Tracy-Widom distribution for the edge eigenvalues of Gram type random matrices

Xiucai Ding *[1] and Fan Yang †[2]

[1]Department of Mathematics, Duke University
[2]Department of Statistics, University of Pennsylvania

Abstract

Large dimensional Gram type matrices are common objects in high-dimensional statistics and machine learning. In this paper, we study the limiting distribution of the edge eigenvalues for a general class of high-dimensional Gram type random matrices, including separable sample covariance matrices, sparse sample covariance matrices, bipartite stochastic block model and random Gram matrices with general variance profiles. Specifically, we prove that under (almost) sharp moment conditions and certain tractable regularity assumptions, the edge eigenvalues, i.e., the largest few eigenvalues of non-spiked Gram type random matrices or the extremal bulk eigenvalues of spiked Gram type random matrices, satisfy the Tracy-Widom distribution asymptotically.

Our results can be used to construct adaptive, accurate and powerful statistics for high-dimensional statistical inference. In particular, we propose data-dependent statistics to infer the number of signals under general noise structure, test the one-sided sphericity of separable matrix, and test the structure of bipartite stochastic block model. Numerical simulations show strong support of our proposed statistics.

The core of our proof is to establish the edge universality and Tracy-Widom distribution for a rectangular Dyson Brownian motion with regular initial data. This is a general strategy to study the edge statistics for high-dimensional Gram type random matrices without exploring the specific independence structure of the target matrices. It has potential to be applied to more general random matrices that are beyond the ones considered in this paper.

1 Introduction

Large dimensional Gram type random matrices play an important role in high-dimensional data analysis and modern statistical learning theory. Consider a $p \times n$ random matrix $Y = (y_{ij})$ whose entries are independent centered random variables with variances

$$\mathbb{E} y_{ij}^2 = s_{ij}, \tag{1.1}$$

*E-mail: xiucai.ding@duke.edu.
†E-mail: fyang75@wharton.upenn.edu

for a sequence of positive numbers s_{ij}, $1 \leqslant i \leqslant p, 1 \leqslant j \leqslant n$, depending on n. In this paper, we shall call $\mathcal{Q} = YY^\top$ a random Gram matrix with variance matrix (or profile) $S = (s_{ij})$ [4, 5]. In wireless communication [20], especially in the Multiple Input Multiple Output (MIMO) system, Y is the channel matrix such that s_{ij} represents the fluctuation of the signal from the j-th transmitter to the i-th receiver antenna.

The above Gram type matrix model covers a lot of important random matrices in high-dimensional statistics and machine learning theory. We now name a few examples. First, when $s_{ij} = (a_i b_j)/n$, where a_i, $1 \leqslant i \leqslant p$, and b_j, $1 \leqslant j \leqslant n$, are deterministic sequences of numbers satisfying certain regularity conditions, we can write $\mathcal{Q} = A^{1/2} Z B Z^\top A^{1/2}$. Here $A = \operatorname{diag}\{a_1, a_2, \cdots, a_p\}$, $B = \operatorname{diag}\{b_1, b_2, \cdots, b_n\}$, and $Z = (z_{ij})$ is a $p \times n$ random matrix whose entries z_{ij} are i.i.d. centered random variables with variance n^{-1}. In this special case, it reduces to the so-called separable covariance matrix, which finds important applications in spatio-temporal data analysis [64], wireless communications [94] and financial economics [85]. Furthermore, when $b_j = 1$, $1 \leqslant j \leqslant n$, it reduces to the standard sample covariance matrix with covariance matrix A, which is a fundamental object in multivariate and high-dimensional statistics. For a detailed review, we refer the reader to [87]. We can also extend the above setting without requiring A and B to be diagonal, that is, A and B can be general positive definite symmetric covariance matrices. Second, if y_{ij} are independent Bernoulli random variables with probabilities $0 < \mathtt{p}_{ij} < 1$, $1 \leqslant i \leqslant p, 1 \leqslant j \leqslant n$, then Y is closely related to the biadjacency matrix of a bipartite stochastic block model (c.f. Definition 2.11). It is a generalization of the stochastic block model (SBM) [1] and has been an important object of study in machine learning theory [45, 46, 82]. In this case, we can write $Y = \mathring{Y} + \mathbb{E}Y$, where $\mathring{Y}\mathring{Y}^\top$ is a random Gram matrix as defined around (1.1). Third, consider $y_{ij} = a_{ij} h_{ij}$, where a_{ij} are i.i.d. Bernoulli random variables with parameter $0 < \mathtt{p} \ll 1$ depending on n, and h_{ij} are independent centered random variables with variances (1.1). In this case, $\mathcal{Q} = YY^\top$ is called a sparse random Gram matrix, where $\mathtt{p}$ characterizes the level of sparsity [54]. This model provides important insights into covariance and precision matrices estimation with randomly missing observations [60, 75].

In this paper, we study the limiting distribution of edge eigenvalues of random Gram matrices with general variance matrices. Here the notion *edge eigenvalue* refers to the extremal eigenvalues of the bulk eigenvalue spectrum of $\mathcal{Q}$. More precisely, if the variance matrix S is non-spiked so that $\mathcal{Q}$ has no outliers (i.e. eigenvalues that are detached from the bulk eigenvalue spectrum), then the edge eigenvalues are the largest few eigenvalues of $\mathcal{Q}$. On the other hand, if the singular value decomposition (SVD) of S contains a finite number of spikes, the edge eigenvalues are the largest non-outlier eigenvalues. We refer the readers to Remark 2.4 below for a more detailed discussion. One of the main goals of this paper is to prove that the edge eigenvalues of a general class of Gram random matrices obey the Tracy-Widom law [89, 90] asymptotically when $n \to \infty$, assuming (almost) sharp moment assumptions and certain tractable regularity conditions.

The edge eigenvalues of random matrices have been applied to various hypothesis testing problems in high-dimensional statistics and machine learning theory. For instance, they have been employed to test the existence and number of spikes for sample covariance matrices [57, 86], test the number of factors in factor model [84], detect the signals in signal-plus-noise model [6, 8, 12, 93], test the structure of covariance matrices [29, 50], and perform the multivariate analysis of variance (MANOVA) [44, 50]. More recently, they have also been used to help to understand the deep learning representations and deep neural networks [28, 48]. The Tracy-Widom law for the largest eigenvalues of non-spiked Gram type random matrices have been proved in many random matrix models. For Wishart matrices with $s_{ij} \equiv n^{-1}$ and i.i.d. centered Gaussian entries y_{ij}, it was proved in [56] that the largest eigenvalue satisfies the Tracy-Widom distribution asymptotically. This result was later extended to general sample random matrices with generally distributed entries (assuming only certain moment assumptions) and variances $s_{ij} = a_i/n$ in a series of papers; see e.g. [8, 22, 29, 59, 71, 83, 88].

Despite the broad application of Gram type random matrices and their edge eigenvalues in statistics and machine learning theory, the literature has been focused on the sample covariance type of variance profile $s_{ij} = a_i/n$. Unfortunately, for more complicated variance profiles, much less is known about the limiting distribution of the edge eigenvalues. In fact, even for the separable covariance matrix with $s_{ij} = (a_i b_j)/n$, the Tracy-Widom law has not been proved yet for the edge eigenvalues as explained in [98]. One purpose of this paper is to solve this problem and prove the Tracy-Widom law for a general class of random Gram matrices (with or without outliers) with variance profiles $S = (s_{ij})$ satisfying certain regularity assumptions. This generality, on one hand, allows us to study more general and realistic random Gram matrices arising from statistical learning theory, such as the separable covariance matrices, bipartite stochastic block model and sparse random Gram matrices with general variance profiles, and apply our result on edge eigenvalues to various hypothesis testing problems in these models. On the other hand, it provides a way to study the extremal non-outlier eigenvalues if the matrices have a few outlier eigenvalues. In particular, we can use our result to test the number of signals instead of only detecting the existence of a signal.

For a random Gram matrix $\mathcal{Q}$ with general variance profile S, its limiting spectral distribution has been studied in [4, 5, 49] via the Stieltjes transform method. It turns out that the Stieltjes transform of the limiting ESD is an average of a sequence of Stieltjes transforms which are the solutions to the so-called vector Dyson equation (c.f. equation (2.14)). Moreover, the local laws on the resolvent of $\mathcal{Q}$ and the rigidity of the eigenvalues—both of them measure the closeness between the ESD of $\mathcal{Q}$ and the limiting ESD—have been established [4, 5]. Based on these inputs, we shall prove the Tracy-Widom distribution for the edge eigenvalues of $\mathcal{Q}$. More precisely, our proof employs the following three step strategy.

Step 1: Local laws. Establishing a local law on the Stieltjes transform of $\mathcal{Q}$, $m_{\mathcal{Q}}(z) := p^{-1}\operatorname{Tr}(\mathcal{Q} - z)^{-1}$. This is needed in order to check the regularity conditions in Definition 2.1 below.

Step 2: Rectangular DBM. Proving the Tracy-Widom distribution for the edge eigenvalues of the Gaussian divisible ensemble $\mathcal{Q}_t := (Y + \sqrt{t}X)(Y + \sqrt{t}X)^\top$, where X is an independent $p \times n$ random matrix whose entries are i.i.d. centered Gaussian random variables with variance n^{-1}. Following the literature, we shall call the evolution of $\mathcal{Q}_t$ with respect to t the rectangular *matrix* Dyson Brownian motion (DBM), while we shall call the evolution of the *eigenvalues* of $\mathcal{Q}_t$ with respect to t the rectangular Dyson Brownian motion.

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

Usually Step 1 depends on the specific random matrix model, and requires a case by case study. Roughly speaking, the purpose of this step is to show that the spectral density 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. (For example, if the spectral density has a regular cubic root behavior, then the corresponding cusp universality is different from the Tracy-Widom law [19, 36].) In our setting of random Gram matrices, this step has been completed in [4, 5]. On the other hand, for Step 3 systematic methods have been developed based on moments matching and resolvent comparison arguments [42, 59, 73, 98], and they work universally for many random matrix ensembles. In this paper, we mainly focus on Step 2, that is, we shall prove that the edge statistics of a rectangular DBM converges to Tracy-Widom as long as we have a regular initial data as described by Definition 2.1. This fills in the last piece of the three step strategy.

The three step strategy has been widely used in the proof of bulk universality of random matrices [37, 38, 39, 41]; for a more extensive review, we refer the reader to [40] and references therein. However, it has been rarely (if any) used in the proof of the Tracy-Widom distribution for the edge eigenvalues of Gram type random matrices. One of the main reasons is that near the spectral edge, one can usually go to Step 3 directly and show that the edge statistics of a random matrix ensemble match its Gaussian counterpart using a simple moment matching condition [22, 42, 59, 73, 88, 98]. If it has been proved that the edge eigenvalues of the corresponding Gaussian matrix satisfies the Tracy-Widom law (as in the Wishart case [56] or the sample covariance case [29, 83]), then the above comparison argument immediately shows that the edge eigenvalues of the original random matrix ensemble also satisfy the Tracy-Widom law. Unfortunately, the Tracy-Widom law for the corresponding Gaussian matrices are not always known. For example, in [98] the second author was able to prove that a general separable covariance matrix has the same edge statistics as a Gaussian separable covariance matrix, whose Tracy-Widom law, however, is still unknown in the literature. Moreover, the conventional way of proving Tracy-Widom law for Gaussian ensembles usually involves some exactly solvable formulas derived from the joint eigenvalue distributions, which are not available for a general Gaussian Gram matrix. On the other hand, our result provides an alternative way to approach this kind of problem. More precisely, for a Gaussian Gram matrix Y, we can decompose it into

$$Y \stackrel{d}{=} Y_t + \sqrt{t}X,$$

where Y_t and X are independent Gaussian matrices. (Here "$\stackrel{d}{=}$" means equal in distribution.) Then our main Theorem 2.3 shows that the edge eigenvalues of Y satisfy the Tracy-Widom law for $t \gg n^{-1/3}$. We also remark that there is another way to show the Tracy-Widom law for general Gaussian Gram ensembles using a delicate continuous comparison argument [44, 70, 71]. However this kind of argument depends greatly on the specific variance profile, and hence requires a case by case study.

Finally, we remark that our main result, Theorem 2.3, is an extension of the one in [67] for the edge statistics of symmetric DBM for Wigner type matrices. Moreover, the result in [67] has been successfully used in the proof of the Tracy-Widom fluctuation for Wigner type random matrices with correlated entries, such as the random d-regular graph [10] and the correlated Wigner matrices [2]. Correspondingly, our result can be also applied to Gram type random matrices with correlated entries, including the random bipartite biregular graph, random regular hypergraphs, or correlated random Gram matrices. However, in these cases we need extra work on Steps 1 and 3 of the three step strategy, and these are possible topics for future study.

Before concluding this introduction, we summarize the main contributions of our work.

- We establish the Tracy-Widom distribution for edge eigenvalues of a general class of Gram type random matrices, which may or may not contain outliers. As important examples, we study the edge eigenvalues of separable covariance matrices, bipartite stochastic block model and sparse random Gram matrices with general variance profiles. These random matrix models are important in high-dimensional statistics, yet their edge fluctuations are not known rigorously in the previous literature.

- We propose a general three step strategy to study the edge statistics of Gram type random matrices. In particular, we establish a general result on the Tracy-Widom distribution for the edge eigenvalues of a rectangular DBM, which completes Step 2 of the three step strategy once for all. Both the strategy and the result on rectangular DBM have potential to be applied to other important and more complicated random matrices in statistics .

- We apply our results to three important hypothesis testing problems: testing the number of signals with general variance structure for the noise (e.g., the doubly-heteroscedastic colored noise and sparse noise), testing the one-sided sphericity of separable covariance matrix, and testing the structure of bipartite stochastic block model.

The rest of this paper is organized as follows. In Section 2, we state the main results: Theorem 2.3 for the edge statistics of rectangular DBM, Theorem 2.7 for a general class of random Gram matrices, Theorem 2.12 for the bipartite stochastic block model, and Theorem 2.15 for the sparse random Gram matrices with general variance profiles. In Section 3, we apply our results to propose some useful statistics for three hypothesis testing problems. In Section A, we prove Theorems 2.7, 2.12 and 2.15. The Sections B and C are devoted to the proof of Theorem 2.3: Section B collects some important estimates for rectangular DBM, whose proofs are given in another paper [23], and Section C includes the core analysis of the rectangular Dyson Brownian motion based on the estimates in Section B.

Convention. The fundamental large parameter is n and we always assume that p is comparable to and depends on n. We use C to denote a generic large positive constant, whose value may change from one line to the next. Similarly, we use ε, τ, δ, etc. to denote generic small positive constants. If a constant depends on a quantity a, we use $C(a)$ or C_a to indicate this dependence. For two quantities a_n and b_n depending on n, the notation $a_n = \mathrm{O}(b_n)$ means that $|a_n| \leqslant C|b_n|$ for some constant $C > 0$, and $a_n = \mathrm{o}(b_n)$ means that $|a_n| \leqslant c_n|b_n|$ for some positive sequence $c_n \downarrow 0$ as $n \to \infty$. We also use the notations $a_n \lesssim b_n$ if $a_n = \mathrm{O}(b_n)$, and $a_n \sim b_n$ if $a_n = \mathrm{O}(b_n)$ and $b_n = \mathrm{O}(a_n)$. For a matrix A, we use $\|A\| := \|A\|_{l^2 \to l^2}$ to denote the operator norm; for a vector $\mathbf{v} = (v_i)_{i=1}^n$, $\|\mathbf{v}\| \equiv \|\mathbf{v}\|_2$ stands for the Euclidean norm. For a matrix A and a positive number a, we write $A = \mathrm{O}(a)$ if $\|A\| = \mathrm{O}(a)$. In this paper, we often write an identity matrix of any dimension as I or 1 without causing any confusions.

