small constants. For $0\le E\le \tau n^{-2\epsilon}t_0^2$, we have that for any constant $D>0$,

$$|\Psi_t(E,\alpha)-\Psi_t(E,0)|\lesssim\left(\frac{n^\epsilon}{t_0}+\frac{n^{-D}}{t}\right)E+t^2. \quad \text{(C.20)}$$

For $0\le E\le\tau n^{-2\epsilon}tt_0$, we have that for any constant $D>0$,

$$\begin{aligned}
&|\Psi_t(-E,\alpha)-\Psi_t(-E,0)|\\
&\lesssim\left(n^\epsilon\frac{t^{1/2}}{t_0^{1/2}}+n^{-D}\frac{t_0^{1/2}}{t^{1/2}}\right)E^{1/2}+t^2. \quad \text{(C.21)}
\end{aligned}$$

Proof: First, we claim that

$$\Psi_t(E,\alpha)=\widetilde{\Psi}_t(E,\alpha)+\mathrm{O}(t^2), \quad \text{(C.22)}$$

where $\widetilde{\Psi}_t(E,\alpha)$ is defined by

$$\begin{aligned}
\widetilde{\Psi}_t(E,\alpha):=&\frac{1-c_n}{2\sqrt{E_+(t,\alpha)}}-\frac{1-c_n}{2\sqrt{E_+(t,\alpha)-E}}\\
&-Re\Big[c_n m_t(E_+(t,\alpha),\alpha)\sqrt{E_+(t,\alpha)}\\
&-c_n m_t(E_+(t,\alpha)-E,\alpha)\sqrt{E_+(t,\alpha)-E}\Big].
\end{aligned}$$

In fact, subtracting $\widetilde{\Psi}_t(x,\alpha)$ from $\Psi_t(x,\alpha)$ and using the definition of $\zeta_t(E_+(t,\alpha),\alpha)$, we get that $\Psi_t(E,\alpha)-\widetilde{\Psi}_t(E,\alpha)$, shown at the bottom of the page. On the other hand, using (B.23) and the fact that $E_+(0,\alpha)=E_\lambda(0)$ for $0\le\alpha\le1$, we get that

$$|E_+(t,\alpha)-E_\lambda(0)|=\mathrm{O}(t),\quad 0\le\alpha\le1. \quad \text{(C.23)}$$

$$\begin{aligned}
&\Psi_t(E,\alpha)-\widetilde{\Psi}_t(E,\alpha)=\frac{1-c_n}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}}-\frac{1-c_n}{2\sqrt{E_+(t,\alpha)}}\\
&+\frac{(1-c_n)c_n t m_t(E_+(t,\alpha),\alpha)}{\sqrt{E_+(t,\alpha)}+\sqrt{E_+(t,\alpha)-t(1-c_n)}}-\frac{(1-c_n)^2 t}{4[E_+(t,\alpha)]^{3/2}}\\
&=\frac{(1-c_n)\left[E_+(t,\alpha)-b_t^2(E_+(t,\alpha),\alpha)\cdot E_+(t,\alpha)\right]}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}\sqrt{E_+(t,\alpha)}\left(\sqrt{E_+(t,\alpha)}+\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}\right)}\\
&+\frac{(1-c_n)^2 t b_t(E_+(t,\alpha),\alpha)}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}\sqrt{E_+(t,\alpha)}\left(\sqrt{E_+(t,\alpha)}+\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}\right)}\\
&+\frac{(1-c_n)c_n t m_t(E_+(t,\alpha),\alpha)}{2\sqrt{E_+(t,\alpha)}}-\frac{(1-c_n)^2 t}{4[E_+(t,\alpha)]^{3/2}}+\mathrm{O}(t^2)\\
&=\frac{-2(1-c_n)c_n t m_t(E_+(t,\alpha),\alpha)\cdot E_+(t,\alpha)+(1-c_n)^2 t}{4[E_+(t,\alpha)]^{3/2}}\\
&+\frac{(1-c_n)c_n t m_t(E_+(t,\alpha),\alpha)}{2\sqrt{E_+(t,\alpha)}}-\frac{(1-c_n)^2 t}{4[E_+(t,\alpha)]^{3/2}}+\mathrm{O}(t^2)\\
&=\mathrm{O}(t^2)
\end{aligned}$$

Hence, we can estimate that

$$\begin{aligned}
&\widetilde{\Psi}_t(E,\alpha)-\widetilde{\Psi}_t(E,0)\\
&=-c_n Re\left[m_t(E_+(t,\alpha),\alpha)-m_t(E_+(t,\alpha)-E,\alpha)\right]\\
&\quad\times\sqrt{E_+(t,\alpha)}\\
&\quad+c_n Re\left[m_t(E_+(t,0),0)-m_t(E_+(t,0)-E,0)\right]\\
&\quad\times\sqrt{E_+(t,0)}+\mathrm{O}(E)\\
&=c_n Re\Big[\left(m_t(E_+(t,0),0)-m_t(E_+(t,0)-E,0)\right)\\
&\quad-\left(m_t(E_+(t,\alpha),\alpha)-m_t(E_+(t,\alpha)-E,\alpha)\right)\Big]\sqrt{E_+(t,\alpha)}\\
&\quad+\mathrm{O}\left(t\left|m_t(E_+(t,0),0)-m_t(E_+(t,0)-E,0)\right|+E\right).
\end{aligned}$$

By (B.18), we have that for $E\geq 0$,

$$t\left|m_t(E_+(t,0),0)-m_t(E_+(t,0)-E,0)\right|\lesssim (Et^{-1})\cdot t=E,$$

and by (B.16) and (B.17), we have that for $E\leq 0$,

$$t\left|m_t(E_+(t,0),0)-m_t(E_+(t,0)-E,0)\right|\lesssim t\sqrt{|E|}\leq t^2+|E|.$$

Using these two estimates and Lemma 14, we can bound $\widetilde{\Psi}_t(E,\alpha)-\widetilde{\Psi}_t(E,0)$ and conclude (C.20) and (C.21). □

Remark 11: Later we will only consider the dynamics after $t=n^{-C}$ for some large constant $C>0$, so that the n^{-D} terms in (C.20) and (C.21) are negligible as long as D is large enough.

Note that the interpolating measures $\mathrm{d}\mu(E,0)$ (resp. $\mathrm{d}\mu(E,1)$) only matches the asymptotic measure $\rho_{\mu,0}(E)\mathrm{d}E$ (resp. $\rho_{\lambda,0}(E)\mathrm{d}E$) for $E\in\mathtt{D}_0$ (resp. $E\in\mathtt{D}_1$). For the random part, we control its effect using the local laws. With the eigenvalues rigidity (C.10), we can check that $|m_0(z,0)-\widetilde{m}_0(z,0)|$ and $|m_0(z,1)-\widetilde{m}_1(z,1)|$ satisfy the two estimates in Assumption 3. (Recall that $\widetilde{m}_0(z,\alpha)$ was defined in (C.7) and $m_0(z,\alpha)$ is the Stieltjes transform of $\mathrm{d}\mu(E,\alpha)$.) Moreover, by Theorem 5 and [7, Theorem 3.2], we also have that $|\widetilde{m}_0(z,0)-m_{\mu,0}(z)|$ and $|\widetilde{m}_0(z,1)-m_{\lambda,0}(z)|$ satisfy the two estimates in Assumption 3. Hence, we have that

$$|m_0(z,0)-m_{\mu,0}(z)|\prec\frac{1}{n\eta},\tag{C.24}$$

for $|E-E_\mu(0)|\leq 3c_V/4$ and $n^{-2/3+\vartheta}\leq\eta\leq 10$;

$$|m_0(z,1)-m_{\lambda,0}(z)|\prec\frac{1}{n\eta},\tag{C.25}$$

for $|E-E_\lambda(0)|\leq 3c_V/4$ and $n^{-2/3+\vartheta}\leq\eta\leq 10$. With the above two estimates, we can control $|E_+(t,1)-E_\lambda(t)|$ and $|E_+(t,0)-E_\mu(t)|$ for $0\leq t\leq 10t_1$.

Lemma 21: We have that

$$\max_{0\leq t\leq 10t_1}|E_+(t,1)-E_\lambda(t)|\prec t^3+n^{-1/2}t,\tag{C.26}$$

and

$$\max_{0\leq t\leq 10t_1}|E_+(t,0)-E_\mu(t)|\prec t^3+n^{-1/2}t.\tag{C.27}$$

Proof: Repeating the proof of Lemma 13 (as given in Lemma A.2 of [19]) with t_0 replaced by 1, we can get that

$$|\zeta_{+,1}-\zeta_{+,\lambda}|\leq Ct^3,$$

where we abbreviate $\zeta_{+,1}\equiv\zeta_t(E_+(t,1),1)$ and $\zeta_{+,\lambda}\equiv\zeta_{\lambda,t}(E_\lambda(t))$. Then with equation (B.3), we get that

$$\begin{aligned}
&|E_+(t,1)-E_\lambda(t)|\\
&\lesssim|\zeta_{+,1}-\zeta_{+,\lambda}|+t|m_0(\zeta_{+,1},1)-m_{\lambda,0}(\zeta_{+,\lambda})|.
\end{aligned}\tag{C.28}$$

Recall that $\zeta_{+,1}-E_\lambda(0)\sim t^2$ and $\zeta_{+,\lambda}-E_\lambda(0)\sim t^2$ by (B.11). Then using (B.16) and (B.17), we get that $|m'_{\lambda,0}(\zeta)|\lesssim t^{-1}$ for ζ between $\zeta_{+,1}$ and $\zeta_{+,\lambda}$. Thus, we can bound (C.28) as

$$\begin{aligned}
&|E_+(t,1)-E_\lambda(t)|\\
&\lesssim t^3+t|m_0(\zeta_{+,1},1)-m_{\lambda,0}(\zeta_{+,1})|\\
&\quad+t|m_{\lambda,0}(\zeta_{+,1})-m_{\lambda,0}(\zeta_{+,\lambda})|\\
&\lesssim t^3+t|m_0(\zeta_{+,1},1)-m_{\lambda,0}(\zeta_{+,1})|.
\end{aligned}\tag{C.29}$$

For the second part, since $\mathrm{d}\mu(E,1)$ matches $\rho_{\lambda,0}(E)$ for $E\in\mathtt{D}_1$, we can bound that

$$\begin{aligned}
&\Big|\left[m_0(\zeta_{+,1},1)-m_{\lambda,0}(\zeta_{+,1})\right]\\
&-\left[m_0(\zeta_{+,1}+\mathrm{i}n^{-1/2},1)-m_{\lambda,0}(\zeta_{+,1}+\mathrm{i}n^{-1/2})\right]\Big|\\
&\leq\sum_{i>c_*n}\frac{n^{-1/2}}{|z_i(0,1)-\zeta_{+,1}||z_i(0,1)-\zeta_{+,1}-\mathrm{i}n^{-1/2}|}\lesssim n^{-1/2}
\end{aligned}$$

with high probability. On the other hand, we have that

$$\left|m_0(\zeta_{+,1}+\mathrm{i}n^{-1/2},1)-m_{\lambda,0}(\zeta_{+,1}+\mathrm{i}n^{-1/2})\right|\prec n^{-1/2}$$

by (C.25). Combining the above two estimates, we obtain that

$$|m_0(\zeta_{+,1},1)-m_{\lambda,0}(\zeta_{+,1})|\prec n^{-1/2}.$$

Plugging it into (C.29), we conclude (C.26). The estimate (C.27) can be proved in the same way. □

In later proof, we will also need to study the evolution of the singular values $y_i(t,\alpha):=\sqrt{z_i(t,\alpha)}$. It is easy to see that the asymptotic density for $y_i(t,\alpha)$ is given by

$$f_t(E,\alpha):=2E\rho_t(E^2,\alpha),\quad 0\leq\alpha\leq 1.$$

Similarly we can define $f_{\lambda,t}$ and $f_{\mu,t}$. Moreover, the quantiles of $f_t(E,\alpha)$ are exactly given by $\sqrt{\gamma_i(t,\alpha)}$. Now with Lemma 18 and Lemma 19, we can easily conclude the following lemma.