2 Main results

In this section, we state the main results of this paper regarding the edge eigenvalues of Gram type random matrices. First, we state a general result on the Tracy-Widom distribution for the edge eigenvalues of a rectangular DBM with regular initial data. It is the central result of this paper, and can be used as a general tool to study the edge statistics of Gram type random matrices. To illustrate the power of this result, in Sections 2.2 and 2.3 we consider its applications to three different models—the random Gram matrices which includes separable covariance matrix as a special example, the bipartite stochastic block model, and the sparse random Gram matrices. For all these models, their Tracy-Widom (TW) laws for the edge eigenvalues are still unknown in the literature.

2.1 Edge statistics of rectangular Dyson Brownian motion

The edge statistics of the symmetric Dyson Brownian motion has been studied in [67] for Wigner type matrix ensembles. In this section, we extend the result there to Gram type matrix ensembles. Let Y be a $p \times n$ data matrix, and X be a $p \times n$ 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

$$Y + \sqrt{t}X \stackrel{d}{=} W + \sqrt{t}X,$$

where W is a $p \times n$ rectangular diagonal matrix,

$$W = \begin{pmatrix} D & 0 \end{pmatrix}, \quad D^2 = \mathrm{diag}(d_1, \cdots, d_p),$$

with $\sqrt{d_1} \geqslant \sqrt{d_2} \geqslant \cdots \geqslant \sqrt{d_p} > 0$ being the singular values of Y. Thus without loss of generality, we can assume that the initial data matrix is W in the following discussion. In this paper, we consider the high-dimension setting, where the aspect ratio $c_n := p/n$ converges to a constant $c \in (0, 1)$. Since the Gram matrices $(W+\sqrt{t}X)(W+\sqrt{t}X)^\top$ and $(W+\sqrt{t}X)^\top(W+\sqrt{t}X)$ have the same nonzero eigenvalues, without loss of generality we can assume that

$$\tau \leqslant c_n \leqslant 1 \tag{2.1}$$

for some constant $\tau > 0$.

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 TW law. Following [67], we state the regularity conditions in terms of the Stieltjes transform of V,

$$m_V(z) := \frac{1}{p}\operatorname{Tr}(V-z)^{-1} = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{d_i - z}, \quad z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}.$$

Definition 2.1 (η_*-regular)**.** *Let η_* be a parameter satisfying $\eta_* := N^{-\phi_*}$ for some constant $0 < \phi_* \leqslant 2/3$. We say V (or equivalently W) is η_*-regular around the largest eigenvalue if the following properties hold for $\lambda_+ := d_1$ (here λ_+ is a standard notation for the right spectral edge in random matrix literature) and some constants $c_V, C_V > 0$.*

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

$$\frac{1}{C_V}\sqrt{|\lambda_+ - E| + \eta} \leqslant \operatorname{Im} m_V(z) \leqslant C_V\sqrt{|\lambda_+ - E| + \eta}, \tag{2.2}$$

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

$$\frac{1}{C_V}\frac{\eta}{\sqrt{|\lambda_+ - E| + \eta}} \leqslant \operatorname{Im} m_V(z) \leqslant C_V\frac{\eta}{\sqrt{|\lambda_+ - E| + \eta}}. \tag{2.3}$$

(ii) We have $2c_V \leqslant \lambda_+ \leqslant C_V/2$.

(iii) We have $\|V\| \leqslant N^{C_V}$.

Remark 2.2. The motivation for item (i) is as follows: if $m(z)$ is the Stieltjes transform of a density ρ with square root behavior around λ_+, $\rho(x) \sim \sqrt{(\lambda_+ - x)_+}$, then (2.2) and (2.3) holds for $\operatorname{Im} m(z)$ with $\eta_* = 0$. For a general $\eta_* > 0$, (2.2) and (2.3) 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 $\eta \leqslant 10$ in the assumption is purely for definiteness of presentation—we can replace 10 with any constant of order 1.

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 for $\mathcal{Q}_t$, and let $m_{w,t}$ be the corresponding Stieltjes transform. It is known that for any $t > 0$, $m_{w,t}$ is the unique solution to

$$m_{w,t} = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{d_i(1+c_n t m_{w,t})^{-1} - (1+c_n t m_{w,t})z + t(1-c_n)}, \tag{2.4}$$

such that $\operatorname{Im} m_{w,t} > 0$ for $z \in \mathbb{C}_+$ [26, 27, 92]. Adopting the notations from free probability theory, we shall call $\rho_{w,t}$ the rectangular free convolution of $\rho_{w,0}$ with Marchenko-Pastur (MP) law at time t. Let $\lambda_{+,t}$ be the rightmost edge of $\rho_{w,t}$. By Lemma B.5 in appendix we know that $\rho_{w,t}$ has a square root behavior near $\lambda_{+,t}$.

We introduce the notation

$$\zeta_t(z) := [1 + c_n t m_{w,t}(z)]^2 z - (1 - c_n)t[1 + c_n t m_{w,t}(z)], \tag{2.5}$$

which is in fact the subordination function for the rectangular free convolution with Marchenko-Pastur (MP) law. Then we define the function

$$\Phi_t(\zeta) = \zeta(1 - c_n t m_{w,0}(\zeta))^2 + (1 - c_n)t(1 - c_n t m_{w,0}(\zeta)), \tag{2.6}$$

and the parameter (need to check this parameter)

$$\gamma_n \equiv \gamma_n(t) = \left(\frac{1}{2}[4\lambda_{+,t}\xi_{+,t} + (1 - c_n)^2 t^2] c_n^2 t^2 \Phi_t''(\zeta_{+,t})\right)^{-1/3}, \tag{2.7}$$

where $\zeta_{+,t} := \zeta(\lambda_{+,t})$. Here we used the short-hand notation $\zeta(\lambda_{+,t}) \equiv \lim_{\eta\downarrow 0} \zeta(\lambda_{+,t} + \mathrm{i}\eta)$.

We now state the central result of this paper for the edge statistics of rectangular DBM. Recall that Gaussian orthogonal ensemble (GOE) refers to symmetric random matrices of the form $H := (Z + Z^\top)/\sqrt{2p}$, where Z is a $p \times p$ matrix with i.i.d. real centered Gaussian entries with variance one. We denote the eigenvalues of H by $\mu_1^g \geqslant \mu_2^g \geqslant \cdots \mu_p^g$ and those of $\mathcal{Q}_t$ by $\lambda_1(t) \geqslant \lambda_2(t) \geqslant \cdots \geqslant \lambda_p(t)$.

Theorem 2.3. *Suppose W is η_*-regular in the sense of Definition 2.1. Suppsoe t satisfies $n^\varepsilon \eta_* \leqslant t^2 \leqslant n^{-\varepsilon}$ for a small constant $\varepsilon > 0$. Fix any integer $\mathfrak{a} \in \mathbb{N}$, and let $f : \mathbb{R}^{\mathfrak{a}} \to \mathbb{R}$ be any test function such that*

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

for some constant $C > 0$. Then for γ_n in (2.7), we have that

$$\left|\mathbb{E} f(\gamma_n p^{2/3}(\lambda_1(t) - \lambda_{+,t}), \cdots, \gamma_n p^{2/3}(\lambda_{\mathfrak{a}}(t) - \lambda_{+,t})) - \mathbb{E} f(p^{2/3}(\mu_1^g - 2), \cdots, p^{2/3}(\mu_{\mathfrak{a}}^g - 2))\right| \leqslant n^{-\mathfrak{c}}, \tag{2.8}$$

for some constant $\mathfrak{c} > 0$.

Since the edge eigenvalues of GOE at ± 2 obeys the type-1 TW fluctuation [89, 90], by Theorem 2.3 and Portmanteau lemma we immediately obtain that

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

where F_1 is the type-1 TW distribution function.

Remark 2.4. In this paper, we focus on the largest few eigenvalues of Gram type random matrices, which are usually the most interesting case from the statistical view of point. However, as in [67] we can also extend our result to the case where the edge eigenvalue is d_{i_0} for some $i_0 > 1$. Specifically, assume that W is η_*-regular in the sense of Definition 2.1 with $\lambda_+ = d_{i_0}$ and that there are no eigenvalues d_i inside the interval $[\lambda_+ + \eta_*, \lambda_+ + c_V]$. Then Theorem 2.3 still holds by replacing (2.8) with

$$\left|\mathbb{E}\Big[f(\gamma_n p^{2/3}(\lambda_{i_0}(t) - \lambda_{+,t}), \cdots, \gamma_n p^{2/3}(\lambda_{i_0+\mathfrak{a}-1}(t) - \lambda_{+,t}))\Big] - \mathbb{E}\left[f(p^{2/3}(\mu_1^g - 2), \cdots, p^{2/3}(\mu_{\mathfrak{a}}^g - 2))\right]\right| \leqslant n^{-\mathfrak{c}}. \tag{2.9}$$

There are several scenarios where this result can be useful. For example, one may have a random matrix model with a few outliers away from the bulk eigenvalue spectrum as in Johnstone's spiked model [56] or the spiked separable covariance model [24]. Also for some models, the eigenvalue spectrum may have more than one disjoint bulk components, and i_0 is the edge of a sub-leading component. The proof of (2.9) is the same as the one for Theorem 2.3, except for an extra complication regarding the reindexing of the eigenvalues; see (3.10)-(3.12) of [67].

2.2 Tracy-Widom law for random Gram matrices

As the first application of Theorem 2.3, we prove the Tracy-Widom law for the edge eigenvalues of a general class of random Gram matrices with heterogeneous variance profiles. Let $Y = (y_{ij})$ be a $p \times n$ matrix whose entries are independent random variables with

$$Ey_{ij} = 0, \quad \mathbb{E}y_{ij}^2 = s_{ij}, \tag{2.10}$$

for some variance matrix $S = (s_{ij})$. Then we shall call $\mathcal{Q} = YY^\top$ a random Gram matrix. Following [4, 5], we make the following regularity assumptions on S.

Assumption 2.5. *Suppose the following conditions hold true.*

(A1) *The dimensions of S are comparable. Without loss of generality we assume that (2.1) holds.*

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

$$\max_{i,j} s_{ij} \leqslant \frac{s_*}{n}, \quad \min_{i,j} s_{ij} \geqslant \frac{n^{-1/3+\varepsilon_*}}{n}. \tag{2.11}$$

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

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

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

$$\Gamma(\varepsilon) \leqslant \min\left\{\inf_{1\leqslant i\leqslant p}\frac{1}{p}\sum_l \frac{1}{\varepsilon + n\|S_i - S_l\|_2^2}, \inf_{1\leqslant j\leqslant n}\frac{1}{n}\sum_l \frac{1}{\varepsilon + n\|S_j^\top - S_l^\top\|_2^2}\right\}, \tag{2.12}$$

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

As explained in (2.22) of [3], assumption (A4) aims to rule out outliers. We refer the reader to Remark 2.10 below for a more detailed explanation in the separable covariance model. We point out that [4, Remark 2.4] provides an easier way to check sufficient (but not necessary) condition for **(A4)**.

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

$$\|S_{i_1} - S_{i_2}\|_2 \leqslant \tau^{-1}\frac{|i_1 - i_2|^{1/2}}{n} \text{ for } i_1, i_2 \in I_\alpha, \quad \|S_{j_1}^\top - S_{j_2}^\top\|_2 \leqslant \tau^{-1}\frac{|j_1 - j_2|^{1/2}}{n} \text{ for } j_1, j_2 \in J_\beta, \tag{2.13}$$

for some small constant $\tau > 0$. The proof of (A4) using (A4-s) is simple using an integral approximate.

If (2.11) holds, then there is 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

$$\frac{1}{\mathbf{m}} = -z\mathbf{1} + S\frac{1}{\mathbf{1} + S^\top\mathbf{m}}, \tag{2.14}$$

such that Im $m_k(z) > 0$, $k = 1, \cdots, p$, for any $z \in \mathbb{C}_+$ [4, 5, 49]. 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 reciprocal. Moreover, if Assumption 2.5 holds, then for each $k = 1, \cdots, p$, there exists a unique probability measure ν_k with support contained in $[0, 4s_*]$, and that is absolutely continuous with respect to the Lebesgue measure. (If we consider the case $p > n$, then ν_k will also have a point mass at zero.) Let ρ_k be the density function associated with ν_k, and m_k be the Stieltjes transform of ν_k:

$$m_k(z) := \int \frac{\nu_k(\mathrm{d}x)}{x - z}, \quad z \in \mathbb{C}_+, \quad k = 1, 2, \cdots, p.$$

Then the asymptotic ESD of $\mathcal{Q} = YY^\top$ is given by $\nu := p^{-1} \sum_k \nu_k$, with the following density and Stieltjes transform:

$$\rho := \frac{1}{p} \sum_k \rho_k, \quad m(z) := \frac{1}{p} \sum_k m_k(z). \tag{2.15}$$

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

Lemma 2.6 (Theorem 2.3 of [4])**.** *Under Assumption 2.5, for any $1 \leqslant k \leqslant p$ there exists a sequence of positive numbers $a_1 > a_2 \cdots > a_{2q} \geqslant 0$ such that*

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

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

$$\rho(a_1 - x) = \pi^{-1} \varpi \sqrt{x} + \mathrm{O}(x), \quad x \downarrow 0, \tag{2.16}$$

where $\varpi > 0$ is some constant determined by S.

We shall call a_k the spectral edges. In particular, we will focus on the right-most edge a_1 and denote $\lambda_+ := a_1$. We now state a necessary and sufficient condition for the Tracy-Widom distribution of the largest eigenvalues of the random Gram matrix $\mathcal{Q}$ with variance matrix S.

Theorem 2.7. *Suppose $Y = (y_{ij})$ is a $p \times n$ random matrix such that $\widehat{y}_{ij} := y_{ij}/\sqrt{s_{ij}}$ are real i.i.d. random variables satisfying $\mathbb{E}\widehat{y}_{11} = 0$, $\mathbb{E}\widehat{y}_{11}^2 = 1$ and*

$$\lim_{x \to \infty} x^4 \mathbb{P}\left(|\widehat{y}_{11}| \geqslant x\right) = 0. \tag{2.17}$$

Suppose the variance matrix $S = (s_{ij})$ satisfies Assumptions 2.5. Denote the eigenvalues of $\mathcal{Q} = YY^\top$ by $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$. Then we have

$$\lim_{n \to \infty} \mathbb{P}(\varpi^{2/3} p^{2/3} (\lambda_1 - \lambda_+) \leqslant x) = F_1(x), \quad \text{for all } \ x \in \mathbb{R}. \tag{2.18}$$

On the other hand, if (2.17) does not hold, then for any fixed $x > \lambda_+$ we have

$$\limsup_{n \to \infty} \mathbb{P}(\lambda_1 > x) > 0. \tag{2.19}$$

Remark 2.8. With Theorem 2.3, it is easy to extend (2.18) to the joint distribution of the largest k eigenvalues for any fixed $k \in \mathbb{N}$:

$$\lim_{n\to\infty}\left[\mathbb{P}\left(\left(\varpi^{2/3}p^{2/3}(\lambda_i-\lambda_+)\leqslant x_i\right)_{1\leqslant i\leqslant k}\right)-\mathbb{P}\left(\left(p^{2/3}(\mu_i^g-\lambda_+)\leqslant x_i\right)_{1\leqslant i\leqslant k}\right)\right]=0, \tag{2.20}$$

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

As an important example of Theorem 2.7, we now show that the edge eigenvalues of separable covariance matrices obey the Tracy-Widom fluctuation. More precisely, we consider the following separable covariance matrix

$$\mathcal{Q}_s := A^{1/2}ZBZ^\top A^{1/2}, \tag{2.21}$$

where A and B are deterministic positive definite symmetric matrices, and $Z = (z_{ij})$ is a $p \times n$ random matrix with $z_{ij} = n^{-1/2}\widehat{z}_{ij}$ for a sequence of i.i.d. random variables $\widehat{z}_{ij}$ with mean zero and variance one. In [98], the edge universality has been proved in the sense that the edge statistics of $\mathcal{Q}_s$ match those of the Gaussian case where the entries of Z are i.i.d. Gaussian. However, the TW law was not proved yet, because the TW law for the corresponding Gaussian case has not been proved in the literature. Now we can solve this problem easily using Theorem 2.7.

Suppose A and B have eigendecompositions

$$A = U\Sigma_a U^\top, \quad B = V\Sigma_b V^\top, \quad \Sigma_a = \mathrm{diag}(a_1,\ldots,a_p), \quad \Sigma_b = \mathrm{diag}(b_1,\ldots,b_n), \tag{2.22}$$

where $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_p > 0$ and $b_1 \geqslant b_2 \geqslant \cdots \geqslant b_n > 0$ are the eigenvalues of A and B, respectively. Then for a Gaussian Z, using the rotational invariance of multivariate Gaussian distribution, we get

$$\mathcal{Q}_s \stackrel{d}{=} U\left(YY^\top\right)U^\top, \quad Y := \Sigma_a^{1/2}Z\Sigma_b^{1/2}.$$

Note that $YY^\top$ is now a random Gram matrix as in Theorem 2.7 with variance matrix $S = ((a_ib_j)/n)$, and its eigenvalues have the same distribution as those of $\mathcal{Q}_s$. Then with this variance profile S, we can define the asymptotic spectral density ρ and the right edge λ_+.

Corollary 2.9. *Consider separable covariance matrices* (2.21)*, where $Z = (z_{ij})$ is a $p \times n$ random matrix with $z_{ij} = n^{-1/2}\widehat{z}_{ij}$ for a sequence of i.i.d. random variables $\widehat{z}_{ij}$ satisfying $\mathbb{E}\widehat{z}_{11} = 0$, $\mathbb{E}\widehat{z}_{11}^2 = 1$ and* (2.17)*. Assume that* (2.1) *holds, and that*

$$\tau \leqslant a_p \leqslant a_1 \leqslant \tau^{-1}, \quad \tau \leqslant b_n \leqslant b_1 \leqslant \tau^{-1}, \tag{2.23}$$

for some constant $\tau > 0$. Suppose $S = ((a_ib_j)/n)$ satisfies **(A4)** *of Assumption 2.5. In addition, assume that*

$$\mathbb{E}(\widehat{z}_{11}^3) = 0. \tag{2.24}$$

Denote the eigenvalues of $\mathcal{Q}_s$ by $\lambda_1^s \geqslant \lambda_2^s \geqslant \cdots \geqslant \lambda_p^s$. Then we have

$$\lim_{n\to\infty}\mathbb{P}(\varpi^{2/3}p^{2/3}(\lambda_1^s-\lambda_+^s)\leqslant x) = F_1(x), \quad \textit{for all } \ x\in\mathbb{R}, \tag{2.25}$$

where ϖ is the factor given in (2.16)*. Moreover, the condition* (2.24) *is not necessary if A or B is diagonal.*

We mention that when $B = I$, Corollary 2.9 gives the Tracy-Widom distribution for the largest eigenvalues of sample covariance matrices as obtained in [8, 22, 29, 56, 71, 83, 88].

Proof of Corollary 2.9. In [98], the following universality estimate was proved under the assumptions of this corollary:

$$\lim_{N\to\infty}\Big[\mathbb{P}(p^{2/3}(\lambda_1-\lambda_+)\leqslant x)-\mathbb{P}^G(p^{2/3}(\lambda_1-\lambda_+)\leqslant x)\Big]=0$$

for all $x\in\mathbb{R}$, where $\mathbb{P}^G$ denotes the law for $Z=(z_{ij})$ with i.i.d. Gaussian entries with mean zero and variance n^{-1}. In particular, the condition (2.24) is not necessary if A or B is diagonal. As discussed above, for Gaussian Z we can reduce the problem to the study of the largest eigenvalues of a random Gram matrix $YY^\top$ with variance matrix $S=((a_ib_j)/n)$. Note that (2.23) is stronger than (2.11) and is equivalent to (A3) of Assumption 2.5. Hence $YY^\top$ satisfies the assumptions of Theorem 2.7, which immediately concludes the proof of this corollary. □

Remark 2.10. For $S=((a_ib_j)/n)$, the condition (2.12) can be reformulated as

$$\Gamma(\varepsilon)\leqslant\min\left\{\inf_{1\leqslant i\leqslant p}\frac{1}{p}\sum_l\frac{1}{\varepsilon+|a_i-a_l|^2},\inf_{1\leqslant j\leqslant n}\frac{1}{n}\sum_l\frac{1}{\varepsilon+|b_j-b_l|^2}\right\}. \tag{2.26}$$

Moreover, (2.13) gives an easier to check sufficient condition

$$|a_{i_1}-a_{i_2}|\leqslant\tau^{-1}\frac{|i_1-i_2|^{1/2}}{n^{1/2}}\ \text{for } i_1,i_2\in I_\alpha,\quad |b_{j_1}-b_{j_2}|\leqslant\tau^{-1}\frac{|j_1-j_2|^{1/2}}{n^{1/2}}\ \text{for } j_1,j_2\in J_\beta. \tag{2.27}$$

Hence if $a_i=f(i/p)$ and $b_j=g(j/n)$ for some piecewise 1/2-Hölder continuous functions f and g, then (2.27) 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 O(1) many different values.

On the other hand, suppose there are some spikes in the spectrum of A or B: $a_1\geqslant a_2\geqslant\cdots\geqslant a_r\geqslant a_{r+1}+\tau$ and $b_1\geqslant b_2\geqslant\cdots\geqslant b_s\geqslant b_{s+1}+\tau$ for some fixed integers r,s and constant $\tau>0$. Then it is easy to check that

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

and (2.26) cannot hold for all n. Hence condition (2.26) rules out the existence of outliers.

However, (2.26) sometimes is too strong because it does not allow any spikes or isolated eigenvalues in the spectrum of A or B. Here by an isolated eigenvalue of A, we mean an a_i such that $a_{i+1}+\tau\leqslant a_i\leqslant a_{i-1}-\tau$ for some $1\leqslant i\leqslant p$ and constant $\tau>0$; we have a similar definition for an isolated eigenvalue of B. On the other hand, in [24] 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 [24, 98]. For $\mathbf{m}(z)$ in (2.14), we define another two holomorphic functions

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

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

$$1+m_{1c}(\lambda_+)b_1\geqslant\tau,\quad 1+m_{2c}(\lambda_+)a_1\geqslant\tau. \tag{2.28}$$

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 (2.23) can be relaxed to some extent. Compared with (2.26) and (2.27), (2.28) is less explicit and harder to check, but it also appears more often in random matrix literature.

2.3 Tracy-Widom law for bipartite stochastic block model

As the second application of Theorem 2.3, we prove the Tracy-Widom law for the edge eigenvalues of a bipartite stochastic block model (SBM) [45, 46].

Definition 2.11 (Bipartite stochastic block model)**.** *Fix two parameters $\delta \in [1,2]$ and $\mathtt{p} \in (0,1/2]$. We take two vertex sets V_1 and V_2 with $|V_1| = p$ and $|V_2| = n$, and assign to each vertex in V_1 and V_2 a label "+" or "−" independently with probability 1/2. We denote by $\sigma_a : V_a \to \{\pm\}$ the label function of the vertices in V_a for $a = 1, 2$. Moreover, the edges between V_1 and V_2 are added independently as follows: for $i \in V_1$ and $j \in V_2$ such that $\sigma_1(i) = \sigma_2(j)$, there is an edge between them with probability $\delta\mathtt{p}$; for $i \in V_1$ and $j \in V_2$ such that $\sigma_1(i) \neq \sigma_2(j)$, there is an edge between them with probability $(2-\delta)\mathtt{p}$.*

In the null case with $\delta = 1$, the above bipartite SBM reduces to the bipartite Erdős-Rényi graph. Here we have defined the bipartite SBM with two communities by allowing for two types of labels. It can be generalized to the case with multiple communities; see e.g. [82].

For the bipartite SBM, on one hand, people are interested in proposing efficient spectral algorithms to recover the partition of the vertex sets. On the other hand, people also want to have a knowledge of the smallest edge probability $\mathtt{p}$ that guarantees the existence of such an algorithm. In practice, such connecting probability $\mathtt{p}$ may tend to zero when the dimensions p and n become very large. In the null case $\delta = 1$, none of the spectral algorithms will be able to detect the partitions no matter how large $\mathtt{p}$ is. This can be understood in the following two ways. Heuristically, when $\delta = 1$, it is indistinguishable between the labels "+" and "−" because all the edges are added with the same probability regardless of the relation between $\sigma_1(i)$ and $\sigma_2(j)$. Theoretically, as discussed in [46], a sufficient lower bound for $\mathtt{p}$ to assure recovery is of order $((\delta-1)^2\sqrt{pn})^{-1}$, which will be invalid when δ tends one. From a statistician's perspective, given a bipartite SBM we need to test whether $\delta = 1$ before implementing any possibly useful algorithms.

Given a bipartite SBM in Definition 2.11, the $p \times n$ biadjacency matrix is given by $Z = (z_{ij} : i \in V_1, j \in V_2)$, where z_{ij} are independent Bernoulli random variables with

$$\mathbb{P}(z_{ij} = 1) = a_{ij}, \quad a_{ij} := \begin{cases} \delta\mathtt{p}, & \text{if } \sigma_1(i) = \sigma_2(j), \\ (2-\delta)\mathtt{p}, & \text{if } \sigma_1(i) \neq \sigma_2(j). \end{cases} \tag{2.29}$$

We are interested in the case where $\delta \in [1,2)$ is fixed, while $\mathtt{p} \in (0,1/2]$ can be either fixed or tend to 0 as $n \to \infty$, which characterizes the sparsity of the model. Furthermore, we always assume that $n\mathtt{p} \gg 1$ such that the number of edges in the graph tends to infinity. Then we rescale the matrix Z as following such that its bulk eigenvalues are of order 1:

$$\frac{Z}{\sqrt{n\mathtt{p}}} = Y + \frac{\mathbb{E}Z}{\sqrt{n\mathtt{p}}}, \quad Y := \frac{Z - \mathbb{E}Z}{\sqrt{n\mathtt{p}}}. \tag{2.30}$$

Note that Y is now a random Gram matrix whose entries satisfy (2.10) with

$$s_{ij} := \begin{cases} n^{-1}\delta(1-\delta\mathtt{p}), & \text{if } \sigma_1(i) = \sigma_2(j), \\ n^{-1}(2-\delta)(1-(2-\delta)\mathtt{p}), & \text{if } \sigma_1(i) \neq \sigma_2(j). \end{cases} \tag{2.31}$$

It is easy to check that S satisfies Assumption 2.5. Using this S, we can define the asymptotic density ρ as in (2.15), which has the spectral edge at λ_+. On the other hand, the mean matrix $(n\mathtt{p})^{-1/2}\mathbb{E}Z$ has two singular values of order $\sqrt{n\mathtt{p}}$ and $(\delta-1)\sqrt{n\mathtt{p}}$, respectively. In particular, these two singular values will give rise to two outliers of order $\sqrt{n\mathtt{p}}$ if $\delta > 1$, and only one outlier if $\delta = 1$.

Theorem 2.12. *Fix a $\delta \in [1, 2)$ and assume that*

$$n^{-1/3+c_\phi} \leqslant \mathtt{p} \leqslant 1/2 \tag{2.32}$$

for some constant $c_\phi > 0$. Let $\lambda_1^{SBM} \geqslant \lambda_1^{SBM} \geqslant \cdots \geqslant \lambda_p^{SBM}$ be the eigenvalues of $ZZ^\top$. Suppose λ_a^{SBM} is the extremal bulk eigenvalue of $ZZ^\top$ ($a = 2$ if $\delta = 1$, and $a = 3$ otherwise). Then we have

$$\lim_{n\to\infty} \mathbb{P}\left(\varpi^{2/3}p^{2/3}\left((n\mathtt{p})^{-1}\lambda_a^{SBM} - \lambda_+\right) \leqslant x\right) = F_1(x), \quad \text{for all } x \in \mathbb{R},$$

where ϖ is the factor given in (2.16).

Remark 2.13. We can obtain a result similar to (2.20) for the joint distribution for the first few extremal bulk eigenvalues. Moreover, our result can be easily extended to the case with more than two communities. Finally in the null case when $\delta = 1$, we can get explicit expressions for ϖ and λ_+; see [54, Theorem 2.10] for more details.

Remark 2.14. We remark that for an $n \times n$ symmetric Erdős–Rényi graphs, if $n^{-2/3} \ll \mathtt{p} \leqslant n^{-1/3}$, the second largest eigenvalue still has the Tracy-Widom fluctuation, but is around a location that is away from the spectral edge $2\sqrt{n\mathtt{p}}$ by a deterministic shift of order $(n\mathtt{p})^{-1/2}$ [72]. A similar result was proved for the bipartite SBM in the null case [54]. Studying the corresponding case for the general bipartite SBM requires a lot of extra work, which is not the focus of this paper. On the other hand, when $\mathtt{p} \ll n^{-2/3}$, the limiting distribution of the second largest eigenvalue of the Erdős–Rényi graph will become Gaussian [51, 53]. We conjecture that a similar phenomenon also happens for bipartite SBM when the sparsity is below $n^{-2/3}$.

We can extend the above setting for matrix Y, and define a more general type of random matrices, called sparse random Gram matrices as proposed in [54]. More precisely, we say that $\mathcal{Q}_{sp} = YY^\top$ is a sparse random Gram matrix if Y satisfies the following properties: the entries y_{ij}, $1 \leqslant i \leqslant p, 1 \leqslant j \leqslant n$, are independent random variables satisfying (2.10) and that

$$\mathbb{E}\left|\frac{y_{ij}}{\sqrt{ns_{ij}}}\right|^k \leqslant \frac{(Ck)^{Ck}}{nq^{k-2}}, \quad k \geqslant 3, \tag{2.33}$$

for some constant $C > 0$ and a sparsity parameter q satisfying $n^\varepsilon \leqslant q \leqslant \sqrt{n}$ for a constant $\varepsilon > 0$. Note that for Y defined in (2.30), $YY^\top$ is a sparse random Gram matrix with $q = \sqrt{n\mathtt{p}}$. We remark that the sparse random Gram matrices can be used as a natural model to study high dimensional data with randomly missing observations and the Markov switching model. For more details, we refer the reader to [54, Section 2.1].