Lemma 22: We have the following rigidity estimate of singular values:

$$\sup_{0\leq\alpha\leq 1}\sup_{0\leq t\leq 10t_1}\left|y_i(t,\alpha)-\sqrt{\gamma_i(t,\alpha)}\right|\prec\frac{1}{i^{1/3}n^{2/3}},\tag{C.30}$$

for $1\leq i\leq c_*n$. Let $\varepsilon,\tau>0$ be sufficiently small constants. For $0\leq E\leq\tau n^{-2\epsilon}t_0^2$, we have that for any constant $D>0$,

$$\begin{aligned}
&f_t\left(\sqrt{E_+(t,\alpha)}-E,\alpha\right)\\
&=f_{\mu,t}\left(\sqrt{E_\mu(t)}-E\right)\left[1+\mathrm{O}(n^\epsilon t/t_0+n^{-D})\right],
\end{aligned}\tag{C.31}$$

and

$$\begin{aligned}
&\max_{1\leq i\leq c_\tau n^{1-3\varepsilon}t_0^3}|\widehat{\gamma}_i(t,\alpha)-\widehat{\gamma}_i(t,0)|\\
&\leq\left(n^\epsilon\frac{t}{t_0}+n^{-D}\right)\frac{i^{2/3}}{n^{2/3}},
\end{aligned}\tag{C.32}$$

where we introduced the short-hand notation $\widehat{\gamma}_i(t,\alpha) := \sqrt{E_+(t,\alpha)} - \sqrt{\gamma_i(t,\alpha)}$.

Proof: The rigidity result (C.30) follows directly from Lemma 19, (C.31) follows from (C.15) together with (C.23), and (C.32) can be proved easily using (C.31). □

B. Short-Range Approximation

As in [51], we will build a short-range approximation for the interpolating processes $\{z_i(t,\alpha)\}$, which is based on the simple intuition that the eigenvalues that are far away from the edge have negligible effect on the edge eigenvalues. It turns out that it is more convenient to study the SDEs for singular values $y_i(t,\alpha)$. By Ito's formula, we get that for $1 \le i \le p$,

$$\begin{aligned}
&\mathrm{d}y_i(t,\alpha)\\
&= \frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2y_i(t,\alpha)}\left(\frac{1}{n}\sum_{j\neq i}\frac{y_i^2(t,\alpha)+y_j^2(t,\alpha)}{y_i^2(t,\alpha)-y_j^2(t,\alpha)} + \frac{n-1}{n}\right)\mathrm{d}t\\
&= \frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n}\sum_{j\neq i}\left[\frac{\mathrm{d}t}{y_i(t,\alpha)-y_j(t,\alpha)} + \frac{\mathrm{d}t}{y_i(t,\alpha)+y_j(t,\alpha)}\right]\\
&\quad + \frac{n-p}{2ny_i(t,\alpha)}\mathrm{d}t. \qquad \text{(C.33)}
\end{aligned}$$

Note that the diffusion term now has a constant coefficient. For convenience, we introduce the shifted processes

$$\begin{aligned}
\widetilde{z}_i(t,\alpha) &:= E_+(t,\alpha) - z_i(t,\alpha),\\
\widetilde{y}_i(t,\alpha) &:= \sqrt{E_+(t,\alpha)} - y_i(t,\alpha). \qquad \text{(C.34)}
\end{aligned}$$

Clearly, we have that $\widetilde{z}_i(t,\alpha) \sim \widetilde{y}_i(t,\alpha)$. We see that $\widetilde{y}_i(t,\alpha)$ obeys the SDE

$$\begin{aligned}
\mathrm{d}\widetilde{y}_i(t,\alpha) &= -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n}\sum_{j\neq i}\frac{\mathrm{d}t}{\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}\\
&\quad - \frac{1}{2n}\sum_{j\neq i}\frac{\mathrm{d}t}{2\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}\\
&\quad + \frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t}\mathrm{d}t - \frac{n-p}{2n[\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)]}\mathrm{d}t, \qquad \text{(C.35)}
\end{aligned}$$

where $\partial_t\sqrt{E_+(t,\alpha)}$ is given by (B.9).

We now define a "short-range" set of indices $\mathcal{A} \subset [1,p]\times[1,p]$. Let $\mathcal{A}$ be a symmetric set of indices in the sense that $(i,j)\in\mathcal{A}$ if and only if $(j,i)\in\mathcal{A}$, and choose a parameter $\ell := n^{\omega_\ell}$, where $\omega_\ell > 0$ is a constant that will be specified later. Then, we define

$$\begin{aligned}
\mathcal{A} := &\Big\{(i,j) : |i-j| \le \ell(10\ell^2 + i^{2/3} + j^{2/3})\Big\}\\
&\bigcup \{(i,j) : i,j > i_*/2\} \qquad \text{(C.36)}
\end{aligned}$$

for $i_* := c_* n$, where c_* is the constant as appeared in Lemma 19. It is easy to check that for each i, the set $\{j : (i,j)\in\mathcal{A}\}$ consists of consecutive integers. For convenience, we introduce the following short-hand notations

$$\sum_j^{\mathcal{A},(i)} := \sum_{j:(i,j)\in\mathcal{A}, j\neq i}, \quad \sum_j^{\mathcal{A}^c,(i)} := \sum_{j:(i,j)\notin\mathcal{A}, j\neq i}.$$

For each i, we denote $[i_-, i_+] := \{j : (i,j)\in\mathcal{A}\}$ and

$$\begin{aligned}
\mathcal{I}_i(t,\alpha) &:= [\widetilde{\gamma}_{i-}(t,\alpha), \widetilde{\gamma}_{i+}(t,\alpha)],\\
\widehat{\mathcal{I}}_i(t,\alpha) &:= [\widehat{\gamma}_{i-}(t,\alpha), \widehat{\gamma}_{i+}(t,\alpha)],
\end{aligned}$$

where we recall that $\widetilde{\gamma}_i(t,\alpha)$ and $\widehat{\gamma}_i(t,\alpha)$ are defined below (C.16) and (C.32), respectively. Finally, we denote

$$\mathcal{J}(t,\alpha) := [-\widetilde{c}_V, \widehat{\gamma}_{3i_*/4}(t,\alpha)],$$

where $\widetilde{c}_V > 0$ is a small constant depending only on c_V.

Let $\omega_a > 0$ be a constant that will be specified later. The short-range approximation to $\widetilde{y}$ is a process $\widehat{y}$ defined as the solution to the following SDEs for $t \ge n^{-C_0}$ with the same initial data (recall Remark 11)

$$\widehat{y}_i(t = n^{-C_0},\alpha) = \widetilde{y}_i(t = n^{-C_0},\alpha), \quad 0 \le \alpha \le 1,$$

where C_0 is an absolute constant (for example, $C_0 = 100$ will be more than enough). For $1 \le i \le n^{\omega_a}$, the SDEs are

$$\begin{aligned}
\mathrm{d}\widehat{y}_i(t,\alpha) &= -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n}\sum_j^{\mathcal{A},(i)}\frac{1}{\widehat{y}_i(t,\alpha)-\widehat{y}_j(t,\alpha)}\mathrm{d}t\\
&\quad - \left[c_n\int_{\mathcal{I}_i^c(t,0)}\frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2}\mathrm{d}E\right]\mathrm{d}t\\
&\quad - \frac{n-p}{2n\sqrt{E_+(t,0)}}\mathrm{d}t + \frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t}\mathrm{d}t; \qquad \text{(C.37)}
\end{aligned}$$

for $n^{\omega_a} < i \le i_*/2$, the SDEs are

$$\begin{aligned}
\mathrm{d}\widehat{y}_i(t,\alpha) &= -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n}\sum_j^{\mathcal{A},(i)}\frac{\mathrm{d}t}{\widehat{y}_i(t,\alpha)-\widehat{y}_j(t,\alpha)}\\
&\quad + \frac{1}{2n}\sum_{j\ge 3i_*/4}\frac{\mathrm{d}t}{\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}\\
&\quad + \frac{c_n}{2}\int_{\widehat{\mathcal{I}}_i^c(t,\alpha)\cap\mathcal{J}(t,\alpha)}\frac{f_t(\sqrt{E_+(t,\alpha)}-E,\alpha)}{\widehat{y}_i(t,\alpha)-E}\mathrm{d}E\mathrm{d}t\\
&\quad - \frac{n-p}{2n[\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)]}\mathrm{d}t + \frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t}\mathrm{d}t\\
&\quad - \frac{1}{2n}\sum_{j\neq i}\frac{\mathrm{d}t}{2\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}; \qquad \text{(C.38)}
\end{aligned}$$

for $i_*/2 < i \le p$, the SDEs are

$$\begin{aligned}
\mathrm{d}\widehat{y}_i(t,\alpha) &= -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n}\sum_j^{\mathcal{A},(i)}\frac{\mathrm{d}t}{\widehat{y}_i(t,\alpha)-\widehat{y}_j(t,\alpha)}\\
&\quad + \frac{1}{2n}\sum_j^{\mathcal{A}^c,(i)}\frac{\mathrm{d}t}{\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}\\
&\quad - \frac{n-p}{2n[\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)]}\mathrm{d}t + \frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t}\mathrm{d}t\\
&\quad - \frac{1}{2n}\sum_{j\neq i}\frac{\mathrm{d}t}{2\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}. \qquad \text{(C.39)}
\end{aligned}$$

Corresponding to (C.34), we denote

$$\widehat{z}_i(t,\alpha) := E_+(t,\alpha)[\sqrt{E_+(t,\alpha)}-\widehat{y}_i(t,\alpha)]^2. \qquad \text{(C.40)}$$

We now choose the hierarchy of the scale parameters in the following quantities:

$$t_0 = n^{-1/3+\omega_0}, \quad t_1 = n^{-1/3+\omega_1}, \quad \ell = n^{\omega_\ell}, \quad \text{and} \quad n^{\omega_a}.$$

In fact, we will choose the constants ω_0, ω_1, ω_ℓ and ω_a such that

$$0 < \omega_1 \le \mathfrak{C}^{-1}\omega_\ell \le \mathfrak{C}^{-2}\omega_a \le \mathfrak{C}^{-3}\omega_0 \le \mathfrak{C}^{-1} \tag{C.41}$$

for some constant $\mathfrak{C} > 0$ that is as large as needed. Here the purpose of the scale ℓ is to cut off the effect of the initial data far away from the right edge, since $\widetilde{y}_i(0, \alpha = 1)$ and $\widetilde{y}_i(0, \alpha = 0)$ only match for small i. Moreover, by choosing scale $\omega_a \ll \omega_0$, we can make use of the matching estimates in Lemma 20 to show that the drifting terms in the SDEs with $1 \le i \le n^{\omega_a}$ are approximately α independent.

Next, we show that $\widetilde{y}_i(t, \alpha)$ are good approximations to $\widehat{y}_i(t, \alpha)$. Before that, we recall the semigroup approach for first order parabolic PDE. Let Ω be a real Banach space with a given norm and $\mathcal{L}(\Omega)$ be the Banach algebra of all linear continuous mappings. We say a family of operators $\{T(t) : t \ge 0\}$ in $\mathcal{L}(\Omega)$ is a semigroup if

$$T(0) = \mathrm{id}, \quad \text{and} \quad T(t+s) = T(t)T(s) \quad \text{for all } t, s \ge 0.$$

For a detailed discussion of semigroups of operators, we refer the readers to [8].

Definition 5: For any operator $\mathcal{W} \in \mathcal{L}(\mathbb{R}^p)$, we denote $\mathcal{U}^{\mathcal{W}}$ as the semigroup associated with $\mathcal{W}$, i.e., $\mathcal{W}$ is the infinitesimal generator of $\mathcal{U}^{\mathcal{W}}$. Moreover, we denote $\mathcal{U}^{\mathcal{W}}(s, t)$ as the semigroup from s to t, that is, $\mathcal{U}^{\mathcal{W}}(s, s) = \mathrm{id}$ and

$$\partial_t \mathcal{U}^{\mathcal{W}}(s, t) = \mathcal{W}(t)\mathcal{U}^{\mathcal{W}}(s, t), \quad \text{for any } t \ge s.$$

For the rest of this subsection, we prove the following short-range approximation estimate.

Lemma 23: With high probability, we have that for any constant $\epsilon > 0$,

$$\sup_{0\le\alpha\le 1} \sup_{n^{-C_0}\le t\le 10t_1} \max_{1\le i\le p} |\widetilde{y}_i(t, \alpha) - \widehat{y}_i(t, \alpha)| \le n^{-2/3+\varepsilon+\omega_1-2\omega_\ell}. \tag{C.42}$$

Proof: We abbreviate $v_i := \widetilde{y}_i - \widehat{y}_i$. Subtracting the SDEs for $\widetilde{y}_i$ and $\widehat{y}_i$, we obtain the following inhomogeneous PDE for v:

$$\partial_t v = (\mathcal{B}_1 + \mathcal{V}_1)v + \zeta,$$

where $\mathcal{B}_1$ is a linear operator defined by

$$(\mathcal{B}_1 v)_i = -\frac{1}{2n} \sum_j^{\mathcal{A},(i)} \frac{v_i - v_j}{(\widetilde{y}_i - \widetilde{y}_j)(\widehat{y}_i - \widehat{y}_j)},$$

and $\mathcal{V}_1$ is a diagonal operator defined as follows: $\mathcal{V}_1(i) = 0$ for $i > i_*/2$; for $1 \le i \le n^{\omega_a}$,

$$\begin{aligned}
&\mathcal{V}_1(i)v_i := c_n\sqrt{E_+(t,0)} \\
&\times \int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2}\mathrm{d}E \\
&- c_n\sqrt{E_+(t,0)} \\
&\times \int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widetilde{y}_i(t,\alpha))^2}\mathrm{d}E \\
&= -v_i \int_{\mathcal{I}_i^c(t,0)} \frac{c_n\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{[E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2]} \\
&\times \frac{2\sqrt{E_+(t,0)} - \widetilde{y}_i(t,\alpha) - \widehat{y}_i(t,\alpha)}{[E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widetilde{y}_i(t,\alpha))^2]}\mathrm{d}E;
\end{aligned}$$

for $n^{\omega_a} < i \le i_*/2$,

$$\mathcal{V}_1(i) = -\frac{c_n}{2} \int_{\widehat{\mathcal{I}}_i^c(t,\alpha)\cap\mathcal{J}(t,\alpha)} \frac{f_t(\sqrt{E_+(t,\alpha)}-E,\alpha)}{(\widetilde{y}_i(t,\alpha)-E)(\widehat{y}_i(t,\alpha)-E)}\mathrm{d}E.$$

The term ζ contains the remaining errors, and we will control its ℓ^∞ norm later.

For the following proof, we assume a rough bound on $\widehat{y}_i(t, \alpha)$:

$$\sup_{0\le\alpha\le 1} \max_{1\le i\le i_*/2} |\widetilde{y}_i(t, \alpha) - \widehat{y}_i(t, \alpha)| \le n^{-2/3}, \tag{C.43}$$

for $n^{-C_0} \le t \le 10t_1$. Later, we will remove it with a simple continuity argument. Since $\mathcal{V}_1(i) \le 0$, the operator $\mathcal{V}_1$ is negative. Then, the semigroup of $\mathcal{B}_1 + \mathcal{V}_1$ is a contraction on every $\ell^q(\{1, \cdots, p\})$ space. To see this, for $u(s) = \mathcal{U}^{\mathcal{B}_1+\mathcal{V}_1}(0, s)u_0$ and $q \ge 1$, we have that

$$\begin{aligned}
&\partial_t \sum_i |u_i(s)|^q \\
&= \sum_i |u_i(s)|^{q-1} \operatorname{sign}(u_i(s)) \left[(\mathcal{B}_1 u(s))_i + \mathcal{V}_1(i)u_i(s)\right] \\
&\le \sum_i |u_i(s)|^{q-1} \operatorname{sign}(u_i(s)) \left(\mathcal{B}_1 u(s)\right)_i \\
&= -\frac{1}{4n} \sum_{i,j\in\mathcal{A}} \frac{|u_i(s)|^{q-1}\operatorname{sign}(u_i(s)) - |u_j(s)|^{q-1}\operatorname{sign}(u_j(s))}{\widetilde{z}_i - \widetilde{z}_j} \\
&\quad \times \frac{u_i(s) - u_j(s)}{\widehat{z}_i - \widehat{z}_j} \\
&\le 0.
\end{aligned}$$

On the space $l^\infty(\{1, \cdots, p\})$, we just need to use $\|u\|_\infty = \lim_{q\to\infty} \|u\|_q$. By Duhamel's principle, we have

$$v(t) = \int_{n^{-C_0}}^t \mathcal{U}^{\mathcal{B}_1+\mathcal{V}_1}(s, t)\zeta(s)\mathrm{d}s,$$

which gives that

$$\|v(t)\|_\infty \le \int_{n^{-C_0}}^t \|\zeta(s)\|_\infty \mathrm{d}s. \tag{C.44}$$

Next, we provide the bounds on $\|\zeta(s)\|_\infty$. Fist, we have that $\zeta_i(t) = 0$ for $i \ge i_*/2$. Second, under (C.43), for

$n^{\omega_a} < i \le i_*/2$, we have that

$$\zeta_i(t) = \frac{1}{2n}\sum_{j\le 3i_*/4}^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)} - \frac{c_n}{2}\int_{\widehat{\mathcal{I}}_i^c(t,\alpha)\cap\mathcal{J}(t,\alpha)} \frac{f_t(\sqrt{E_+(t,\alpha)}-E,\alpha)}{\widetilde{y}_i(t,\alpha)-E}\mathrm{d}E. \quad \text{(C.45)}$$

Decomposing the integral in (C.45) according to the quantiles of f_t as $\sum_j \int_{\sqrt{\gamma_{j+1}}}^{\sqrt{\gamma_j}}$ and using (C.30), we obtain that for any constant $\varepsilon > 0$,

$$|\zeta_i| \le \frac{n^{\varepsilon}}{n^{5/3}}\sum_{j\le 3i_*/4}^{\mathcal{A}^c,(i)} \frac{1}{(\widehat{\gamma}_i-\widehat{\gamma}_j)^2 j^{1/3}} \le \frac{Cn^{\epsilon}}{n^{1/3}}\sum_{j\le 3i_*/4}^{\mathcal{A}^c,(i)} \frac{i^{2/3}+j^{2/3}}{(i-j)^2 j^{1/3}} \quad \text{(C.46)}$$

with high probability, where for the second inequality we used

$$|\widehat{\gamma}_i-\widehat{\gamma}_j| \sim |i^{2/3}-j^{2/3}|n^{-2/3} \gtrsim |i-j|(i+j)^{-1/3}n^{-2/3}$$

for $(i,j)\notin\mathcal{A}$. Using the inequalities (3.67) and (3.68) of [51], we can bound (C.46) by

$$|\zeta_i| \le C\frac{n^{\epsilon}}{n^{1/3}n^{2\omega_\ell}} \quad \text{with high probability.} \quad \text{(C.47)}$$

For $1 \le i \le n^{\omega_a}$, through a lengthy but straightforward calculation, we find that

$$\zeta_i =: A_1 + A_2 + A_3 + A_4 + A_5 + A_6,$$

where A_i, $1 \le i \le 6$, are defined as (recall (C.34))

$$A_1 := -\frac{\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}}{n}\sum_{j}^{\mathcal{A}^c,(i)} \frac{1}{z_i(t,\alpha)-z_j(t,\alpha)} + c_n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i^c(t,\alpha)} \frac{\rho_t(E_+(t,\alpha)-E,\alpha)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E,$$

$$A_2 := c_n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E - c_n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i^c(t,\alpha)} \frac{\rho_t(E_+(t,\alpha)-E,\alpha)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E,$$

$$A_3 := c_n\left(\sqrt{E_+(t,0)}-\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}\right) \times \int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E,$$

$$A_4 := c_n\sqrt{E_+(t,0)} \times \left[\int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widetilde{y}_i(t,\alpha))^2}\mathrm{d}E - \int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right],$$

$$A_5 := \frac{n-p}{2n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}} - \frac{n-p}{2n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}} + \frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t} - \frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t},$$

$$A_6 := -\frac{n-p}{2n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}} + \frac{n-p}{2n\sqrt{E_+(t,0)}} - \frac{1}{2n}\sum_{j}^{\mathcal{A},(i)} \frac{1}{2\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}.$$

First, for term A_6, we notice that for $1 \le i \le n^{\omega_a}$ there are at most $\mathrm{O}(n^{2\omega_a/3+\omega_\ell})$ many indices j such that $(i,j)\in\mathcal{A}$. On the other hand, by (C.17) and (C.43), we have $|\widetilde{z}_i(t,\alpha)| \le n^{-2/3+2\omega_a/3+\varepsilon}$ with high probability. Hence, we can bound that

$$|A_6| \le n^{-2/3+\omega_a}, \quad \text{with high probability.} \quad \text{(C.48)}$$

Next, using $\rho_t(E_+(t,0)-E,0) = \mathrm{O}(\sqrt{E})$, we can bound that

$$\int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{|E-\widetilde{z}_i(t,\alpha)|}\mathrm{d}E = \mathrm{O}(1),$$

which immediately gives

$$|A_3| \lesssim |\widetilde{z}_i(t,\alpha)| \le n^{-2/3+\omega_a} \text{with high probability.} \quad \text{(C.49)}$$

For A_4, we have

$$|A_4| \lesssim \int_{\mathcal{I}_i^c(t,0)} \frac{|\widetilde{y}_i(t,\alpha)||\sqrt{E_+(t,0)}-\sqrt{E_+(t,\alpha)}|}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widetilde{y}_i(t,\alpha))^2} \times \int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E. \quad \text{(C.50)}$$