Theorem 2.15. *Suppose $\mathcal{Q}_{sp} = YY^\top$ is a random Gram matrix with Y satisfying (2.10) and (2.33) with $q \geqslant n^{1/3+c_\phi}$ for some constant $c_\phi > 0$. Suppose the variance matrix S satisfies Assumption 2.5. Let ρ be the density defined in* (2.15) *with spectral edge at λ_+. Denote the eigenvalues of $\mathcal{Q}_{sp}$ by $\lambda_1^{sp} \geqslant \lambda_2^{sp} \geqslant \cdots \geqslant \lambda_p^{sp}$. Then we have*

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

where ϖ is the factor given in (2.16).

Remark 2.16. Our result can be applied to other types of bipartite random graph models. One important class of random network models is the famous *latent space model* proposed in the influential paper [52]. We can define similarly a bipartite latent space model with biajacency matrix Z, where the connecting probabilities between vertices depend on the latent variables in some low dimensional space. Then the edge eigenvalues of $(Z - \mathbb{E}Z)(Z - \mathbb{E}Z)^\top$ can be studied using Theorem 2.15. However, the mean matrix $\mathbb{E}Z$ will be much more complicated than the bipartite SBM case, and its effect on the spectral properties of Z needs an extensive study.

The edge statistics of symmetric DBM for Wigner type matrices in [67] have been applied to the proof of the Tracy-Widom fluctuation for random matrices with correlated entries, such as the random d-regular graph [10] and the correlated Wigner matrices [2]. We expect that our result can be also applied to Gram type random matrices with correlated entries. One important example is the random bipartite biregular graph, which is also closely related to random regular hypergraphs. We leave this to future study.

2.4 Main strategy for the proof

Before concluding this section, we briefly describe the main strategy for the proof of the main results. The proofs for Theorems 2.7, 2.12 and 2.15 are based on Theorem 2.3 and the three step strategy described in introduction. Here we mainly focus on the basic ideas for the proof of Theorem 2.3, which is the main technical part of this paper. Our proof essentially follows the same strategy as the one for the symmetric DBM of Wigner type random matrices [67]. But compared to [67], there are extra difficulties to deal with in our setting; we will mention some of them below.

Given a rectangular matrix DBM $\mathcal{Q}_t = (W + \sqrt{t}X)(W + \sqrt{t}X)^\top$, its eigenvalues satisfy a rectangular DBM, which is described by the system of SDEs in (C.1). Heuristically, (C.1) can be regarded as describing a systems of p interacting particles on the positive half real line with a random driving force (given by independent Brownian motions) and a deterministic external potential. Furthermore, the particles repulse each other through a logarithmic potential, and the repulsions will push the particles to their equilibrium locations such that the eigenvalue statistics around the edge will converge to the *local equilibrium* very quickly. (On the other hand, the eigenvalues statistics of $\mathcal{Q}_t$ reach the *global equilibrium* only when $t \gg 1$.) To justify the above intuition, we need to provide the following three ingredients.

Analysis of the rectangular free convolution. First, we need to know what the local equilibrium looks like. It is known that the asymptotic spectral density of $\mathcal{Q}_t$ is given by the rectangular free convolution with MP law, i.e., $\rho_{w,t}$ defined above (2.4), and the quantiles of $\rho_{w,t}$ give the equilibrium locations for the rectangular DBM. The properties of $\rho_{w,t}$ has been studied in [26, 27, 92] for the case $t \sim 1$, but the estimates proved there are not sufficient for our purpose when $t \ll 1$. In Section B, we collect the necessary estimates on $\rho_{w,t}$ and $m_{w,t}$ when the initial data matrix is η_*-regular and t satisfies $\sqrt{\eta_*} \ll t \ll 1$. In particular, Lemma B.5 shows that $\rho_{w,t}$ has a regular square root behavior near the right-most edge, and Lemma B.8 provides a key matching estimate that will be used in the analysis of the rectangular DBM. The proof of these estimates will be given in a separate paper [23]. Compared with [67], the analysis of this step is much more complicated. In fact the subordination function in [67] takes the simple form $\zeta = z + tm_{w,t}$, while in our case we need to deal with the quadratic subordination function in (2.5), which brings into a lot of technical difficulties.

Local law and rigidity for rectangular DBM. In this step, we establish a rigidity estimate on the eigenvalues of the rectangular DBM, which will serve as an a priori estimate for the more detailed analysis in next step. This step is relatively standard, and has been explored repeatedly in the study of Gram type random matrices [9, 13, 22, 59, 88, 95, 98]. Roughly speaking, using the local laws on the resolvent of the rectangular DBM, we can show that the edge eigenvalues are sticked to the quantiles of $\rho_{w,t}$ up to an error of order $\mathrm{O}(n^{-2/3+\varepsilon})$, which is called the rigidity estimate. Note that this error is still too large for the Tracy-Widom fluctuation, which is of order $n^{-2/3}$. The local laws and rigidity estimate of this step will be proved in a separate paper [23].

Analysis of the rectangular DBM. With the above results as the main inputs, in Section C we analyze the system of SDEs in (C.1) for the rectangular DBM $\{\lambda_i(t) : 1 \leqslant i \leqslant p\}$ around the spectral edge of the initial data matrix W. We will define another rectangular DBM with initial data matrix being a $p \times n$ matrix, say W^g, with i.i.d. Gaussian entries of mean zero and variance n^{-1}. Then the corresponding rectangular matrix DBM $\mathcal{Q}_t^g := (W^g + \sqrt{t}X)(W^g + \sqrt{t}X)^\top$ is a Wishart matrix for any t, and its eigenvalues $\{\lambda_i^g(t) : 1 \leqslant i \leqslant p\}$ are known to satisfy the Tracy-Widom fluctuation at the spectral edge [56]. We shift and rescale $\{\lambda_i(t)\}$ such that at $t = 0$, its spectral edge λ_+ matches the edge of the MP law for $\{\lambda_i^g(0)\}$, and its asymptotic density $\rho_{w,0}$ matches the MP density near λ_+. Then we adopt the coupling idea of [16] and couple the two rectangular DBMs $\{\lambda_i(t)\}$ and $\{\lambda_i^g(t)\}$ by using a common multidimensional Brownian motion as the random driving force. Now the difference between the two systems of SDEs for $\{\lambda_i(t)\}$ and $\{\lambda_i^g(t)\}$ is governed by a simple discrete parabolic equation, which we shall refer to as the coupled equation. Then for our purpose, it suffices to show that for any fixed $\mathfrak{a} \in \mathbb{N}$,

$$\max_{1 \leqslant i \leqslant \mathfrak{a}} |\lambda_i(t) - \lambda_i^g(t)| \leqslant n^{-\mathfrak{c}} \quad \text{in probability,} \tag{2.34}$$

for some constant $\mathfrak{c} > 0$ as long as $t \gg \sqrt{\eta_*}$.

The analysis of the coupled equation mainly consists of two parts. The first part is a localization of the analysis to the eigenvalues around the edge by introducing a *short-range approximation* to the original rectangular DBM. The intuition behind this approximation is that the evolutions of the edge eigenvalues $\{\lambda_i(t) : 1 \leqslant i \leqslant \mathfrak{a}\}$ should mainly depend on the nearby eigenvalues, whereas the dependence on the eigenvalues that are far away from the edge (say $\lambda_i(t)$ with $i \geqslant n^{\varepsilon}$) should be very weak. This is justified rigorously through a so-called finite speed estimate, which shows that the effect of the eigenvalues with $i \geqslant n^{\varepsilon}$ on the edge statistics is exponentially small. Such a finite speed estimate was first developed in [17], and later used in [67] to study the edge statistics of a symmetric DBM. It is advantageous in the following aspects. On one hand, our Assumption 2.1 are local, hence the rectangular DBM evolution away from the edge can be pretty irregular, but the finite speed estimate shows that it does not affect the behavior at the edge. On the other hand, our initial eigenvalues $\{\lambda_i(0)\}$ and $\{\lambda_i^g(0)\}$ match with each other only locally, and the finite speed estimate introduces a natural cut-off on the non-matching eigenvalues away from the edge. The second part of the analysis is an energy estimate of the localized coupled equation developed in [15]. It shows that for $t \gg \sqrt{\eta_*}$, the l^∞ norm of the solution to the localized coupled equation is much smaller than $n^{-2/3}$, which then implies (2.34).

We remark that our analysis of the rectangular DBM is more complicated than the symmetric DBM in [67], because the SDEs in (C.1) has a more complicated eigenvalues repulsion part than the symmetric DBM. In particular, for the short range approximation argument, there is a pretty subtle cancellation hidden inside the rectangular DBM, which is harder to identify than the symmetric DBM case; see the proofs for Lemmas C.5 and C.10 below.

3 Statistical applications

In this section, we describe statistical applications of our results to three hypothesis testing problems.

3.1 Testing the number of signals

The detection of unknown noisy signals is a fundamental task in many signal processing and wireless communication applications [6, 58, 62, 79, 86]. Consider the following generic signal-plus-noise model

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{x}, \tag{3.1}$$

where $\boldsymbol{s}$ and $\boldsymbol{x}$ are p-dimensional centered signal and noise vectors, respectively, and we assume that $\boldsymbol{s}$ and $\boldsymbol{x}$ are independent. In practice, we can independently observe n samples generated from (3.1), say $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$. In signal processing applications, $\boldsymbol{s}$ is usually generated from a low-dimensional MIMO filter [58] such that $\boldsymbol{s} = \Gamma\boldsymbol{\nu}$, where Γ is a $p \times r$ deterministic matrix and $\boldsymbol{\nu}$ is an r-dimensional centered signal vector. Here r is a fixed positive integer that does not depend on n. In signal processing, the first step is to test whether there exists any signal, that is, whether $r = 0$. This problem has been considered under different assumptions on $\boldsymbol{x}$ in the high-dimensional setting. For instance, the white noise case was considered in [12, 79], while the correlated case was addressed in [8, 80, 93]. However, when $r_0 > 0$ is a pre-specified integer, the testing problem of whether $r = r_0$ is less studied in the literature, where we aim to test the number of existing signals.

On the other hand, a closely related problem is the low-rank matrix denoising [7, 21, 25, 78, 91, 97]. Consider the model

$$\widetilde{\boldsymbol{S}} = \boldsymbol{S} + Y, \tag{3.2}$$

where $\boldsymbol{S}$ is a low-rank matrix with SVD $\boldsymbol{S} = \sum_{i=1}^r d_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ for some fixed integer r and Y is the noise matrix. Several statistics have been proposed in the literature to estimate the number of signals, i.e., the rank r of $\boldsymbol{S}$, under various assumptions on Y; see e.g. [99, 63, 80]. Here we shall perform statistical testing on r under a more general settings of Y, which have not been studied in detail in the literature. In what follows, we focus on the matrix denoising model (3.2) and study the hypothesis testing problem

$$\mathbf{H}_0^{(1)} : r = r_0 \quad \text{vs} \quad \mathbf{H}_a^{(1)} : r > r_0, \tag{3.3}$$

where r_0 is a pre-specified integer. In particular, when $r_0 = 0$, the problem aims to detect the existence of signals. We consider the following two different scenarios for Y with the choices $c_n = 0.5,\ 1,\ 2$.

(1) Y is a doubly-heteroscedastic colored noise matrix. Specifically, $Y = A^{1/2} X B^{1/2}$, where $X = (x_{ij})$ is a $p \times n$ white noise matrix with i.i.d. Gaussian entries $\mathcal{N}(0, n^{-1})$, and A and B are diagonal matrices with entries

$$a_i = 2 - \mathbf{1}_{\{1 \leqslant i \leqslant p/2\}}, \quad b_j = 4 - \mathbf{1}_{\{1 \leqslant j \leqslant n/2\}}, \quad 1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant n.$$

(2) Y is a sparse white noise matrix. Specifically, $Y = (y_{ij})$ with $y_{ij} = h_{ij} z_{ij}$, where z_{ij} are i.i.d. $\mathcal{N}(0,1)$ random variables, and h_{ij} are independent i.i.d. (rescaled) Bernoulli random variables with $\mathrm{P}(h_{ij} = (n\mathtt{p})^{-1/2}) = \mathtt{p}$ and $\mathrm{P}(h_{ij} = 0) = 1 - \mathtt{p}$. Moreover, we choose $\mathtt{p} = n^{-1/4}$.

First by Corollary 2.9 and Theorem 2.15, we find that the extremal bulk eigenvalues in both scenarios (1) and (2) obey the TW law asymptotically. Then we propose to use the following statistic, which slightly modifies the proposed quantity in [84] with $r_0 = 0$:

$$\mathbb{T}_1 := \frac{\lambda_{r_0+1} - \lambda_{r_0+2}}{\lambda_{r_0+2} - \lambda_{r_0+3}}. \tag{3.4}$$

As in [84], our theoretical results show that the statistic $\mathbb{T}_1$ is asymptotically pivotal under the null hypothesis of (3.3), and should match the corresponding statistic for a $p \times p$ Wishart matrix $XX^\top$:

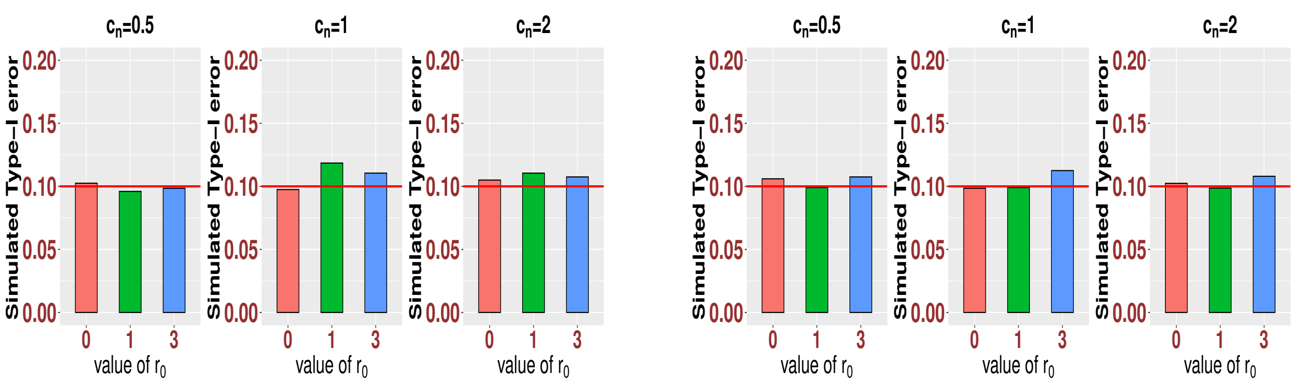
$$\frac{\lambda_1(XX^\top) - \lambda_2(XX^\top)}{\lambda_2(XX^\top) - \lambda_3(XX^\top)}.$$

This choice allows us to provide critical values without detailed knowledge of the variance structure of Y. In Table 1, we report the simulated critical values corresponding to type I error rate $\alpha = 0.1$ for different values of $c_n = 0.5,\ 1,\ 2$, based on $5,000$ simulations. All the simulations below will be based on these critical values.

$\alpha/(p,n)$	$(100, 200)$	$(250, 500)$	$(200, 200)$	$(500, 500)$	$(400, 200)$	$(1000, 500)$
0.1	4.77	4.68	4.71	4.53	4.51	4.51

Table 1: Critical values for different combination of (p, n) under nominal level 0.1.

Next, we conduct Monte-Carlo simulations to illustrate the behavior of our statistic (3.4). First, we check the accuracy of the statistic (3.4) under the nominal level 0.1 for different values of r_0 under scenarios (1) and (2). We consider the three null hypotheses with $r_0 = 0$, $r_0 = 1$ and $r_0 = 3$, for the following two choices of the signal matrix: $\boldsymbol{S}_1 = 14\mathbf{e}_{1p}\mathbf{e}_{1n}^\top$ and $\boldsymbol{S}_3 = 18\mathbf{e}_{1p}\mathbf{e}_{1n}^\top + 16\mathbf{e}_{2p}\mathbf{e}_{2n}^\top + 14\mathbf{e}_{3p}\mathbf{e}_{3n}^\top$, where $\mathbf{e}_{ip}$ and $\mathbf{e}_{in}$ are 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 under the nominal level 0.1 using the critical values from Table 1 for scenario (1). We present our results for both sample sizes $n = 200$ and $n = 500$. We conclude that our proposed statistic combined with the critical values in Table 1 can attain reasonable accuracy even for not so large n. In Figure 2, we provide the results for scenario (2), which also show that our statistic is quite accurate.



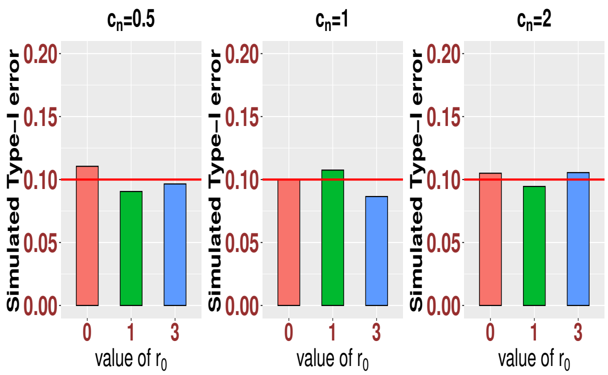
(a) $n = 200$.

(b) $n = 500$.

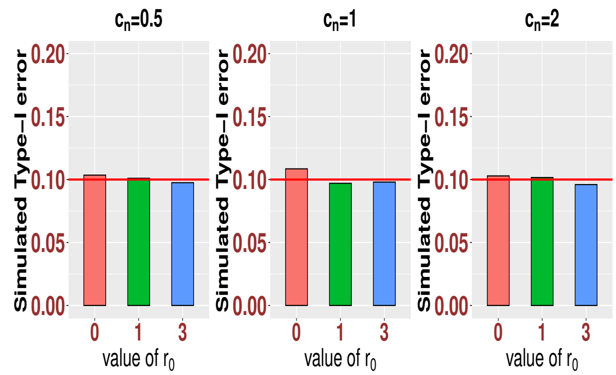
Figure 1: Scenario (1): simulated type I error rates under nominal level 0.1 for (3.3) using $\mathbb{T}_1$. We report the results based on 2,000 Monte-Carlo simulations. Here we used the critical values from Table 1.

Second, we examine the power of our statistic when $r_0 = 0$ in (3.3) under the nominal level 0.1. We set the alternative as

$$\mathbf{H}_a^{(1)}: \ \mathbf{S} = d\mathbf{e}_1\mathbf{e}_1^\top, \quad d \geqslant 0, \tag{3.5}$$



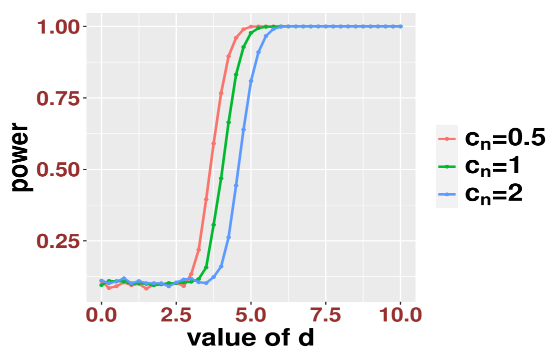
(a) $n = 200$.



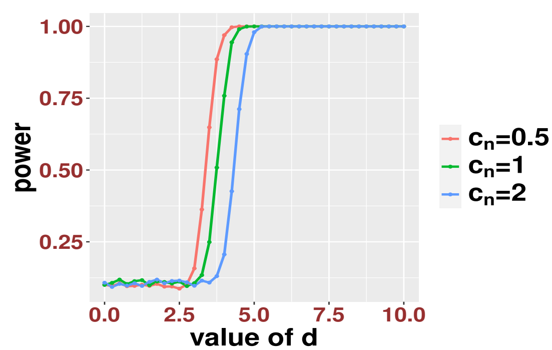
(b) $n = 500$.

Figure 2: Scenario (2): simulated type I error rates under nominal level 0.1 for (3.3) using $\mathbb{T}_1$. We report the results based on 2,000 Monte-Carlo simulations. Here we used the critical values from Table 1.

where $d = 0$ corresponds to the null case. In Figure 3, we provide the simulated power as d increases for Y generated from scenario (1). We see that our proposed statistic can have a high power even for a small n as long as d is above certain threshold. For Y generated from scenario (2), we can make the same conclusion with the simulation results in Figure 4.



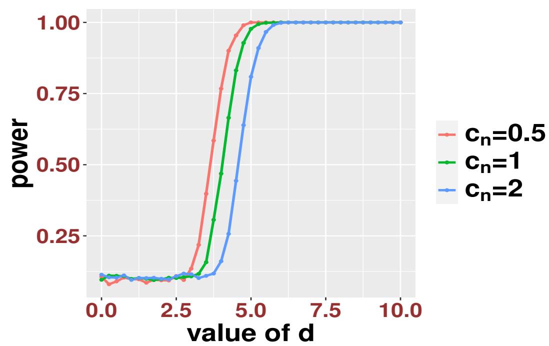
(a) $n = 200$.



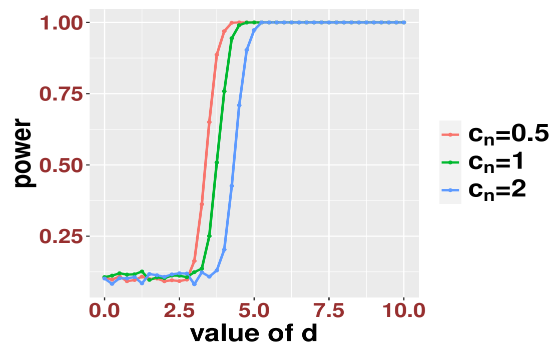
(b) $n = 500$.

Figure 3: Scenario (1): simulated power under nominal level 0.1 for the alternative (3.5) when d changes. We report our results based on 2,000 Monte-Carlo simulations.

Finally, we remark that the above statistic $\mathbb{T}_1$ and calibration procedure can be applied to the testing of number of spikes in general random Gram matrix defined in Section 2.2. Specifically, we can test the total number of spikes in spiked separable covariance matrix [24], whereas the testing of spiked sample covariance matrix have been considered in [57, 84, 86].



(a) $n = 200$.



(b) $n = 500$.

Figure 4: Scenario (2): simulated power under nominal level 0.1 for the alternative (3.5) when d changes. We report our results based on 2,000 Monte-Carlo simulations.

3.2 Testing one-sided sphericity of separable covariance matrix

Consider the separable model $Y = A^{1/2} X B^{1/2}$, where A and B are two deterministic positive definite symmetric matrices representing the covariances between spatial locations and temporal points, respectively. We are interested in testing the separable structure of Y. More precisely, we want to determine whether the temporal identity holds and study the hypothesis testing problem

$$\mathbf{H}_0^{(2)} : B = I \quad \text{vs} \quad \mathbf{H}_a^{(2)} : B \neq I. \tag{3.6}$$

Under null hypothesis, the matrix $YY^\top$ reduces to the conventional sample covariance matrix.

In [96, Section 2.2.2], the authors studied this problem by providing a visualization tool that projects the eigenvector empirical spectral distribution to multiple testing directions. However, their method cannot be applied to conduct a formal statistical hypothesis testing due to the multiple testing problem. In [8, Section 2.1], the authors used the statistic (3.4) with $r_0 = 0$ to conduct the testing. However, with both simulations and theoretical justifications, we will show that their statistic is asymptotically powerless if the alternative of B does not contain any spikes.

When the null hypothesis of (3.6) holds, the separable model is reduced to the sample covariance matrix $\mathcal{Q} = A^{1/2} X X^\top A^{1/2}$. In the literature, several efficient algorithms have been proposed to estimate A, such as [30, 61, 68, 77]. Here we shall use the method proposed in [68], and employ the `QuEST` proposed in [69] and the R package `nlshrink` to estimate A. We then propose the following procedure to conduct a two-sample test for the hypothesis testing problem (3.6). Here we remark that it is hard to use $\lambda_1 - \lambda_+$ directly in the test because we cannot estimate the right edge λ_+ up to a small enough error (i.e. much smaller than $n^{-2/3}$).

Step 1: Generate l, say $l = 5{,}000$, independent copies of $p \times n$ Wishart matrices $W_j W_j^\top$, $1 \leqslant j \leqslant l$. Fix a value $K \equiv K(n)$. For each Gaussian matrix W_j, we partition the matrix W_j into a sequence of K^2 independent submatrices $\widetilde{W}_j^i$, $i = 1, 2, \cdots, K^2$, where each $\widetilde{W}_j^i$ is a $(p/K) \times (n/K)$ matrix. Then every

$(\widetilde{W}_j^i)(\widetilde{W}_j^i)^\top$ is still a Wishart matrix, and we denote its eigenvalues by $\{\lambda_k^{ij}\}$. Next we calculate the statistic

$$\mathtt{T}_{ij}^g := \frac{1}{K}\left(\frac{n}{K}\right)^{2/3}\gamma_{\mathtt{Mp}}\left(\lambda_1^{ij}-\lambda_2^{ij}\right), \ \gamma_{\mathtt{Mp}} := (1+\sqrt{c_n})(1+c_n^{-1/2})^{1/3}. \tag{3.7}$$

Note that (3.7) is asymptotically pivotal. For each i, $1 \leqslant i \leqslant K^2$, we define

$$\overline{\mathtt{T}}_i^g := \frac{1}{l}\sum_{j=1}^{l}\mathtt{T}_{ij}^g \ .$$

Then we construct the sample $\mathbb{T}^g \in \mathbb{R}^{K^2}$ with $\mathbb{T}_i^g = \overline{\mathtt{T}}_i^g$, and we shall call it sample one.

Step 2: Use the `QuEST` function from the R package **`nlshrink`** to estimate A and denote the estimator by $\widehat{A}$. For the above given value K, we partition the matrix $\widehat{A}^{-1/2}Y$ into a sequence of K^2 submatrices $\widetilde{Y}_i$, $i = 1, 2, \cdots, K^2$, where each $\widetilde{Y}_i$ is a $(p/K) \times (n/K)$ matrix. For every $\widetilde{Y}_i\widetilde{Y}_i^\top$, we denote its eigenvalues as $\{\mu_k^i\}$. Then we calculate the statistic

$$\mathtt{T}_i := \frac{1}{K}\left(\frac{n}{K}\right)^{2/3}\gamma_{\mathtt{Mp}}\left(\mu_1^i - \mu_2^i\right).$$

We construct the sample $\mathbb{T} \in \mathbb{R}^{K^2}$ with $\mathbb{T}_i = \mathtt{T}_i$, and we shall call $\mathbb{T}$ sample two.

Step 3: Conduct two sample test for $\mathbb{T}^g$ and $\mathbb{T}$. For instance, we use **`t.test`** function in `R`.

Note that if the null hypothesis of (3.6) holds, then the eigenvalue gap $(\lambda_1 - \lambda_2)$ of $\widehat{A}^{-1/2}YY^\top\widehat{A}^{-1/2}$ should have distribution that is close to the one of the Wishart case. Hence, the above two sample test will not be able to reject the null hypothesis at a given nominal level.

Now we report the accuracy of the test proposed above and the Onatski statistic used in [8]. We shall refer to our method as BM and the method of [8] as OM. We next use Monte-Carlo simulations to illustrate the usefulness of our proposed method and statistic. For the simulations, we take A to be a diagonal matrix with eigenvalues

$$a_i = 2 - \mathbf{1}_{1 \leqslant i \leqslant p/2}, \quad 1 \leqslant i \leqslant p.$$

We pick the entries of X to be i.i.d. Gaussian random variables, and choose $K = 5$. In Figure 5, we report the results of the simulated type-I error rates under the nominal level 0.1. We find that under the null hypothesis $B = I$, both of the methods are quite accurate for different values of c_n.

To compare the power of our proposed statistic method with the one in [8], we consider the following two alternatives.

(i) For $\delta \geqslant 0$, we set $B = \operatorname{diag}\{b_1, \cdots, b_n\}$, where

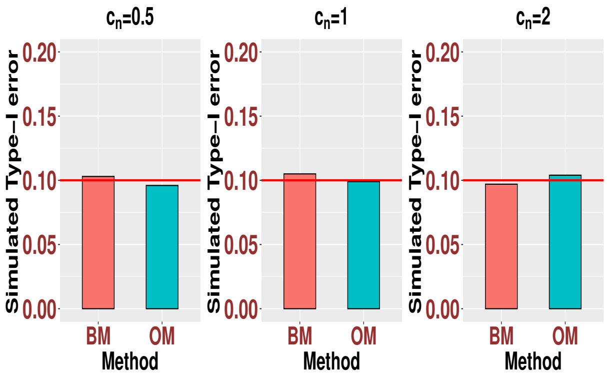
$$b_i = 1 + \delta\mathbf{1}_{\{n/2+1 \leqslant i \leqslant n\}}, \quad 1 \leqslant i \leqslant n. \tag{3.8}$$

Note that $\delta = 0$ corresponds to the null case $\mathbf{H}_0^{(2)}$. When $\delta > 0$, the variances will have two different clusters and no spikes will be generated.

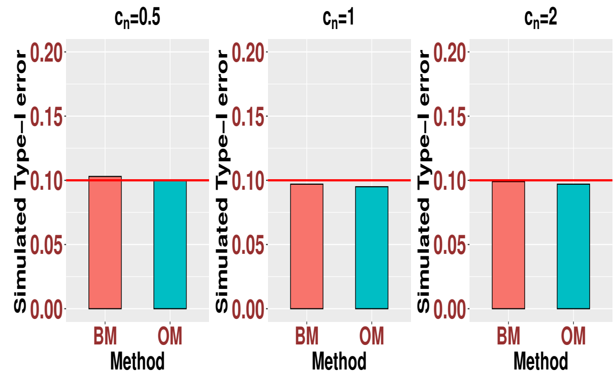
(ii) For $\delta \geqslant 0$, we let

$$B = \delta\mathbf{e}_1\mathbf{e}_1^\top + I. \tag{3.9}$$

Again $\delta = 0$ is the null case, and in the $\delta > 0$ case we have a single spike, which will potentially generate an outlier in the eigenvalue spectrum of of $YY^\top$.



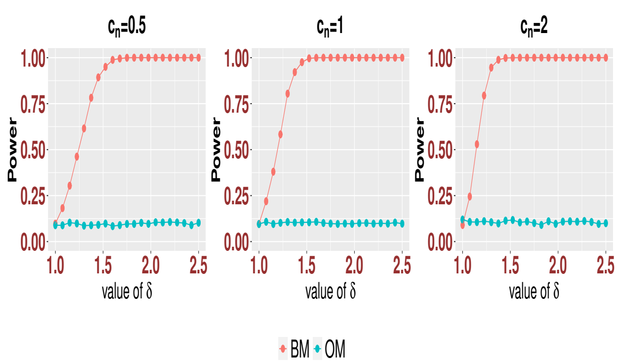
(a) $n = 500$.