Note that for $E\in\mathcal{I}_i^c(t,\alpha)$, we have

$$|E-\widetilde{z}_i(t,\alpha)| \gtrsim n^{-2/3+2\omega_\ell} + i^{2/3}n^{-2/3},$$

and

$$|E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widetilde{y}_i(t,\alpha))^2| \gtrsim n^{-2/3+2\omega_\ell} + i^{2/3}n^{-2/3}.$$

Thus, we can bound the integral on the right-hand side of (C.50) by

$$\int_{\mathcal{I}_i^c(t,0)} \frac{\rho_t(E_+(t,0)-E,0)\mathrm{d}E}{[E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widetilde{y}_i(t,\alpha))^2][E-\widetilde{z}_i(t,\alpha)]} \lesssim n^{1/3-\omega_\ell}.$$

Together with $|\sqrt{E_+(t,0)}-\sqrt{E_+(t,\alpha)}| \lesssim t$ by (C.23) and the rigidity estimate (C.30) for $|\widetilde{y}_i(t,\alpha)|$, we get that for any constant $\varepsilon > 0$,

$$|A_4| \lesssim t|\widetilde{y}_i(t,\alpha)|n^{1/3-\omega_\ell} \lesssim n^{-1/3+\omega_1}\left(\frac{i^{2/3}}{n^{2/3}} + \frac{n^{\varepsilon}}{i^{1/3}n^{2/3}}\right)n^{1/3-\omega_\ell} \le n^{-2/3+\omega_a}, \quad \text{(C.51)}$$

with high probability. The term A_1 can be handled in exactly the same way as B_1 in (3.71) of [51] and we get that for any constant $\varepsilon > 0$,

$$|A_1| \le n^{-1/3-2\omega_\ell+\varepsilon} + n^{-1/2+\varepsilon} \quad \text{(C.52)}$$

with high probability. Finally, using the definitions of $m_t(\widetilde{z}_i(t,\alpha),0)$ and $m_t(\widetilde{z}_i(t,\alpha),\alpha)$ we can write A_2+A_5 as

$$\begin{aligned}
A_2+A_5 &= \frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t} - \frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t} \\
&+ c_n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}Rem_t(E_+(t,\alpha)-\widetilde{z}_i(t,\alpha),\alpha) \\
&- c_n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}Rem_t(E_+(t,0)-\widetilde{z}_i(t,\alpha),0) \\
&- \frac{1-c_n}{2\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}} + \frac{1-c_n}{2n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}} \\
&+ c_n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i(t,\alpha)}\frac{\rho_t(E_+(t,\alpha)-E,\alpha)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&- c_n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&:= B_1+B_2+\mathrm{O}(t^2), \qquad \text{(C.53)}
\end{aligned}$$

where

$$\begin{aligned}
B_1 &:= c_n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)} \\
&\quad\times\int_{\mathcal{I}_i(t,\alpha)}\frac{\rho_t(E_+(t,\alpha)-E,\alpha)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&\quad - c_n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E, \\
B_2 &:= \Psi_t(\widetilde{z}_i(t,\alpha),\alpha)-\Psi_t(\widetilde{z}_i(t,\alpha),0).
\end{aligned}$$

In the second step of (C.53), we used (C.18) and (C.19). Now using Lemma 20, we can bound that for any constant $\varepsilon>0$,

$$\begin{aligned}
|B_2| &\le \frac{n^{\varepsilon}t^{1/2}}{t_0^{1/2}}|\widetilde{z}_i(t,\alpha)|^{1/2} + \frac{n^{\varepsilon}|\widetilde{z}_i(t,\alpha)|}{t_0} + \mathrm{O}(t^2) \\
&\lesssim n^{-1/3+\varepsilon+\omega_a/3+\omega_1/2-\omega_0/2} \qquad \text{(C.54)}
\end{aligned}$$

with high probability, where in the second step we used $|\widetilde{z}_i(t,\alpha)|\lesssim n^{-2/3+2\omega_a/3}$ by (C.17) because the largest index i_+ is at most $\mathrm{O}(n^{\omega_a})$.

It remains to bound B_1:

$$B_1 = B_{11}+B_{12},$$

where

$$\begin{aligned}
B_{11} &:= c_n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)} \\
&\quad\times\left[\int_{\mathcal{I}_i(t,\alpha)}\frac{\rho_t(E_+(t,\alpha)-E,\alpha)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right. \\
&\quad\left.-\int_{\mathcal{I}_i(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right], \\
B_{12} &:= c_n\left[\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}-\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}\right] \\
&\quad\times\int_{\mathcal{I}_i(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E.
\end{aligned}$$

The term B_{11} can be bounded in the same way as (3.82) of [51], which gives that for any constant $\varepsilon>0$,

$$|B_{11}|\le n^{-1/3+\varepsilon+2\omega_a/3+\omega_1-\omega_0} \qquad \text{(C.55)}$$

with high probability. For B_{12}, we need to obtain a bound on

$$\int_{\mathcal{I}_i(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E. \qquad \text{(C.56)}$$

Note that this is a principal value, so we need to deal with the logarithmic singularity at $\widetilde{z}_i(t,\alpha)$. First assume that $i\ge n^{\delta}$ for some $\delta<\omega_\ell/10$. Then with (C.17) and (C.23), it is easy to check that $\widetilde{z}_i(t,\alpha)$ is away from the boundary of $\mathcal{I}_i(t,0)$ at least by a distance n^{-2} with high probability, and that $|\widetilde{z}_i(t,\alpha)|\ge n^{-2}$ with high probability. Then, we can bound that

$$\begin{aligned}
&\left|\int_{\mathcal{I}_i(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right| \\
&\le \int_{\mathcal{I}_i(t,0),|E-\widetilde{z}_i(\alpha,t)|>n^{-50}}\frac{\sqrt{|E|}}{|\widetilde{z}_i(t,\alpha)-E|}\mathrm{d}E \\
&\quad+\left|\int_{|E-\widetilde{z}_i(t,\alpha)|\le n^{-50}}\frac{\rho_t(E_+(t,0)-E,0)}{\widetilde{z}_i(t,\alpha)-E}\right. \\
&\quad\left.-\frac{\rho_t(E_+(t,0)-\widetilde{z}_i(t,\alpha),0)}{\widetilde{z}_i(t,\alpha)-E}\mathrm{d}E\right| \\
&=: D_1+D_2.
\end{aligned}$$

For the term D_1, using $\sqrt{|E|}=\mathrm{O}(n^{-1/3+\omega_a/3})$ for $E\in\mathcal{I}_i(0,t)$, we obtain that

$$\begin{aligned}
D_1 &\lesssim n^{-1/3+\omega_a/3}\int_{\mathcal{I}_i(t,0),|E-\widetilde{z}_i(\alpha,t)|>n^{-50}}\frac{\mathrm{d}E}{|\widetilde{z}_i(t,\alpha)-E|} \\
&\lesssim n^{-1/3+\omega_a/3}\log n.
\end{aligned}$$

For the term D_2, using Lemma 12 we obtain that

$$\begin{aligned}
&|\rho_t(E_+(t,0)-E,0)-\rho_t(E_+(t,0)-\widetilde{z}_i(t,\alpha),0)| \\
&\lesssim \frac{|E-\widetilde{z}_i(t,\alpha)|}{\min(t,|E_+(t,0)-\widetilde{z}_i(t,\alpha)|^{1/2})}\le n|E-\widetilde{z}_i(t,\alpha)|,
\end{aligned}$$

which gives that

$$D_2\lesssim n\int_{|\widetilde{z}_i(t,\alpha)-E|<n^{-50}}\mathrm{d}E\le 2n^{-49}.$$

Next, we consider the case with $i<n^{\delta}$. It suffices to assume that $|\widetilde{z}_i(t,\alpha)|\le n^{-100}$, because otherwise we can obtain an estimate in the same way as the case $i\ge n^{\delta}$. Then we decompose (C.56) as

$$\begin{aligned}
&\int_{\mathcal{I}_i(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&=\int_{\mathcal{I}_i(t,0),E\ge n^{-50}}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&\quad+\int_{3\widetilde{z}_i(t,\alpha)/2\le E<n^{-50}}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&\quad+\int_{\widetilde{z}_i(t,\alpha)\le E<3\widetilde{z}_i(t,\alpha)/2}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&\quad+\int_{0\le E<\widetilde{z}_i(t,\alpha)/2}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&=: F_1+F_2+F_3+F_4.
\end{aligned}$$

The term F_1 can be estimated in the same way as D_1. For term F_2, we used that $\rho_t(E_+(t,0)-E,0)=\mathrm{O}(\sqrt{E})$ to bound the integral as $|F_2|\lesssim n^{-25}$. If $\widetilde{z}_i(t,\alpha)\le 0$, then we have $F_3=F_4=0$. Otherwise, for F_4 we have

$$|F_4|\le\int_{0\le E<\widetilde{z}_i(t,\alpha)/2}E^{-1/2}\mathrm{d}E\lesssim|\widetilde{z}_i(t,\alpha)|^{1/2}\le n^{-50},$$

and for F_3 we have

$$|F_3| \le \int_{\widetilde{z}_i(t,\alpha)\le E<3\widetilde{z}_i(t,\alpha)/2} \mathrm{d}E \times \frac{|\rho_t(E_+(t,0)-E,0)-\rho_t(E_+(t,0)-\widetilde{z}_i(t,\alpha),0)|}{|E-\widetilde{z}_i(t,\alpha)|} \lesssim |\widetilde{z}_i(t,\alpha)|^{1/2} \le n^{-50},$$

where in the second step we used that $|\rho_t(E_+(t,0)-E,0)-\rho_t(E_+(t,0)-\widetilde{z}_i(t,\alpha),0)| \le |\widetilde{z}_i(t,\alpha)|^{-1/2}|E-\widetilde{z}_i(t,\alpha)|$. Combining the above estimates, we get that for any constant $\varepsilon>0$, with high probability

$$\left|\int_{\mathcal{I}_i(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right| \le n^{-1/3+\omega_a+\varepsilon},$$

which further implies that

$$|B_{12}| \lesssim n^{-1/3+\omega_a+\varepsilon}t = n^{-2/3+\varepsilon+\omega_a+\omega_1}. \tag{C.57}$$

In sum, combining (C.48), (C.49), (C.51), (C.52), (C.54), (C.55) and (C.57) and using the hierarchy of parameters (C.41), we obtain that for $1\le i\le n^{\omega_a}$,

$$|\zeta_i(t)| \le n^{-1/3-2\omega_\ell+\varepsilon} \quad \text{with high probability,} \tag{C.58}$$

for any constant $\varepsilon>0$. Then, combining (C.47) and (C.58), we obtain that for any constant $\varepsilon>0$,

$$\|\zeta(t)\|_\infty \le n^{-1/3-2\omega_\ell+\varepsilon} \quad \text{with high probability,} \tag{C.59}$$

uniformly in all $n^{-C_0}\le t\le t_1$ under the assumption (C.43). Plugging it into (C.44), we get

$$\|v(t)\|_\infty \le tn^{-1/3-2\omega_\ell+\varepsilon} = n^{-2/3+\varepsilon+\omega_1-2\omega_\ell},$$

which concludes (C.42) under (C.43). Note that the right hand side of (C.42) is much smaller than $n^{-2/3}$ on the right-hand side of (C.43). Then using a simple continuity argument we can remove the assumption (C.43). In fact, the continuity argument is deterministic in nature because v satisfies a system of deterministic equations conditioning on the trajectories of $\{\widetilde{y}_i(t,\alpha)\}$ and $\{\widehat{y}_i(t,\alpha)\}$. In fact, we can pick a high probability event Ξ, on which the rigidity (C.17) and the local law, Theorem 6, hold for all $n^{-C_0}\le t\le t_1$. Then, we can perform the continuity argument on Ξ. □

Before concluding this section, we record the following rigidity estimates.

Corollary 4: Let $i\le n^{3\omega_\ell+\delta}$ for a constant $0<\delta<\omega_\ell-\omega_1$. Then we have

$$\sup_{0\le t\le 10t_1} |\widehat{y}_i(t,\alpha)-\widehat{\gamma}_i(t,\alpha)| \prec i^{-1/3}n^{-2/3}.$$

Proof: This is an immediate consequence of Lemma 23 and (C.30). □

C. Proof of Theorem 7

Our goal is to bound $|\widehat{y}_i(t_1,\alpha=1)-\widehat{y}_i(t_1,\alpha=0)|$. For this purpose, we shall study the partial derivative $u_i(t,\alpha):=\partial_\alpha\widehat{y}_i(t,\alpha)$. With (C.37)–(C.39), we find that $u=(u_i(t,\alpha): 1\le i\le p)$ satisfies the PDE

$$\partial_t u = \mathcal{L}u+\zeta^{(0)}. \tag{C.60}$$

Here the operator $\mathcal{L}$ is defined as $\mathcal{L}=\mathcal{B}+\mathcal{V}$, where $\mathcal{B}$ is defined by

$$(\mathcal{B}u)_i = -\frac{1}{2n}\sum_j^{\mathcal{A},(i)} \frac{u_i-u_j}{[\widehat{y}_i(t,\alpha)-\widehat{y}_j(t,\alpha)]^2}. \tag{C.61}$$

$\mathcal{V}$ is a diagonal operator with $(\mathcal{V}u)_i=\mathcal{V}_iu_i$, where $\mathcal{V}_i$'s are defined as

$$\mathcal{V}_i = -2c_n\sqrt{E_+(t,0)}\int_{\mathcal{I}_i^c(t,0)}\mathrm{d}E \times \frac{[\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha)]\rho_t(E_+(t,0)-E,0)}{[E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2]^2} \tag{C.62}$$

for $1\le i\le n^{\omega_a}$,

$$\mathcal{V}_i = -\frac{c_n}{2}\int_{\widehat{\mathcal{I}}_i^c(t,\alpha)\cap\mathcal{J}(t,\alpha)} \frac{f_t(\sqrt{E_+(t,\alpha)}-E,\alpha)}{[\widehat{y}_i(t,\alpha)-E]^2}\mathrm{d}E \tag{C.63}$$

for $n^{\omega_a}<i\le i_*/2$, and $\mathcal{V}_i=0$ for $i_*/2<i\le p$. With the same discussion as the one below (C.43), we know that the semigroup of $\mathcal{L}$ is a contraction on every $\ell^q(\{1,\cdots,p\})$ space. The random forcing term $\zeta^{(0)}$ comes from the ∂_α derivatives of all the other terms, and we notice that $\zeta_i^{(0)}=0$ when $1\le i\le n^{\omega_a}$. For $i>n^{\omega_a}$, it is easy to check that for some constant $C>0$,

$$\max_{i>n^{\omega_a}}|\zeta_i^{(0)}| \le n^C \quad \text{with high probability.} \tag{C.64}$$

Next, we define a long range cut-off of u. Fix a small constant $\delta_v>0$ and let v be the solution to the following homogeneous equation

$$\partial_t v = \mathcal{L}v, \quad v_i(n^{-C_0}) = u_i(n^{-C_0})\mathbf{1}_{\{1\le i\le \ell^3n^{\delta_v}\}}. \tag{C.65}$$

Then, we have the following proposition, which essentially states that the u_i's with indices far away from the edge have a negligible effect on the solution.

Proposition 1: With high probability, we have

$$\sup_{n^{-C_0}\le t\le 10t_1}\sup_{1\le i\le\ell^3}|u_i(t,\alpha)-v_i(t,\alpha)| \le n^{-100}.$$

One can see that Proposition 1 is an immediate consequence of the following finite speed of propagation estimate, whose proof is postponed to Section C-D.

Lemma 24: For any small constant $\delta>0$, we have that for $a\ge\ell^3n^{2\delta}$ and $b\le\ell^3n^\delta$,

$$\sup_{n^{-C_0}\le s\le t\le 10t_1}\left[\mathcal{U}_{ab}^{\mathcal{L}}(s,t)+\mathcal{U}_{ba}^{\mathcal{L}}(s,t)\right] \le n^{-D}$$

with high probability for any large constant $D>0$.

Remark 12: In fact, we have $\mathcal{U}_{ab}^{\mathcal{L}}(s,t)\ge0$ and $\mathcal{U}_{ba}^{\mathcal{L}}(s,t)\ge0$ by maximum principle. More precisely, define $v_i(t)=\exp(-\int_0^t\mathcal{V}_i(s)\mathrm{d}s)u_i(t)$. Then $v=(v_i:1\le i\le p)$ satisfies the equation $\partial_t v=\mathcal{B}v$. If $v_i(s)\ge0$ for all i at time s, we claim that $v_i(t)\ge0$ for all i at any time $t\ge s$. To see this, at any time $t'\in[s,t]$, suppose $v_j(t')=\min\{v_i(t'):1\le i\le p\}$ is the smallest entry of $v(t')$. Then with (C.61), we can check that $\partial_t v_j(t')=(\mathcal{B}v(t'))_j\ge0$, i.e. the smallest entry of v will always increase. Hence the entries of v can never be negative at any time $t\ge s$.

Proof of Proposition 1: Fix a $n^{-C_0} \le t \le 10t_1$, by Duhamel's principle we have that

$$u(t,\alpha) - v(t,\alpha) = \mathcal{U}^{\mathcal{L}}(n^{-C_0},t)[u(n^{-C_0},\alpha) - v(n^{-C_0},\alpha)] + \int_{n^{-C_0}}^{t} \mathcal{U}^{\mathcal{L}}(s,t)\zeta^{(0)}(s)\mathrm{d}s.$$

Since $u_i(n^{-C_0},\alpha) - v_i(n^{-C_0},\alpha) = 0$ for $i \le \ell^3 n^{\delta_v}$ and $\zeta_i^{(0)}(s) = 0$ for $i \le n^{\omega_a}$, we can conclude the proof using Lemma 24 and (C.64). □

Another key ingredient is the following energy estimate. We postpone its proof until we complete the proof of Theorem 7. Here we have fixed the starting time point to be n^{-C_0}, but the same conclusion holds for any other starting time by the semigroup property.

Proposition 2: For any small constant $\delta_1 > 0$, consider a vector $w \in \mathbb{R}^p$ with $w_i = 0$ for $i \ge \ell^3 n^{\delta_1}$. Then, for any constants $\epsilon, \eta > 0$ and fixed $q \ge 1$, there exists a constant $C_q > 0$ independent of ϵ and η such that for all $2n^{-C_0} \le t \le 2t_1$,

$$\|\mathcal{U}^{\mathcal{L}}(n^{-C_0},t)w\|_\infty \le C(q,\eta)\left(\frac{n^{C_q\eta+\epsilon}}{n^{1/3}t}\right)^{3(1-6\eta)/q}\|w\|_q. \tag{C.66}$$

With all the above preparations, we are now ready to give the proof of Theorem 7.

Proof of Theorem 7: Fix any $1 \le i \le \mathfrak{a}$, by (C.26) and (C.27) we have that with high probability,

$$\begin{aligned}&|[\lambda_i(t_1) - E_\lambda(t_1)] - [\mu_i(t_1) - E_\mu(t_1)]| \\ &\le |\widetilde{z}_i(t_1,1) - \widetilde{z}_i(t_1,0)| + |E_+(t_1,1) - E_\lambda(t_1)| \\ &\quad + |E_+(t_1,0) - E_\mu(t_1)| \\ &\le |\widetilde{z}_i(t_1,1) - \widetilde{z}_i(t_1,0)| + n^{-2/3-\tau}\end{aligned}$$

for some small constant $\tau > 0$. Recalling (C.34), we have that

$$\begin{aligned}&|\widetilde{z}_i(t_1,1) - \widetilde{z}_i(t_1,0)| \\ &\le |\widetilde{y}_i(t_1,1) - \widetilde{y}_i(t_1,0)|\left(2\sqrt{E_+(t_1,0)} - y_i(t_1,0)\right) \\ &\quad + |\widetilde{y}_i(t_1,1)|\left|2\sqrt{E_+(t_1,1)} - y_i(t_1,1) - 2\sqrt{E_+(t_1,0)} + y_i(t_1,0)\right| \\ &\lesssim |\widetilde{y}_i(t_1,1) - \widetilde{y}_i(t_1,0)| + \mathrm{O}_\prec\left(n^{-2/3}t_1\right),\end{aligned}$$

where in the second step we used (B.23) and the rigidity estimate (C.30). Together with Lemma 23, we obtain that with high probability,

$$\begin{aligned}&|[\lambda_i(t_1) - E_\lambda(t_1)] - [\mu_i(t_1) - E_\mu(t_1)]| \\ &\lesssim |\widehat{y}_i(t_1,1) - \widehat{y}_i(t_1,0)| + n^{-2/3-\tau}\end{aligned} \tag{C.67}$$

for some small constant $\tau > 0$. Now, we write that

$$\widehat{y}_i(t_1,1) - \widehat{y}_i(t_1,0) = \int_0^1 u_i(t_1,\alpha)\mathrm{d}\alpha.$$

Applying Proposition 1 (together with a simple stochastic continuity argument to pass to all $0 \le \alpha \le 1$), we get that

$$|\widehat{y}_i(t_1,0) - \widehat{y}_i(t_1,1)| \le n^{-50} + \left|\int_0^1 v_i(t_1,\alpha)\mathrm{d}\alpha\right|, \tag{C.68}$$

with high probability. By (C.10) and (C.12), we have that at $t = 0$, for $1 \le j \le \ell^3 n^{\delta_v}$,

$$|z_j(t=0,0) - z_j(t=0,1)| \prec n^{-2/3-\omega_0} + j^{-1/3}n^{-2/3},$$

where $\delta_v > 0$ is a small enough constant. Moreover, at $t = n^{-C_0}$ the eigenvalues are perturbed at most by $n^{-C_0/2}$, so we can calculate that for $0 \le \alpha \le 1$,

$$\|v(n^{-C_0},\alpha)\|_4 \prec n^{-2/3-\omega_0}(\ell^3 n^{\delta_v})^{1/4} + n^{-2/3} \le 2n^{-2/3}.$$

Finally, using Proposition 2 with $q = 4$, we find that

$$\left|\int_0^1 v_i(t_1,\alpha)\mathrm{d}\alpha\right| \prec n^{-2/3-\omega_1/2}.$$

Inserting it into (C.68) and further into (C.67), we conclude the proof. □

The proof of Proposition 2 is almost the same as the one for Lemma 3.11 in [51], so we only give an outline of it.

Proof of Proposition 2: The proof relies on Lemma 24 and the estimates in the following lemma.