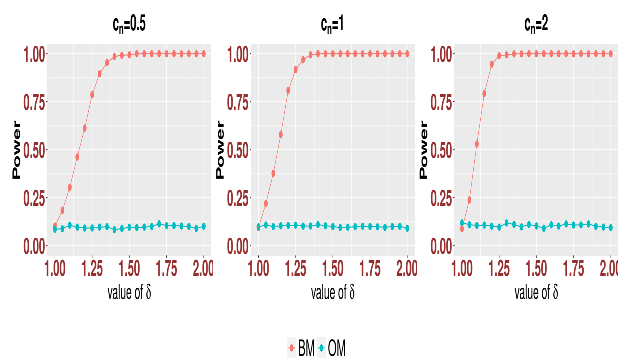
(b) $n = 800$.

Figure 5: Simulated type I error rates under nominal level 0.1. Here BM stands for our proposed methodology and statistics, whereas OM stands for the statistics used in [8]. We report the results based on 2,000 Monte-Carlo simulations and the choice $K = 5$ for our proposed method.

We pick the entries of X to be i.i.d. Gaussian random variables, and choose $K = 5$. In Figure 6, we report the simulated power under scenario (i). We find that the method in [8] is asymptotically powerless, which can be understood with the result of this paper as follows: under the alternative of scenario (i), by (2.25) the edge eigenvalues of $YY^\top$ still obey the TW fluctuation, such that the statistic used in [8] will have a similar behavior as the one under the null hypothesis of (3.6). However, our proposed statistic and procedure is able to detach such an alternative. In Figure 7, we report the results under the alternative of Scenario (ii). In this case, we find that both of the methods can attain high power once the spike is above some value, where an outlier will be generated.

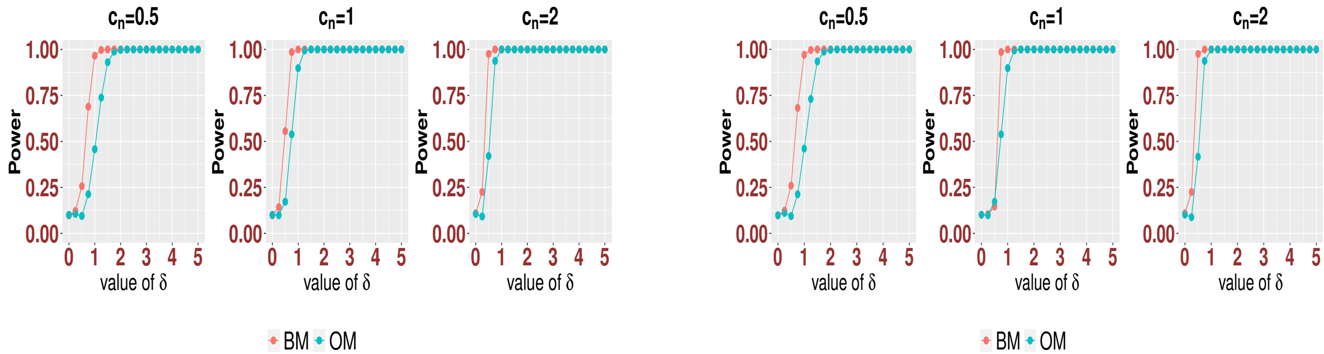


(a) $n = 500$.



(b) $n = 800$.

Figure 6: Scenario (i): simulated power under nominal level 0.1 for the alternative (3.8) for different values of δ . We report our results based on 2,000 Monte-Carlo simulations and $K = 5$.



(a) $n = 500$. (b) $n = 800$.

Figure 7: Scenario (ii): simulated power under nominal level 0.1 for the alternative (3.9) for different values of δ . We report our results based on 2,000 Monte-Carlo simulations and $K = 5$.

3.3 Testing of bipartite stochastic block model

Clustering and community detection are recurrent problems in data science and machine learning. In practice, we observe data sets with interactions represented by a network or graph models [55, 47, 76]. The most known and well studied framework for clustering is the stochastic block model (SBM), which is a random graph model with planted clusters. For a detailed review of SBM, we refer the readers to [1].

A non-symmetric generalization of the SBM is the bipartite SBM introduced in Section 2.3 (c.f. Definition 2.11), which has been used to study the user recommendation system [81]. In the bipartite SBM, people are interested in establishing useful spectral algorithms and theoretical guarantees for exact recovery of the clusters (i.e., all the cluster labels are recovered correctly with probability $1 - \mathrm{o}(1)$ as $n \to \infty$) [45, 46, 81]. As mentioned in Section 2.3, in the null case $\delta = 1$ none of the algorithms will be able to partition the vertices correctly. Now we are interested in testing

$$\mathbf{H}_0^{(3)} : \ \delta = 1 \quad \text{vs} \quad \mathbf{H}_a^{(3)} : \ \delta > 1. \tag{3.10}$$

By Theorem 2.12, we have that under the null hypothesis $\mathbf{H}_0^{(3)}$, the second eigenvalue of $ZZ^\top$ obeys the TW fluctuation, whereas under the hypothesis $\mathbf{H}_a^{(3)}$ the second largest eigenvalue is an outlier and the TW fluctuation takes effect starting from the third eigenvalue. Inspired by this observation, we construct the following statistic

$$\mathbb{T}_3 := \frac{\lambda_2 - \lambda_3}{\lambda_3 - \lambda_4},$$

which is asymptotically pivotal if the null hypothesis of (3.10) holds. In the simulations, under nominal level 0.1, we will reject the null hypothesis $\mathbf{H}_0^{(3)}$ of (3.10) if $\mathbb{T}_3$ is larger than the critical values given in Table 1.

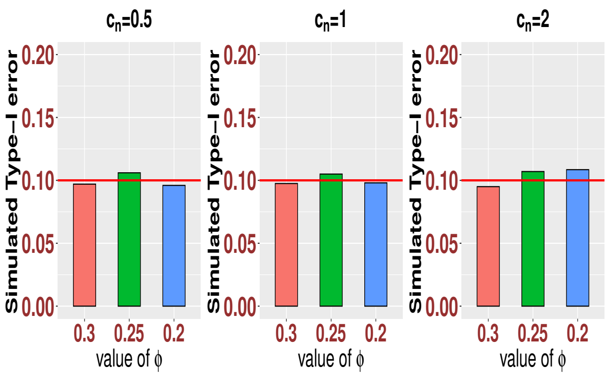
Now we conduct Monte-Carlo simulations to check the accuracy and power of the above statistic. We will choose the probability $\mathtt{p}$ as

$$\mathtt{p} = n^{-\phi}, \tag{3.11}$$

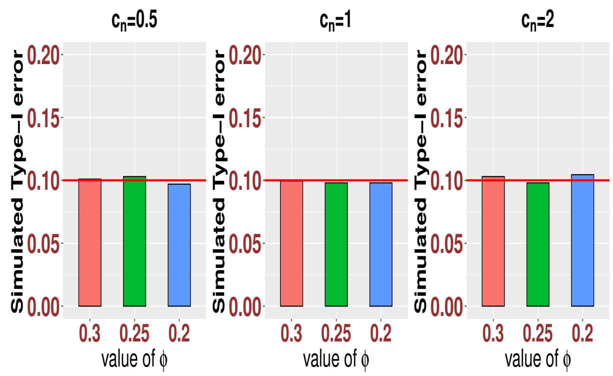
where $\phi > 0$ will be specified in the simulations. Moreover, we make the following assignment of labels

$$\sigma_1(i) = \begin{cases} +, & 1 \leqslant i \leqslant p/2 \\ -, & p/2+1 \leqslant i \leqslant p \end{cases}, \quad \sigma_2(j) = \begin{cases} +, & 1 \leqslant j \leqslant n/2 \\ -, & n/2+1 \leqslant j \leqslant n \end{cases}, \tag{3.12}$$

and the connecting probabilities are given by (2.29). In Figure 8, we report the simulated type-I error rates for different values of p. We observe that our statistic $\mathbb{T}_3$ is reasonably accurate under the null hypothesis even for not so large n. Then in Figure 9 we check the power of our statistic. We find that the statistic $\mathbb{T}_3$ can detect the deviation of δ from 1 even for small values of $(\delta - 1)$ and n.

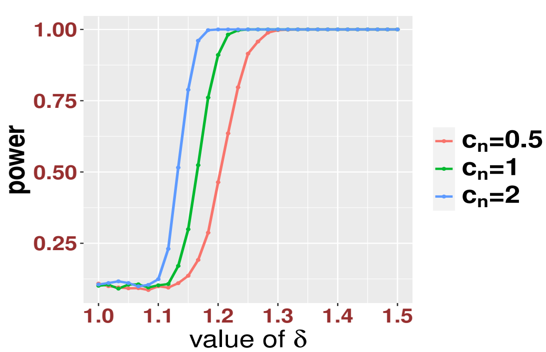


(a) $n = 200$.

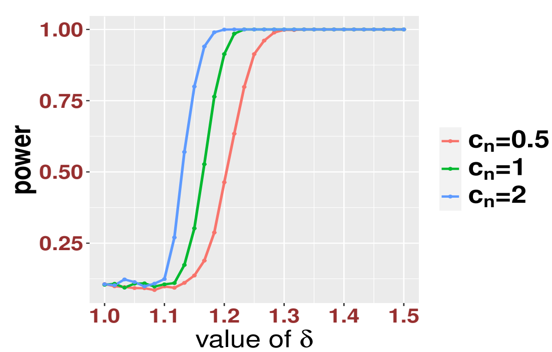


(b) $n = 500$.

Figure 8: Simulated type I error rates under nominal level 0.1 for (3.10) using $\mathbb{T}_3$. We report the results based on 2,000 Monte-Carlo simulations. Here we used the critical values from Table 1.



(a) $n = 200$.



(b) $n = 500$.

Figure 9: Simulated power under nominal level 0.1 for the alternative in (3.10) with $\mathtt{p} = n^{-1/4}$ and different values of δ. We report the results based on 2,000 Monte-Carlo simulations.

A Proof of Theorem 2.7, Theorem 2.12 and Theorem 2.15

In this section, we prove Theorem 2.7, Theorem 2.12 and Theorem 2.15 for the three random matrices model—random Gram matrices with heterogeneous variances, bipartite SBM, and sparse random Gram matrices—based on the three step strategy (or some simple variants of it) described in the introduction.

We will use the following notion of stochastic domination, which was first introduced in [31] and subsequently used in many works on random matrix theory, such as [13, 14, 18, 33, 34, 59]. 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 A.1 (Stochastic domination)**.** *(i) Let*

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

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) $\varepsilon > 0$ and (large) $D > 0$,

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

for large enough $n \geqslant n_0(\varepsilon, D)$, and we shall use the notation $\xi \prec \zeta$. Throughout this paper, the stochastic domination will always be uniform in all parameters that are not explicitly fixed, such as matrix indices and spectral parameter z that takes values in some compact set. 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 extend the definition of $\mathrm{O}_{\prec}(\cdot)$ to matrices in the operator norm sense as follows. Let A be a family of random matrices and ζ a family of nonnegative random variables. Then $A = \mathrm{O}_{\prec}(\zeta)$ means that $\|A\| \prec \zeta$.

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

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

Lemma A.2 (Lemma 3.2 of [13])**.** *Let ξ and ζ be two families of nonnegative random variables.*

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

(iii) Suppose that $\Psi(u) \geqslant n^{-C}$ is deterministic and $\xi(u)$ satisfies $\mathbb{E}\xi(u)^2 \leqslant n^C$ for all u. Then if $\xi(u) \prec \Psi(u)$ uniformly in u, we have $\mathbb{E}\xi(u) \prec \Psi(u)$ uniformly 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. [22, 24, 73, 98].

Definition A.3 (Bounded support condition)**.** *We say a random matrix Y satisfies the bounded support condition with ϕ_n if*

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

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

Then we introduce the following $(p+n)\times(p+n)$ self-adjoint 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 the resolvents

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

Using Schur complement formula, it is easy to check that

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

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 as

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

In [5], it is shown that 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 $M_1 : \mathbb{C}_+ \to \mathbb{C}^p$ and $M_2 : \mathbb{C}_+ \to \mathbb{C}^n$ 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 $\mathrm{Im}\, M_{1,i}(z) > 0$, $i = 1,2,\cdots,p$, and $\mathrm{Im}\, M_{2,j}(z) > 0$, $j = 1,2,\cdots,n$, 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 (2.14), which shows that $M_1(z) = \mathbf{m}(z)$. We now define the asymptotic matrix limit of G as

$$\Pi(z) := \mathrm{diag}\left(M_{1,1}(z), \cdots M_{1,p}(z), M_{2,1}(z), \cdots M_{2,n}(z)\right). \tag{A.8}$$

Then we define the following spectral domains: for constants $c_0, \vartheta > 0$,

$$\mathcal{D}(c_0,\vartheta) := \left\{z = E + \mathrm{i}\eta : \lambda_+ - c_0 \leqslant E \leqslant \lambda_+ + c_0^{-1}, n^{-1+\vartheta} \leqslant \eta \leqslant c_0^{-1}\right\};$$
$$\mathcal{D}_{out}(c_0,\vartheta) := \mathcal{D}(c_0,\vartheta) \cap \{z = E + \mathrm{i}\eta : E \geqslant \lambda_+, N\eta\sqrt{\kappa+\eta} \geqslant N^\vartheta\}.$$

Finally, we define the distance to the rightmost edge as

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

The following local law has been proved in [4].

Lemma A.4 (Theorem 2.6 of [4]). *Assume that Y is a $p \times n$ random matrix with real independent entries satisfying (2.10) and that for any fixed $k \in \mathbb{N}$,*

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

for some constant $C_k > 0$. Moreover, suppose that the variance matrix $S = (s_{ij})$ satisfies Assumption 2.5. Then there exists a constant $c_0 > 0$ such that following estimates hold for any (small) constant $\vartheta > 0$.

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

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

where recall that $m(z)$ *is defined in* (2.15)*. Moreover, for* $z \in \mathcal{D}_{out}(c_0, \vartheta)$ *we have the stronger estimate*

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

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

$$\max_{1\leqslant i,j \leqslant p+n} |G_{ij}(z) - \Pi_{ij}(z)| \prec \sqrt{\frac{\operatorname{Im} m(z)}{n\eta}} + \frac{1}{n\eta}. \tag{A.13}$$

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

Remark A.5. Strictly speaking, the estimate (A.12) was not proved in [4]. However, its proof is standard by combing the results in [4] with a separate argument for $z \in \mathcal{D}_{out}(c_0, \vartheta)$; see e.g. the proof of (2.20) in [34].

As a consequence of (A.11) and (A.12), we obtain the following rigidity estimates for the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$ of $\mathcal{Q}_1$ near the right edge λ_+. We define the classical location γ_j of the j-th eigenvalue as

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} \rho(x)\mathrm{d}x > \frac{j-1}{p} \right\}, \tag{A.14}$$

where recall that ρ is defined in (2.15), and it gives the asymptotic spectral density of $\mathcal{Q}_1$. By the above definition, we have $\gamma_1 = \lambda_+$.