Lemma 25: Fix a constant $0 < \delta_1 < \omega_\ell - \omega_1$. Let $w \in \mathbb{R}^p$ be a vector such that $w_i = 0$ for $i \ge \ell^3 n^{\delta_1}$. For any constants $\eta, \epsilon > 0$, there is a constant $C > 0$ independent of ϵ and η, and a constant $c_\eta > 0$ such that the following estimates hold with high probability for all $n^{-C_0} \le s \le t \le 5t_1$:

$$\|\mathcal{U}^{\mathcal{L}}(s,t)w\|_2 \le \left(\frac{n^{C\eta+\epsilon}}{c_\eta n^{1/3}(t-s)}\right)^{\frac{3}{2}(1-6\eta)}\|w\|_1, \tag{C.69}$$

and

$$\|(\mathcal{U}^{\mathcal{L}}(s,t))^\top w\|_2 \le \left(\frac{n^{C\eta+\epsilon}}{c_\eta n^{1/3}(t-s)}\right)^{\frac{3}{2}(1-6\eta)}\|w\|_1. \tag{C.70}$$

Proof: The proof is very similar to the ones for [51, Lemma 3.13], [12, Proposition 10.4] and [36, Section 10]. More precisely, our operator $\mathcal{L}$ is almost the same as the operator $\mathcal{L}$ in [51, Lemma 3.13], where the only difference is the form of $\mathcal{V}$. However, the $\mathcal{V}_i$'s in (C.62) and (C.63) satisfy exactly the same estimates as the $\mathcal{V}_i$'s in [51]. So we omit the details of the proof. □

Now, we complete the proof of Proposition 2. Fix constants $0 < \delta_1 < \delta_2 < \omega_\ell - \omega_1$. We define the indicator function $\mathcal{X}_2(i) = \mathbf{1}_{\{1 \le i \le \ell^3 n^{\delta_2}\}}$ and let $\mathcal{X}_2$ be the associated digonal operator. For any $v \in \mathbb{R}^p$ with $\|v\|_1 = 1$, we decompose that

$$\begin{aligned}\langle \mathcal{U}^{\mathcal{L}}w, v\rangle &= \langle w, (\mathcal{U}^{\mathcal{L}})^\top v\rangle \\ &= \langle w, (\mathcal{U}^{\mathcal{L}})^\top \mathcal{X}_2 v\rangle + \langle w, (\mathcal{U}^{\mathcal{L}})^\top (1-\mathcal{X}_2)v\rangle,\end{aligned}$$

where we have abbreviated $\mathcal{U}^{\mathcal{L}} \equiv \mathcal{U}^{\mathcal{L}}(n^{-C_0},t)$. For the second term, with Lemma 24, we obtain that

$$\left|\langle w, (\mathcal{U}^{\mathcal{L}})^\top (1-\mathcal{X}_2)v\rangle\right| \le n^{-100}\|w\|_1\|v\|_1 \le n^{-99}\|w\|_2\|v\|_1$$

with high probability. For the first term, with Lemma 25 and Cauchy-Schwarz inequality, we get that for any constant $\eta > 0$,

$$\begin{aligned}\langle w, (\mathcal{U}^{\mathcal{L}})^\top \mathcal{X}_2 v\rangle &\le \|w\|_2\|(\mathcal{U}^{\mathcal{L}})^\top \mathcal{X}_2 v\|_2 \\ &\le \|w\|_2\left(\frac{n^{C\eta+\epsilon}}{c_\eta n^{1/3}(t-n^{-C_0})}\right)^{\frac{3}{2}(1-6\eta)}\|v\|_1.\end{aligned}$$

By ℓ^1–ℓ^∞ duality and using $t \geq 2n^{-C_0}$, we find that

$$\|\mathcal{U}^{\mathcal{L}} w\|_\infty \leq C(\eta) \left(\frac{n^{C\eta+\epsilon}}{n^{1/3}t} \right)^{\frac{3}{2}(1-6\eta)} \|w\|_2.$$

Consequently, by the semigroup property, we find that

$$\begin{aligned} \|\mathcal{U}^{\mathcal{L}}(n^{-C_0}, t)w\|_\infty &= \|\mathcal{U}^{\mathcal{L}}(2t/3, t)\mathcal{U}^{\mathcal{L}}(n^{-C_0}, 2t/3)w\|_\infty \\ &\leq C(\eta) \left(\frac{n^{C\eta+\epsilon}}{n^{1/3}t} \right)^{\frac{3}{2}(1-6\eta)} \|\mathcal{U}^{\mathcal{L}}(n^{-C_0}, 2t/3)w\|_2 \\ &\leq C(\eta) \left(\frac{n^{C\eta+\epsilon}}{n^{1/3}t} \right)^{3(1-6\eta)} \|w\|_1, \end{aligned}$$

where we used Lemma 25 again in the last step. Finally, the estimate (C.66) for general q follows from a standard interpolation argument. □

D. Proof of Lemma 24

Finally, in this section, we prove the finite speed of propagation estimate, Lemma 24. For simplicity of notations, we shift the time such that the starting time point is $t = 0$. We first prove a result for fixed s.

Lemma 26: Fix a small constant $0 < \delta < \omega_\ell - \omega_1$. For any $a \geq \ell^3 n^\delta$, $b \leq \ell^3 n^\delta/2$ and fixed $0 \leq s \leq 10t_1$, we have that for any large constant $D > 0$, with high probability,

$$\sup_{t:s\leq t\leq 10t_1} \left[\mathcal{U}^{\mathcal{L}}_{ab}(s,t) + \mathcal{U}^{\mathcal{L}}_{ba}(s,t)\right] \leq n^{-D}.$$

We postpone its proof until we complete the proof of Lemma 24. We need to use the following lemma in order to extend the result in Lemma 26 to all $0 \leq s \leq t \leq 10t_1$ simultaneously.

Lemma 27: Let $u \in \mathbb{R}^p$ be a solution of $\partial_t u = \mathcal{L}u$ with $u_i(0) \geq 0$ for $1 \leq i \leq p$. Then, for $0 \leq t \leq 10t_1$, we have

$$\frac{1}{2}\sum_i u_i(0) \leq \sum_i u_i(t) \leq \sum_i u_i(0).$$

Proof: Summing over i and using $\sum_i (\mathcal{B}u)_i = 0$, we get that

$$\partial_t \sum_i u_i = \sum_i \mathcal{V}_i u_i.$$

We now bound (C.62) and (C.63). Using (C.30) and Lemma 23, we have that with high probability, for $1 \leq i \leq n^{\omega_a}$ and $E \in \mathcal{I}_i^c(t,0)$,

$$E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{\gamma}_i(t,\alpha))^2 \gtrsim E + n^{-2/3+2\omega_\ell}.$$

Together with the estimate $\rho_t(E_+(t,0) - E, 0) \sim \sqrt{E}$, we get that for $1 \leq i \leq n^{\omega_a}$,

$$0 \leq -\mathcal{V}_i \lesssim \int_{\mathcal{I}_i^c(t,0)} \frac{\sqrt{E}}{|E + n^{-2/3+2\omega_\ell}|^2} \mathrm{d}E \lesssim n^{1/3-\omega_\ell}.$$

We can get the same bound for (C.63). Then, applying Gronwall's inequality to

$$-\left(Cn^{1/3-\omega_\ell}\right)\sum_i u_i \leq \partial_t \sum_i u_i \leq 0,$$

we can conclude the proof. □

Now, we can complete the proof of Lemma 24.

Proof of Lemma 24: Fix any constant $0 < \varepsilon < \delta$, $a \geq \ell^3 n^{2\delta}$ and $b \leq \ell^3 n^\delta$. By the semigroup property, we have

$$\begin{aligned} \mathcal{U}^{\mathcal{L}}_{bi}(n^{-C_0}, t) &= \sum_j \mathcal{U}^{\mathcal{L}}_{bj}(s,t)\mathcal{U}^{\mathcal{L}}_{ji}(n^{-C_0}, s) \\ &\geq \mathcal{U}^{\mathcal{L}}_{ba}(s,t)\mathcal{U}^{\mathcal{L}}_{ai}(n^{-C_0}, s). \end{aligned} \tag{C.71}$$

By Lemma 27, we find that $\sum_i \mathcal{U}^{\mathcal{L}}_{ai}(n^{-C_0}, s) \geq 1/2$. Moreover, by Lemma 26 we have that $\mathcal{U}^{\mathcal{L}}_{ai}(n^{-C_0}, s) \leq n^{-100}$ for any $i \leq \ell^3 n^{\delta+\epsilon}$. This implies that there exists an $i_* \geq \ell^3 n^{\delta+\varepsilon}$ such that $\mathcal{U}^{\mathcal{L}}_{ai_*}(n^{-C_0}, s) \geq (4n)^{-1}$. However, by Lemma 26 we have that $\mathcal{U}^{\mathcal{L}}_{bi_*}(0,t) \leq n^{-D}$ for any large constant $D > 0$. Thus picking $i = i_*$ in (C.71), we get that $\mathcal{U}^{\mathcal{L}}_{ba}(s,t) \leq n^{-D+2}$. This finishes the proof for the estimate on $\mathcal{U}^{\mathcal{L}}_{ba}(s,t)$. The estimate on $\mathcal{U}^{\mathcal{L}}_{ab}(s,t)$ can be proved in a similar way. □

It remains to prove Lemma 26. The strategy was first developed in [14], and later used in [50], [51] to study the symmetric DBM for Wigner type matrices. Our proof is similar to the ones for [50, Lemma 4.2] and [51, Lemma 4.1], so we will not write down all the details.

Proof of Lemma 26: We focus on the case $s = 0$ and the general case can be dealt with similarly using a simple time shift. Let ψ be a smooth function satisfying the following properties: (i) $\psi(x) = -x$ for $|x| \leq \ell^2 n^{-2/3+2\delta/3}$, (ii) $\psi'(x) = 0$ for $|x| > 2\ell^2 n^{-2/3+2\delta/3}$, (iii) ψ is decreasing, (iv) $|\psi(x) - \psi(y)| \leq |x-y|$ and $|\psi'(x)| \leq 1$, and (v) $|\psi''(x)| \leq C\ell^{-2} n^{2/3-2\delta/3}$ for some constant $C > 0$. Similar to [51, Lemma 4.1], we now consider a solution of

$$\partial_t f = \mathcal{L}f, \quad \text{with} \quad f_i(0) = \delta_{q_*},$$

for any $q_* \geq q := \ell^3 n^\delta$. Let $\nu > 0$ be a fixed constant and define the functions

$$\phi_k := \exp\left[\nu\psi(\widehat{y}_k(t,\alpha) - \widehat{\gamma}_q(t,\alpha))\right], \quad v_k := \phi_k f_k,$$

and

$$F(t) := \sum_k v_k^2.$$

For our proof, we will choose a specific ν later. By Ito's formula, we find that F satisfies the SDE

$$\mathrm{d}F = -\sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij}(v_i - v_j)^2 \mathrm{d}t + 2\sum_i \mathcal{V}_i v_i^2 \mathrm{d}t \tag{C.72}$$

$$+ \sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij} v_i v_j \left(\frac{\phi_i}{\phi_j} + \frac{\phi_j}{\phi_i} - 2 \right) \mathrm{d}t \tag{C.73}$$

$$+ 2\nu \sum_i v_i^2 \psi'(\widehat{y}_i - \widehat{\gamma}_q) \mathrm{d}(\widehat{y}_i - \widehat{\gamma}_q) \tag{C.74}$$

$$+ \sum_i v_i^2 \left(\frac{\nu^2}{n} [\psi'(\widehat{y}_i - \widehat{\gamma}_q)]^2 + \frac{\nu}{n}\psi''(\widehat{y}_i - \widehat{\gamma}_q) \right) \mathrm{d}t, \tag{C.75}$$

where we denoted

$$\mathcal{B}_{ij} = \frac{1}{2n} \frac{1}{(\widehat{y}_i(t,\alpha) - \widehat{y}_j(t,\alpha))^2}.$$

Now, we choose a proper stopping time. Let τ_1 be the stopping time such that for $t < \tau_1$, Lemmas 19 and 23 hold true for a sufficiently small constant $0 < \epsilon < \delta/100$. Note that $\tau_1 \geq 10t_1$

with high probability. Let τ_2 be the first time such that $F \geq 10$. Then, we define the stopping time

$$\tau := \min\{\tau_1, \tau_2, 10t_1\}.$$

For the rest of the proof, we only consider times with $t < \tau$. We will show that with a suitable choice of ν, we actually have $\tau = 10t_1$ with high probability.