Lemma A.6. *For any* j *such that* $\lambda_+ - c_0/2 < \gamma_j \leqslant \lambda_+$*, we have*

$$|\lambda_j - \gamma_j| \prec j^{-1/3} n^{-2/3}. \tag{A.15}$$

Proof. The estimate (A.15) follows from the estimates (A.11) and (A.12) combined with a standard argument using Helffer-Sjöstrand calculus. The details are already given in [35], [42] and [88]. □

Note that combining (A.10) with Markov's inequality, we get that the matrix Y in Lemma A.4 has support $n^{-1/2}$. Now combining the analysis of equation (2.14) in [4] with the arguments for local law in [22], we can relax the moment condition (A.10) to a weaker bounded support condition.

Lemma A.7. *Suppose all the assumptions of Lemma A.4 hold except for* (A.10)*. Moreover, assume that* Y *satisfies the bounded support condition (A.1) with* $\phi_n \leqslant n^{-c_\phi}$ *for some constant* $c_\phi > 0$*. Then there exists a constant* $c_0 > 0$ *such that following estimates hold for any (small) constant* $\vartheta > 0$*.*

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

$$|g_1(z) - m(z)| \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + (n\eta)^{-1}. \tag{A.16}$$

where recall that $m(z)$ *is defined in* (2.15)*. Moreover, for* $z \in \mathcal{D}_{out}(c_0, \vartheta)$ *we have the stronger estimate*

$$|g_1(z) - m(z)| \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}. \tag{A.17}$$

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

$$\max_{1 \leqslant i,j \leqslant p+n} |G_{ij}(z) - \Pi_{ij}(z)| \prec \phi_n + \sqrt{\frac{\operatorname{Im} m(z)}{n\eta}} + \frac{1}{n\eta}. \tag{A.18}$$

Proof. With the stability analysis of equation (2.14) in [4, Section 3], we can repeat the arguments for Lemma 3.11 of [22] and Theorem 3.6 of [98] to conclude the proof. We omit the details. □

Now we are ready to give the proof of Theorem 2.7.

Proof of Theorem 2.7. With the local laws in Lemma A.7, we can repeat the proof for Theorem 2.7 of [22] almost verbatim to conclude (2.19) and the following universality estimate:

$$\lim_{n\to\infty} \left[\mathbb{P}(p^{2/3}(\lambda_1 - \lambda_+) \leqslant x) - \mathbb{P}^G(p^{2/3}(\lambda_1 - \lambda_+) \leqslant x)\right] = 0 \tag{A.19}$$

for all $x \in \mathbb{R}$, where $\mathbb{P}^G$ denotes the law for $Y = (y_{ij})$ with independent Gaussian entries satisfying (2.10). To conclude the proof, it remains to show that $\varpi p^{2/3}(\lambda_1 - \lambda_+)$ converges weakly to type-1 Tracy-Widom for a Gaussian Y.

Let $t_0 = n^{-1/3+\varepsilon_0}$ for some small constant $\varepsilon_0 < \varepsilon_*$, where recall that ε_* is given in (2.11). Then we pick the initial data matrix W as a matrix with independent Gaussian entries satisfying

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

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

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

We regard $W + \sqrt{t}X$ as a rectangular matrix Dyson Brownian motion starting at 0, and at time t_0 it has the same distribution as Y. We now fix the notations for the proof.

First, for $\mathcal{Q} := (W + \sqrt{t_0}X)(W + \sqrt{t_0}X)^\top$ we denote its eigenvalues by $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$. We follow the notations in Section 2.2 and define its asymptotic spectral density ρ and the corresponding Stieltjes transform $m(z)$ as in (2.15). Then let λ_+ be the right edge of ρ, and γ_j be the quantiles of ρ defined as in (A.14). For W, we denote its variance matrix as $S_w = (s_{ij} - t_0/n : 1 \leqslant i \leqslant p, 1 \leqslant j \leqslant n)$, 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

$$\frac{1}{\mathbf{M}_w} = -z\mathbf{1} + S_w \frac{1}{\mathbf{1} + S_w^\top \mathbf{M}_w}, \tag{A.20}$$

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

$$\gamma_{j,w} := \sup_x \left\{\int_x^{+\infty} \rho_w(x)dx > \frac{j-1}{p}\right\}, \quad 1 \leqslant j \leqslant p. \tag{A.21}$$

Finally, following the notations in Section 2.1, we denote $m_V(z) = m_{w,0}(z) := p^{-1}\operatorname{Tr}(WW^\top - z)^{-1}$ and the eigenvalues of $WW^\top$ by $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_p$. Then we define $m_{w,t}$ as in (2.4), 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 some small enough constant $0 < \varepsilon_1 < \varepsilon_0$. We first verify that m_V is η_*-regular as in Definition 2.1. Notice that W is also a random Gram matrix satisfying the assumptions of Lemma A.4. Hence by (A.11) and (A.12) we have

$$|m_{w,0}(z) - M_w(z)| \prec (n\eta)^{-1}, \quad \lambda_{+,w} - c_0 \leqslant E \leqslant \lambda_{+,w}, \quad n^{-2/3+\vartheta} \leqslant \eta \leqslant 10, \tag{A.22}$$

and

$$|m_{w,0}(z) - M_w(z)| \prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}, \quad \lambda_{+,w} \leqslant E \leqslant \lambda_{+,w} + c_0, \quad n^{-2/3+\vartheta} \leqslant \eta \leqslant 10. \tag{A.23}$$

Moreover, as a consequence of the square root behavior of ρ_w around λ_+ in (2.16), it is easy to show that for $z \in \mathcal{D}(c_0, \vartheta)$,

$$|M_w(z)| \sim 1, \quad \operatorname{Im} M_w(z) \sim \begin{cases} \eta/\sqrt{\kappa+\eta}, & \text{if } E \geqslant \lambda_+ \\ \sqrt{\kappa+\eta}, & \text{if } E \leqslant \lambda_+ \end{cases}. \tag{A.24}$$

Finally, using (A.15) we get that

$$|d_j - \gamma_{j,w}| \prec j^{-1/3} n^{-2/3} \tag{A.25}$$

for any j such that $\lambda_{+,w} - c_0/2 < \gamma_j \leqslant \lambda_{+,w}$. Combining the above estimates, we obtain that for some constants $c_V < c_0/2$ and $C_V > 0$, the following estimates hold on a high probability event Ξ:

$$\frac{1}{C_V}\sqrt{|d_1 - E| + \eta} \leqslant \operatorname{Im} m_{w,0}(E + \mathrm{i}\eta) \leqslant C_V\sqrt{|d_1 - E| + \eta}, \quad d_1 - c_V \leqslant E \leqslant d_1, \quad \eta_* \leqslant \eta \leqslant 10,$$

and

$$\frac{1}{C_V}\frac{\eta}{|d_1 - E| + \eta} \leqslant \operatorname{Im} m_{w,0}(E + \mathrm{i}\eta) \leqslant C_V\frac{\eta}{|d_1 - E| + \eta}, \quad d_1 \leqslant E \leqslant d_1 + c_V, \quad \eta_* \leqslant \eta \leqslant 10.$$

Thus on the event Ξ, m_V is η_*-regular. Then by Theorem 2.3, there exists a parameter $\gamma_0 \equiv \gamma_0(t_0) \sim 1$ such that

$$\gamma_0 p^{2/3}(\lambda_1 - \lambda_{+,t_0}) \to F_1 \tag{A.26}$$

in distribution. Now to conclude the proof, it remains to show that

$$n^{2/3}|\lambda_{+,t_0} - \lambda_+| \to 0 \quad \text{in probability.} \tag{A.27}$$

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 in general, but we can control their difference using the local laws (A.22) and (A.23).

Recalling the notation in (2.5), we denote

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

and

$$\zeta_+ := [1 + c_n t_0 m(\lambda_+)]^2 \lambda_+ - (1 - c_n)t_0[1 + c_n t_0 m(\lambda_+)].$$

Then repeating the proof of Lemma B.7 (which was given in Lemma 3.20 of [23]), we can obtain that

$$|\lambda_{+,t_0} - \lambda_+| \lesssim |\zeta_{+,t_0} - \zeta_+| + t_0|M_w(\zeta_{+,t_0}) - M_w(\zeta_+)| + t_0\left|m_{w,0}(\zeta_{+,t_0}) - M_w(\zeta_{+,t_0})\right|, \tag{A.28}$$

and

$$|\zeta_{+,t_0} - \zeta_+| \lesssim t_0^3 \left| m'_{w,0}(\zeta_{+,t_0}) - M'_w(\zeta_{+,t_0}) \right|, \tag{A.29}$$

where by (B.12) and (A.25) we have

$$|\zeta_+ - \lambda_{+,w}| \sim |\zeta_{+,t_0} - \lambda_{+,w}| \sim t_0^2. \tag{A.30}$$

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

$$\begin{aligned}
\left| m'_{w,0}(\zeta_{+,t_0}) - M'_w(\zeta_{+,t_0}) \right| &= \left| \frac{1}{p} \sum_j \frac{1}{(d_j - \zeta_{+,t_0})^2} - \int_0^{\lambda_{+,w}} \frac{\rho_w(x)}{(x - \zeta_{+,t_0})^2} \mathrm{d}x \right| \\
&\leqslant \sum_{j \geqslant 2: \gamma_{j,w} > \lambda_{+,w} - c_0/2} \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \left| \frac{\rho_w(x)}{(d_j - \zeta_{+,t_0})^2} - \frac{\rho_w(x)}{(x - \zeta_{+,t_0})^2} \right| \mathrm{d}x + \mathrm{O}(1) \\
&\prec \sum_{j \geqslant 2: \gamma_{j,w} > \lambda_{+,w} - c_0/2} \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \frac{j^{-1/3} n^{-2/3} (|\lambda_{+,w} - x| + t_0^2) \rho_w(x)}{|x - \zeta_{+,t_0}|^4} \mathrm{d}x + \mathrm{O}(1) \\
&\lesssim n^{-1} \int_{\lambda_{+,w} - c_0/2}^{\lambda_{+,w}} \frac{(\lambda_{+,w} - x) + t_0^2}{|(\lambda_{+,w} - x) + t_0^2|^4} \mathrm{d}x + \mathrm{O}(1) \lesssim \frac{1}{n t_0^4},
\end{aligned} \tag{A.31}$$

where in the third step we used that for $\gamma_{j,w} \leqslant x \leqslant \gamma_{j-1,w}$,

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

by (A.25), (A.30) and $\lambda_{+,w} - \gamma_{j,w} \gtrsim j^{2/3} n^{-2/3}$, and in the fourth step we used $\rho_w(x) \sim \sqrt{\lambda_{+,w} - x}$, $j^{-1/3} \sim n^{-1/3} (\lambda_{+,w} - x)^{-1/2}$. Thus by (A.29) we have

$$|\zeta_{+,t_0} - \zeta_+| \prec n^{-1} t_0^{-1}. \tag{A.32}$$

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

$$t_0 |M_w(\zeta_{+,t_0}) - M_w(\zeta_+)| \lesssim t_0 \frac{|\zeta_{+,t_0} - \zeta_+|}{\min\{|\zeta_+ - \lambda_{+,w}|^{1/2}, |\zeta_{+,t_0} - \lambda_{+,w}|^{1/2}\}} \lesssim n^{-1} t_0^{-1}, \tag{A.33}$$

where we used (A.30) in the last step. Finally we bound $|m_{w,0}(\lambda_{+,t_0}) - M_w(\lambda_{+,t_0})|$. Denote $z_0 := \lambda_{+,t_0} + \mathrm{i}\eta_0$ with $\eta_0 := n^{-2/3+\vartheta}$ for some small constant $\vartheta > 0$. We now decompose $m_{w,0}(\lambda_{+,t_0}) - M_w(\lambda_{+,t_0})$ as

$$m_{w,0}(\lambda_{+,t_0}) - M_w(\lambda_{+,t_0}) = m_{w,0}(z_0) - M_w(z_0) + \mathcal{K}_1 + \mathcal{K}_2,$$

where

$$\mathcal{K}_1 := \sum_{j \geqslant 2: \gamma_{j,w} > \lambda_{+,w} - c_0/2} \left(\frac{1}{d_j - \lambda_{+,t_0}} - \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \frac{\rho_w(x)}{x - \lambda_{+,t_0}} \right) - \sum_{j \geqslant 2: \gamma_{j,w} > \lambda_{+,w} - c_0/2} \left(\frac{1}{d_j - z_0} - \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \frac{\rho_w(x)}{x - z_0} \right),$$

$$\mathcal{K}_2 := \sum_{j: \gamma_{j,w} \leqslant \lambda_{+,w} - c_0/2} \left(\frac{1}{d_j - \lambda_{+,t_0}} - \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \frac{\rho_w(x)}{x - \lambda_{+,t_0}} \right) - \sum_{j: \gamma_{j,w} \leqslant \lambda_{+,w} - c_0/2} \left(\frac{1}{d_j - z_0} - \int_{\gamma_{j,w}}^{\gamma_{j-1,w}} \frac{\rho_w(x)}{x - z_0} \right).$$

By (A.23), 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.25), it is easy to bound $\mathcal{K}_2 \lesssim \eta_0$ with high probability. Finally, using a similar argument as for (A.31), we can bound

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

Combining the above estimates, we get

$$|m_{w,0}(\lambda_{+,t_0}) - M_w(\lambda_{+,t_0})| \prec \eta_0 + \frac{1}{nt_0^2} + \frac{1}{(n\eta_0)^2 t_0}. \tag{A.34}$$

Now with (A.28), (A.32), (A.33) and (A.34), we can bound that

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

Plugging into $t_0 = n^{-1/3+\varepsilon_0}$ and $\eta_0 = n^{-2/3+\vartheta}$, we conclude (A.27) as long as ε_0 and ϑ are sufficiently small.

Combining (A.26) and (A.27), we obtain that $\gamma_0 p^{2/3}(\lambda_1 - \lambda_{+,t_0})$ converges to the Tracy-Widom law. Furthermore, matching the gap between the quantiles γ_1 and γ_2 (recall (A.14)) for the density in (2.16) and the one for the semicircle law $\rho_{sc}(2-x) = \pi^{-1}\sqrt{x} + \mathrm{O}(x)$ around the right edge at $\lambda_+ = 2$, we see that γ_0 must be $\varpi^{2/3}$. This concludes the proof. □

Next we prove Theorem 2.15, which then implies Theorem 2.12 as a special case. We remark that Theorem 2.15 is not simply a special case of Theorem 2.7. In fact, in the setting of Theorem 2.7 the distribution of $\widehat{y}_{ij} = y_{ij}/\sqrt{s_{ij}}$ is independent of n, while for sparse random Gram matrices, the distribution of $\widehat{y}_{ij}$ may change with respect to n under (2.33).