We now deal with each term in (C.72)-(C.75). First, (C.72) is a dissipative term, so it only decreases the size of $F(t)$. By Corollary 4, we see that $\psi'(\widehat{y}_i - \widehat{\gamma}_q) = 0$ when $i > C'\ell^3 n^\delta$ for a large enough constant $C' > 0$. Moreover, if $i \leq C'\ell^3 n^\delta$ and $(i,j) \in \mathcal{A}$, then $j \leq C\ell^3 n^\delta$ for some constant $C > 0$ depending on C'. Thus the nonzero terms in (C.73) must satisfy that $i, j \leq C\ell^3 n^\delta$ for a large enough constant $C > 0$. Then, by Corollary 4, for $i, j \leq C\ell^3 n^\delta$ satisfying $(i,j) \in \mathcal{A}$, we have

$$|\widehat{y}_i - \widehat{y}_j| \lesssim \ell^2 n^{-2/3+\delta/3}.$$

Now, with the Taylor expansion of $e^{-x} + e^x - 2$, we get that if

$$\nu\ell^2 n^{-2/3+\delta/3} \leq C_1 \tag{C.76}$$

for some constant $C_1 > 0$, then

$$\begin{aligned}(C.73) &\lesssim \frac{\nu^2}{n} \sum_{(i,j)\in\mathcal{A}} (v_i^2 + v_j^2)\mathbf{1}_{\{\phi_j \neq \phi_i\}} \mathrm{d}t \\ &\leq \frac{\nu^2 \ell^3 n^{2\delta/3}}{n} F(t)\mathrm{d}t. \end{aligned} \tag{C.77}$$

The term (C.75) can be easily bounded as

$$(C.75) \lesssim \left(\frac{\nu^2}{n} + \frac{\nu\ell^{-2}}{n^{1/3+2\delta/3}}\right) F(t)\mathrm{d}t. \tag{C.78}$$

It remains to control (C.74). Since $\psi'(\widehat{y}_i - \zeta_q) \neq 0$ only when $i \leq C\ell^3 n^\delta \ll n^{\omega_a}$, thus $\widehat{y}_i$ satisfies the SDE (C.37), which gives that

$$\begin{aligned}&\mathrm{d}\left[\widehat{y}_i(t,\alpha) - \widehat{\gamma}_q(t,\alpha)\right] \\ &= -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n}\sum_j^{\mathcal{A},(i)} \frac{1}{\widehat{y}_i(t,\alpha) - \widehat{y}_j(t,\alpha)}\mathrm{d}t \\ &+ \left(\frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t} - \frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t}\right)\mathrm{d}t - \frac{n-p}{2nE_+(t,0)}\mathrm{d}t \\ &- \left[\int_{\mathcal{I}_i^c(t,0)} \frac{c_n\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2}\mathrm{d}E\right]\mathrm{d}t. \end{aligned} \tag{C.79}$$

By Burkholder-Davis-Gundy inequality and Markov's inequality, we find that for any constant $\varepsilon > 0$,

$$\sup_{0\leq t\leq\tau} \nu \left|\int_0^t \sum_i v_i^2 \psi'(\widehat{y}_i - \widehat{\gamma}_q)\frac{\mathrm{d}B_i}{\sqrt{n}}\right| \leq n^\epsilon \nu \frac{n^{\omega_1/2}}{n^{2/3}} \tag{C.80}$$

with high probability. Moreover, with the same arguments for (4.17) of [51], we obtain that

$$\begin{aligned}\frac{\nu}{n}\sum_{(i,j)\in\mathcal{A}} \frac{v_i^2\psi'(\widehat{y}_i - \widehat{\gamma}_q)}{\widehat{y}_i - \widehat{y}_j} &\leq \sum_{(i,j)\in\mathcal{A}} \frac{\mathcal{B}_{ij}}{100}(v_i - v_j)^2 \\ &+ C\left(\frac{\nu n^{\omega_\ell}}{n^{1/3}} + \frac{\nu^2 n^{3\omega_\ell + 2\delta/3}}{n}\right) F(t) \end{aligned} \tag{C.81}$$

for large enough constant $C > 0$. The main difference from the argument in [51] is about the term

$$\begin{aligned}&\frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t} - \frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t} - \frac{n-p}{2n\sqrt{E_+(t,0)}} \\ &- c_n \int_{\mathcal{I}_i^c(t,0)} \frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2}\mathrm{d}E \\ &= c_n\sqrt{E_+(t,0)}\Big[m_t\left(\left(\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha)\right)^2, 0\right) \\ &\qquad - m_t\left(E_+(t,0),0\right)\Big] \\ &+ c_n \int_{\mathcal{I}_i(t,0)} \frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2}\mathrm{d}E \\ &- \frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t} + \mathrm{O}(t), \end{aligned} \tag{C.82}$$

where we used (C.18) and $\zeta_t(E_+(t,0),0) = E_+(t,0) + \mathrm{O}(t)$ in the derivation. By the square root behavior of m_t around the right edge, we have that

$$\begin{aligned}&|m_t((\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2, 0) - m_t(E_+(t,0),0)| \\ &\lesssim \sqrt{|\widehat{y}_i(t,\alpha)|} \lesssim n^{-1/3+\omega_\ell+\delta/3}\end{aligned}$$

with high probability, where we used (C.30) in the last step. For the second term on the right-hand side of (C.82), it can be bounded in the same way as (C.56) and we can get that

$$\begin{aligned}&\left|\int_{\mathcal{I}_i(t,0)} \frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2}\mathrm{d}E\right| \\ &\lesssim n^{-1/3+\omega_\ell+\delta/3},\end{aligned}$$

with high probability. Finally, we know that $\widehat{\gamma}_q(t,\alpha)$ satisfies

$$\int_0^{\widehat{\gamma}_q(t,\alpha)} \widetilde{\rho}(t,E)\mathrm{d}E = \frac{q}{p}, \quad \widetilde{\rho}(t,E) := f_t(\sqrt{E_+(t,\alpha)} - E, \alpha).$$

Taking the derivative of this equation, we get

$$\frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t} = \frac{-1}{\widetilde{\rho}(t,\widehat{\gamma}_q(t,\alpha))}\int_0^{\widehat{\gamma}_q(t,\alpha)} \partial_t\widetilde{\rho}(t,E)\mathrm{d}E.$$

It is trivial to check that $\partial_t\widetilde{\rho}(t,E) = \mathrm{O}(1)$, and we have $\widetilde{\rho}(t,\widehat{\gamma}_q) \sim \sqrt{\widehat{\gamma}_q(t,\alpha)}$ by (B.12). Thus, we obtain from the above equation that

$$\left|\frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t}\right| \lesssim \sqrt{\widehat{\gamma}_q(t,\alpha)} \lesssim n^{-1/3+\omega_\ell+\delta/3}.$$

Combining the above estimates, we get

$$|(C.82)| = \mathrm{O}(n^{-1/3+\omega_\ell+\delta/3}). \tag{C.83}$$

Now, combining (C.77), (C.78),(C.80), (C.81) and (C.83), we find that if ν satisfies the condition of (C.76), then with high probability,

$$\partial_t F(t) \leq C\left(\frac{\nu^2 n^{3\omega_l+2\delta/3}}{n} + \frac{\nu n^{\omega_\ell+\delta/3}}{n^{1/3}}\right) F(t).$$

Then, by Gronwall's inequality, we get that

$$\sup_{0\leq s\leq\tau} F(s) \leq F(0) + C\left(\frac{\nu^2 n^{3\omega_l+2\delta/3+\omega_1}}{n^{4/3}} + \frac{\nu n^{\omega_\ell+\omega_1+\delta/3}}{n^{2/3}}\right)$$

with high probability. Hence, choosing $\nu = n^{2/3-2\omega_l-\delta/3}$, we obtain by continuity that $\tau = 10t_1$ with high probability, i.e.,

$$\sup_{0\le s\le 10t_1} F(s) \le 10, \quad \text{with high probability.}$$

Now, notice that if $i \le \ell^3 n^\delta/2$, we have that

$$\nu|\widehat{y}_i(t,\alpha) - \widehat{\gamma}_q(t,\alpha)| \gtrsim n^{\delta/3}, \quad \text{with high probability.}$$

Then, by the definition of $F(t)$ and Markov's inequality, we obtain that $\mathcal{U}^{\mathcal{L}}_{iq_*}(0,t) \le n^{-D}$ for any large constant $D > 0$ if $i \le \ell^3 n^\delta/2$ and $q_* \ge \ell^3 n^\delta$. The proof for $\mathcal{U}^{\mathcal{L}}_{q_* i}$ is the same by setting $\psi \to -\psi$. □

ACKNOWLEDGMENT

The authors would like to thank the associate editor and three anonymous reviewers for many insightful comments and suggestions, which have resulted in a significant improvement of the paper.

REFERENCES

[1] O. Ajanki, L. Erdős, and T. Krüger, "Quadratic vector equations on complex upper half-plane," *Memoirs Amer. Math. Soc.*, vol. 261, no. 1261, pp. 1–135, Sep. 2019.

[2] J. Alt, "Singularities of the density of states of random Gram matrices," *Electron. Commun. Probab.*, vol. 22, p. 13, Jan. 2017.

[3] J. Alt, L. Erdős, and T. Krüger, "Local law for random Gram matrices," *Electron. J. Probab.*, vol. 22, p. 41, Jan. 2017.

[4] N. Asendorf and R. R. Nadakuditi, "Improved detection of correlated signals in low-rank-plus-noise type data sets using informative canonical correlation analysis (ICCA)," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3451–3467, Jun. 2017.

[5] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, 2005.

[6] Z. Bao, X. Ding, and A. K. Wang, "Singular vector and singular subspace distribution for the matrix denoising model," *Ann. Statist.*, vol. 49, no. 1, pp. 370–392, Feb. 2021.

[7] Z. Bao, G. Pan, and W. Zhou, "Universality for the largest eigenvalue of sample covariance matrices with general population," *Ann. Statist.*, vol. 43, no. 1, pp. 382–421, Feb. 2015.

[8] A. Bensoussan, G. D. Prato, M. C. Delfour, and S. K. Mitter, "Semigroups of operators and interpolation," in *Representation and Control of Infinite Dimensional Systems*. Boston, MA, USA: Birkhäuser, 2007, pp. 87–172.

[9] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of statistical tests for single-source detection using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2400–2419, Apr. 2011.

[10] B. Alex, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, "Isotropic local laws for sample covariance and generalized Wigner matrices," *Electron. J. Probab.*, vol. 19, no. 33, pp. 1–53, Jan. 2014.

[11] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin, "On the principal components of sample covariance matrices," *Probab. Theory Rel. Fields*, vol. 164, nos. 1–2, pp. 459–552, Feb. 2016.

[12] P. Bourgade, L. Erdős, and H.-T. Yau, "Edge universality of beta ensembles," *Commun. Math. Phys.*, vol. 332, no. 1, pp. 261–353, Nov. 2014.

[13] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin, "Fixed energy universality for generalized Wigner matrices," *Commun. Pure Appl. Math.*, vol. 69, no. 10, pp. 1815–1881, Oct. 2016.

[14] P. Bourgade and H.-T. Yau, "The eigenvector moment flow and local quantum unique ergodicity," *Commun. Math. Phys.*, vol. 350, no. 1, pp. 231–278, Feb. 2017.

[15] L. D. Chamain, P. Dharmawansa, S. Atapattu, and C. Tellambura, "Eigenvalue-based detection of a signal in colored noise: Finite and asymptotic analyses," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6413–6433, Oct. 2020.

[16] G. Cipolloni, L. Erdős, T. Krüger, and D. Schröder, "Cusp universality for random matrices, II: The real symmetric case," *Pure Appl. Anal.*, vol. 1, no. 4, pp. 615–707, Oct. 2019.

[17] X. Ding, "High dimensional deformed rectangular matrices with applications in matrix denoising," *Bernoulli*, vol. 26, no. 1, pp. 387–417, Feb. 2020.

[18] X. Ding and F. Yang, "A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices," *Ann. Appl. Probab.*, vol. 28, no. 3, pp. 1679–1738, Jun. 2018.

[19] X. Ding and F. Yang, "Edge statistics of large dimensional deformed rectangular matrices," 2020, *arXiv:2009.00389*.

[20] X. Ding and F. Yang, "Spiked separable covariance matrices and principal components," *Ann. Statist.*, vol. 49, no. 2, pp. 1113–1138, Apr. 2021.

[21] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, Mar. 1995.

[22] R. B. Dozier and J. W. Silverstein, "Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices," *J. Multivariate Anal.*, vol. 98, no. 6, pp. 1099–1122, Jul. 2007.

[23] R. B. Dozier and J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices," *J. Multivariate Anal.*, vol. 98, no. 4, pp. 678–694, Apr. 2007.

[24] N. El Karoui, "Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices," *Ann. Probab.*, vol. 35, no. 2, pp. 663–714, Mar. 2007.

[25] L. Erdős, A. Knowles, and H.-T. Yau, "Averaging fluctuations in resolvents of random band matrices," *Annales Henri Poincaré*, vol. 14, no. 8, pp. 1837–1926, Dec. 2013.

[26] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, "Spectral statistics of Erdős-Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues," *Commun. Math. Phys.*, vol. 314, no. 3, pp. 587–640, Sep. 2012.

[27] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, "The local semicircle law for a general class of random matrices," *Electron. J. Probab.*, vol. 18, pp. 1–58, Jan. 2013.

[28] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, "Spectral statistics of Erdős–Rényi graphs I: Local semicircle law," *Ann. Probab.*, vol. 41, no. 3B, pp. 2279–2375, May 2013.

[29] L. Erdős, T. Krüger, and D. Schröder, "Cusp universality for random matrices I: Local law and the complex Hermitian case," *Commun. Math. Phys.*, vol. 378, no. 2, pp. 1203–1278, Sep. 2020.

[30] L. Erdős, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau, "Bulk universality for Wigner matrices," *Commun. Pure Appl. Math.*, vol. 63, no. 7, pp. 895–925, 2010.

[31] L. Erdős, B. Schlein, and H.-T. Yau, "Universality of random matrices and local relaxation flow," *Inventiones Mathematicae*, vol. 185, no. 1, pp. 75–119, Jul. 2011.

[32] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, "The local relaxation flow approach to universality of the local statistics for random matrices," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 48, no. 1, pp. 1–46, Feb. 2012.

[33] L. Erdős and H.-T. Yau, *A Dynamical Approach to Random Matrix Theory* (Courant Lecture Notes in Mathematics), vol. 28. Providence, RI, USA: AMS, 2017.

[34] L. Erdős, H.-T. Yau, and J. Yin, "Bulk universality for generalized Wigner matrices," *Probab. Theory Rel. Fields*, vol. 154, nos. 1–2, pp. 341–407, Oct. 2012.

[35] L. Erdős, H.-T. Yau, and J. Yin, "Rigidity of eigenvalues of generalized Wigner matrices," *Adv. Math.*, vol. 229, no. 3, pp. 1435–1515, Feb. 2012.

[36] L. Erdos and H.-T. Yau, "Gap universality of generalized Wigner and β-ensembles," *J. Eur. Math. Soc.*, vol. 17, no. 8, pp. 1927–2036, 2015.

[37] Z. Fan and I. M. Johnstone, "Tracy-Widom at each edge of real covariance and MANOVA estimators," 2017, *arXiv:1707.02352*.

[38] P. J. Forrester, "The spectrum edge of random matrix ensembles," *Nucl. Phys. B*, vol. 402, no. 3, pp. 709–728, Aug. 1993.

[39] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Ann. Appl. Probab.*, vol. 17, no. 3, pp. 875–930, Jun. 2007.

[40] X. Han, G. Pan, and B. Zhang, "The Tracy–Widom law for the largest eigenvalue of F type matrices," *Ann. Statist.*, vol. 44, no. 4, pp. 1564–1592, Aug. 2016.

[41] Y. He and A. Knowles, "Fluctuations of extreme eigenvalues of sparse Erdős–Rényi graphs," *Probab. Theory Rel. Fields*, vol. 180, no. 3, pp. 985–1056, Apr. 2021.

[42] J. Huang, B. Landon, and H.-T. Yau, "Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs," *Ann. Probab.*, vol. 48, no. 2, pp. 916–962, Mar. 2020.

[43] J. Y. Hwang, J. O. Lee, and K. Schnelli, "Local law and Tracy–Widom limit for sparse sample covariance matrices," *Ann. Appl. Probab.*, vol. 29, no. 5, pp. 3006–3036, Oct. 2019.

[44] J. Jin, Z. T. Ke, S. Luo, and M. Wang, "Optimal estimation of the number of network communities," *J. Amer. Stat. Assoc.*, pp. 1–16, Mar. 2022. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.2022.2035736
[45] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, 2001.
[46] I. M. Johnstone and A. Onatski, "Testing in high-dimensional spiked models," *Ann. Statist.*, vol. 48, no. 3, pp. 1231–1254, Jun. 2020.
[47] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[48] A. Knowles and J. Yin, "Anisotropic local laws for random matrices," *Probab. Theory Rel. Fields*, vol. 169, pp. 1–96, Oct. 2016.
[49] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.
[50] B. Landon, P. Sosoe, and H.-T. Yau, "Fixed energy universality of Dyson Brownian motion," *Adv. Math.*, vol. 346, pp. 1137–1332, Apr. 2019.
[51] B. Landon and H.-T. Yau, "Edge statistics of Dyson Brownian motion," 2017, *arXiv:1712.03881*.
[52] J. O. Lee and K. Schnelli, "Edge universality for deformed Wigner matrices," *Rev. Math. Phys.*, vol. 27, no. 8, Sep. 2015, Art. no. 1550018.
[53] J. O. Lee and K. Schnelli, "Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population," *Ann. Appl. Probab.*, vol. 26, no. 6, pp. 3786–3839, Dec. 2016.
[54] J. O. Lee and J. Yin, "A necessary and sufficient condition for edge universality of Wigner matrices," *Duke Math. J.*, vol. 163, no. 1, pp. 117–173, Jan. 2014.
[55] W. E. Leeb, "Matrix denoising for weighted loss functions and heterogeneous signals," *SIAM J. Math. Data Sci.*, vol. 3, no. 3, pp. 987–1012, 2021.
[56] P. Loubaton and P. Vallet, "Almost sure localization of the eigenvalues in a Gaussian information plus noise model. Application to the spiked models," *Electron. J. Probab.*, vol. 16, pp. 1934–1959, Jan. 2011.
[57] K. Lounici, "High-dimensional covariance matrix estimation with missing observations," *Bernoulli*, vol. 20, no. 3, pp. 1029–1058, Aug. 2014.
[58] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967.
[59] E. M. Marcotte, M. Pellegrini, H.-L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg, "Detecting protein function and protein-protein interactions from genome sequences," *Science*, vol. 285, no. 5428, pp. 751–753, Jul. 1999.
[60] R. R. Nadakuditi, "OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3002–3018, May 2014.
[61] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, Jul. 2008.
[62] R. R. Nadakuditi and J. W. Silverstein, "Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 468–480, Jun. 2010.
[63] A. Onatski, "The Tracy–Widom limit for the largest eigenvalues of singular complex Wishart matrices," *Ann. Appl. Probab.*, vol. 18, no. 2, pp. 470–490, Apr. 2008.
[64] A. Onatski, "Testing hypotheses about the number of factors in large factor models," *Econometrica*, vol. 77, no. 5, pp. 1447–1479, 2009.
[65] A. Onatski, M. J. Moreira, and M. Hallin, "Signal detection in high dimension: The multispiked case," *Ann. Statist.*, vol. 42, no. 1, pp. 225–254, Feb. 2014.
[66] N. S. Pillai and J. Yin, "Universality of covariance matrices," *Ann. Appl. Probab.*, vol. 24, no. 3, pp. 935–1001, 2014.
[67] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
[68] C. A. Tracy and H. Widom, "Level-spacing distributions and the Airy kernel," *Commun. Math. Phys.*, vol. 159, no. 1, pp. 151–174, 1994.
[69] C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Commun. Math. Phys.*, vol. 177, no. 3, pp. 727–754, 1996.
[70] D. W. Tufts and A. A. Shah, "Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix," *IEEE Trans. Signal Process.*, vol. 41, no. 4, pp. 1716–1721, Apr. 1993.
[71] P. Vallet, P. Loubaton, and X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1043–1068, Feb. 2012.
[72] J. Vinogradova, R. Couillet, and W. Hachem, "Statistical inference in large antenna arrays under unknown noise pattern," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5633–5645, Nov. 2013.
[73] D. Yang, Z. Ma, and A. Buja, "Rate optimal denoising of simultaneously sparse and low rank matrices," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 3163–3189, 2016.
[74] F. Yang, "Edge universality of separable covariance matrices," *Electron. J. Probab.*, vol. 24, p. 57, Jan. 2019.
[75] Z. Zhang and G. Pan, "Tracy-Widom law for the extreme eigenvalues of large signal-plus-noise matrices," 2020, *arXiv:2009.12031*.

Xiucai Ding received the Ph.D. degree in statistics from the University of Toronto in 2018. He worked as a Post-Doctoral Research Associate at Duke University from 2018 to 2020. He is currently an Assistant Professor with the Department of Statistics, University of California at Davis, Davis, CA, USA. His research interests include random matrix theory, high-dimensional statistics, time series analysis, and statistical learning theory.

Fan Yang received the Ph.D. degree in physics from The Chinese University of Hong Kong in 2014, and the Ph.D. degree in mathematics from the University of California at Los Angeles, Los Angeles, CA, USA, in 2019. He is currently a Post-Doctoral Researcher with the Department of Statistics and Data Science, University of Pennsylvania. He will be an Associate Professor with the Yau Mathematical Sciences Center, Tsinghua University, in 2022 Fall. His research interests include random matrix theory, high-dimensional statistics, statistical machine learning, and mathematical physics.