Proof of Theorem 2.15. Combining (2.33) with Markov's inequality, we see that Y satisfies the bounded support condition (A.1) with $\phi_n = q^{-1} \leqslant n^{-1/3-c_\phi}$. Then Lemma A.7 holds, and in [22, Lemma 3.11] we have shown that (A.16)-(A.17) imply the following weaker rigidity estimate than (A.15):

$$|\lambda_j - \gamma_j| \prec j^{-1/3}n^{-2/3} + \phi_n^2. \tag{A.35}$$

Then with (A.16), (A.17) and (A.35) as the main inputs, using the same arguments as for [32, Theorem 2.7] we can show that the edge statistics of $\mathcal{Q}_{sp}$ match those of the Gaussian case in the sense of (A.19) as long as $\phi_n \leqslant n^{-1/3-c_\phi}$. Then the rest of the proof proceeds as in the above proof of Theorem 2.7. □

Proof of Theorem 2.12. By Theorem 2.15, we immediately obtain that the edge eigenvalues of $YY^\top$ satisfy desired Tracy-Widom law for Y in (2.30). On the other hand, with the same arguments as for [32, Theorem 6.2], we can show that largest *bulk eigenvalue* (not the outliers) of $(n\mathtt{p})^{-1}ZZ^\top$ are sticked to the largest eigenvalue of $YY^\top$ up to an error of order $n^{-1+\varepsilon}$. This concludes Theorem 2.12. Alternatively, one can also use the version of Theorem 2.3 with outliers; see Remark 2.4. □

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 W as in Definition 2.1. Furthermore, we will state an (almost) sharp local law on the resolvent of the rectangular matrix DBM $\mathcal{Q}_t = (W + \sqrt{t}X)(W + \sqrt{t}X)^\top$, and a rigidity estimate on the rectangular DBM $\{\lambda_i(t) : 1 \leqslant i \leqslant p\}$. As described in Section 2.4, 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 [23] under more general assumptions on X, and we shall provide the exact reference for each of them.

B.1 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 (2.4) 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 (2.5), 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 density associated with $m_{w,t}$, and let $\mu_{w,t}$ be the corresponding probability measure. Moreover, we denote the support of $\rho_{w,t}$ as $S_{w,t}$, with a right-most edge at $\lambda_{+,t}$. We first summarize some basic properties of these quantities, which have been proved in previous works [26, 27, 92].

Lemma B.1 (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 (2.4) *satisfying that* $\operatorname{Im} m_{w,t}(z) > 0$ *and* $\operatorname{Im} z m_{w,t}(z) > 0$ *if* $\operatorname{Im} z > 0$.

(ii) For all $x \in \mathbb{R}\backslash\{0\}$, $\lim_{\eta \downarrow 0} m_{w,t}(x + \mathrm{i}\eta)$ exists, and we denote it as $m_{w,t}(x)$. The function $m_{w,t}$ is continuous on $\mathbb{R}\backslash\{0\}$. The measure $\mu_{w,t}$ has a continuous density $\rho_{w,t}$ given by $\rho_{w,t}(x) = \pi^{-1} \operatorname{Im} m_{w,t}(x)$ on $\mathbb{R}\backslash\{0\}$. Moreover, $m_{w,t}(x)$ is a solution to (2.4) for $z = x$.

(iii) For all $x \in \mathbb{R}\backslash\{0\}$, $\lim_{\eta \downarrow 0} \zeta_t(x + \mathrm{i}\eta)$ exists, and we denote it as $\zeta_t(x)$. Moreover, we have $\operatorname{Im} \zeta_t(z) > 0$ if $\operatorname{Im} z > 0$.

(iv) We have $\operatorname{Re} b_t(z) > 0$ for all $z \in \mathbb{C}_+$ and $|m_{w,t}(z)| \leqslant (c_n t|z|)^{-1/2}$.

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

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

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

Proof. (i) follows from [27, Theorem 4.1], (ii) and (iii) follow from [26, Lemma 2.1] and [92, Proposition 1], (iv) follows from [26, Lemma 2.1] and (v) follows from [92, Propositions 1 and 2]. ☐

The following lemma characterizes the right-most edge of $S_{w,t}$. One can rewrite the equation (B.2) as

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

where Φ_t is defined in (2.6) and is an analytic function on $\mathbb{C}_+$. We recall that by definition

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

In [92], the authors characterize the support of $\mu_{\omega,t}$ and its edges using the local extrema of Φ_t on $\mathbb{R}$.

Lemma B.2. *Fix any $t > 0$. The function $\Phi_t(x)$ on $\mathbb{R}\backslash\{0\}$ admits $2q$ positive local extrema counting multiplicities, where $q \geqslant 1$ is some integer. The preminages of these extrema are denoted by $\zeta_{1,-}(t) < 0 < \zeta_{1,+}(t) \leqslant \zeta_{2,-}(t) \leqslant \zeta_{2,+}(t) \leqslant \cdots \leqslant \zeta_{q,-}(t) \leqslant \zeta_{q,+}(t)$, and they 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 $(-\infty, \zeta_{1,-}(t)]$, $[\zeta_{1,+}(t), \zeta_{2,-}(t)]$, $\cdots$, $[\zeta_{q-1,+}(t), \zeta_{q,-}(t)]$, $[\zeta_{q,+}(t), \infty)$. 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 $\zeta_{q,-}(t) < d_1 < \zeta_{q,+}(t)$.*

Proof. See [92, Proposition 3] and the discussion below [92, Theorem 2], or [74, Lemma 1]. □

Now we rewrite (B.2) into a equation for ζ only. We focus on $z \in \mathbb{C}_+$ with $\operatorname{Re} z > 0$. Then we can solve from (2.5) 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 B.1 (iv) holds. Together with (B.2), we find that the pair (z, b_t) is a solution to (B.2) if and only if (z, ζ_t) is a solution to

$$F_t(z,\zeta) = 0, \quad \text{with} \quad F_t(z,\zeta) := 1 + \frac{t(1-c_n) - \sqrt{t^2(1-c_n)^2 + 4\zeta z}}{2\zeta} - c_n t m_{w,0}(\zeta). \tag{B.6}$$

Since $\Phi_t(\zeta_t) = x$ and $F_t(x, \zeta_t) = 0$ are essentially the same equation, from Lemma B.2 we can derive the following characterization of the edges of $S_{w,t}$.

Lemma B.3. *Denote $a_{k,\pm}(t) := \Phi_t(\zeta_{k,\pm}(t))$, $1 \leqslant k \leqslant q$. Then $(a_{k,\pm}(t), \zeta_{k,\pm}(t))$ are real solutions to*

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

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

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

By Lemma B.2, 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 it 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 B.3 to derive an expression for the partial derivative $\partial_t \lambda_{+,t}$, which will be used in the analysis of rectangular DBM. 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 $F(t, z, \zeta) \equiv F_t(z, \zeta)$. Thus we can solve that

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

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 2.1. For the following results, we always assume that

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

for some constant $\varepsilon > 0$. Note that under this condition, we have $n^{\varepsilon}\eta_* \leqslant t^2 \leqslant n^{-\varepsilon}$. We centralize ζ at the right-most edge λ_+ of the spectrum of V, that is, we define

$$\xi_t(z) := \zeta_t(z) - \lambda_+, \quad \text{and} \quad \xi_{+,t} := \zeta_{+,t} - \lambda_+, \tag{B.11}$$

where $\zeta_{+,t} := \zeta(\lambda_{+,t})$.

Lemma B.4 (Lemma 3.5 of [23]). *Suppose $V = WW^\top$ is η_*-regular, and t satisfies* (B.10)*. Then we have $\xi_{+,t} \geqslant 0$ and*

$$\xi_{+,t} \sim t^2. \tag{B.12}$$

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

Lemma B.5 (Lemmas 3.16 and 3.17 of [23]). *Suppose $V = WW^\top$ is η_*-regular, and t satisfies* (B.10)*. Then for $\kappa := |E - \lambda_+| \leqslant 3c_V/4$, the asymptotic density satisfies*

$$\rho_{w,t}(E) \sim \sqrt{(\lambda_{+,t} - E)_+}. \tag{B.13}$$

Moreover, if $-\tau t^2 \leqslant E - \lambda_{+,t} \leqslant 0$ for some sufficiently small constant $\tau > 0$, we have

$$\rho_{w,t}(E) = \frac{1}{\pi}\sqrt{\frac{2(\lambda_{+,t} - E)}{[4\lambda_{+,t}^2 + (1-c_n)^2 t^2] c_n^2 t^2 \Phi_t''(\zeta_{+,t})}} \left[1 + \mathrm{O}\left(\frac{|E - \lambda_{+,t}|}{t^2}\right)\right], \tag{B.14}$$

where $t^2\Phi''(\xi_+(t))| \sim 1$. Finally, as a consequence of (B.13)*, the following estimates hold:*

$$|m_{w,t}| \lesssim 1, \quad \operatorname{Im} m_{w,t}(z) \sim \begin{cases} \sqrt{\kappa + \eta}, & E \leqslant \lambda_{+,t} \\ \dfrac{\eta}{\sqrt{\kappa+\eta}}, & E \geqslant \lambda_{+,t} \end{cases}. \tag{B.15}$$

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

$$|\partial_z m_{w,t}(z)| = \left| \int \frac{\mathrm{d}\mu_{w,t}(x)}{(x-z)^2} \right| \leqslant \frac{\operatorname{Im}\, m_{w,t}}{\eta}. \tag{B.16}$$

Moreover, we claim the following estimates.

Lemma B.6 (Lemma 3.18 of [23])**.** *Suppose $V = WW^\top$ is η_*-regular, and t satisfies* (B.10)*. For $\kappa + \eta \leqslant t^2$, we have*

$$|\partial_z m_{w,t}(z)| \lesssim \frac{1}{\sqrt{\kappa+\eta}}. \tag{B.17}$$

Moreover, if $\kappa + \eta \geqslant t^2$, we have that for $E \geqslant \lambda_{+,t}$,

$$|\partial_z m_{w,t}(z)| \lesssim \frac{1}{\sqrt{\kappa+\eta}}, \tag{B.18}$$

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

$$|\partial_z m_{w,t}(z)| \lesssim \frac{\sqrt{\kappa+\eta}}{t\sqrt{\kappa+\eta}+\eta}. \tag{B.19}$$

Finally, we compare the edge behaviors of two free rectangular convolution measures satisfying certain matching properties given in Section C. Specifically, let $t_0 = N^{-1/3+\omega_0}$ for some constant $0 < \omega_0 < 1/3$. We consider two 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:

$$\rho_1(\psi - x) = \rho_2(\psi - x)\left[1 + \mathrm{O}\left(\frac{|x|}{t_0^2}\right)\right], \quad 0 \leqslant x \leqslant c_\psi t_0^2, \tag{B.20}$$

and

$$\rho_1(x) = \rho_2(x) = 0 \ \text{ on } \ [\psi, 2\psi], \quad \rho_1(x) \sim \rho_2(x) \sim \sqrt{\psi - x} \ \text{ on } \ [\psi - c_\psi, \psi]. \tag{B.21}$$

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 $\rho_{i,t}$, $m_{i,t}$, 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

$$b_{i,t}(z) := 1 + c_n t m_{i,t}(z), \quad \zeta_{i,t}(z) := b_{i,t}^2 z - (1 - c_n) t b_{i,t}. \tag{B.22}$$

As in (B.11), we introduce the notation $\xi_{i,t}(z) := \zeta_{i,t}(z) - \psi$. For $i = 1, 2$, let $\lambda_{+,i}(t)$ be the right edge of $\rho_{i,t}$, and denote $\xi_{+,i}(t) := \xi_{i,t}(\lambda_{+,i}(t))$. Due to the matching condition (B.20), we can show that $\xi_{+,1}(t)$ and $\xi_{+,2}(t)$ are close to each other with a distance of order $\mathrm{o}(t^2)$.

Lemma B.7 (Lemma 3.20 of [23])**.** *Suppose* (B.20) *and* (B.21) *hold. Then there exists a constant $C > 0$ such that for any $0 \leqslant t \leqslant t_0$,*

$$|\xi_{+,1}(t) - \xi_{+,2}(t)| \leqslant \frac{Ct^3}{t_0}, \tag{B.23}$$

and

$$|\lambda_{+,1}(t) - \psi| + |\lambda_{+,2}(t) - \psi| \leqslant Ct. \tag{B.24}$$

The following matching estimates will play an important role in the analysis of rectangular DBM.

Lemma B.8 (Lemmas 3.22 and 3.23 of [23])**.** *Suppose* (B.20) *and* (B.21) *hold, and* $0 < t \leqslant t_0 n^{-\varepsilon_0}$ *for some constant* $\varepsilon_0 > 0$. *If* $0 \leqslant x \leqslant \tau n^{-2\varepsilon} t_0^2$, *where* $\tau > 0$ *is fixed and* $\varepsilon > 0$ *is a sufficiently small constant, then for any (large) constant* $D > 0$ *we have*

$$\rho_{1,t}(\lambda_{+,1} - x) = \rho_{2,t}(\lambda_{+,2} - x) \left[1 + \mathrm{O}\left(\frac{n^{\varepsilon} t}{t_0} + n^{-D}\right)\right], \tag{B.25}$$

and

$$|\mathrm{Re}[m_{1,t}(\lambda_{+,1} - x) - m_{1,t}(\lambda_{+,1})] - \mathrm{Re}[m_{2,t}(\lambda_{+,2} - x) - m_{2,t}(\lambda_{+,2})]| \lesssim \left(\frac{n^{\varepsilon}}{t_0} + \frac{n^{-D}}{t}\right) x. \tag{B.26}$$

If $0 \leqslant x \leqslant \tau n^{-2\varepsilon} t_0 t$, *where* $\tau > 0$ *is a sufficiently small constant, the for any (large) constant* $D > 0$ *we have*

$$|\mathrm{Re}\ [m_{1,t}(\lambda_{+,1} + x) - m_{1,t}(\lambda_{+,1})] - \mathrm{Re}\ [m_{2,t}(\lambda_{+,2} + x) - m_{2,t}(\lambda_{+,2})]| \lesssim \left(n^{\varepsilon} \frac{t^{1/2}}{t_0^{1/2}} + n^{-D} \frac{t_0^{1/2}}{t^{1/2}}\right) |x|^{1/2}. \tag{B.27}$$

B.2 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 $(p+n) \times (p+n)$ self-adjoint block matrix

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

Definition B.9 (Resolvents)**.** *We define the resolvent of* H_t *as*

$$G(z) \equiv G_t(X, W, z) := (z^{1/2} H_t - z)^{-1}, \quad z \in \mathbb{C}_+. \tag{B.28}$$

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*

$$\mathcal{G}_1(z) \equiv \mathcal{G}_{1,t}(X, W, z) := (\mathcal{Q}_{1,t} - z)^{-1}, \quad \mathcal{G}_2(z) \equiv \mathcal{G}_{2,t}(X, W, z) := (\mathcal{Q}_{2,t} - z)^{-1}. \tag{B.29}$$

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

$$\rho_1 \equiv \rho_{1,t}(X, W, z) := \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i(\mathcal{Q}_{1,t})}, \quad m_1(z) \equiv m_{1,t}(X, W, z) := \int \frac{1}{x - z} \rho_1(\mathrm{d}x) = \frac{1}{p} \mathrm{Tr}\, \mathcal{G}_1(z). \tag{B.30}$$

Similarly, we denote the empirical spectral density $\rho_{2,t}$ *of* $\mathcal{Q}_{2,t}$ *and its Stieltjes transform as*

$$\rho_2 \equiv \rho_{2,t}(X, W, z) := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(\mathcal{Q}_{2,t})}, \quad m_2(z) \equiv m_{2,t}(X, W, z) := \int \frac{1}{x - z} \rho_2(\mathrm{d}x) = \frac{1}{n} \mathrm{Tr}\, \mathcal{G}_2(z). \tag{B.31}$$

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

$$\mathcal{D}_\vartheta := \left\{ z = E + \mathrm{i}\eta : \lambda_{+,t} - \frac{3}{4}c_V \leqslant E \leqslant \lambda_{+,t}, \frac{n^\vartheta}{n\eta} \leqslant \sqrt{\kappa + \eta} \leqslant 10 \right\} \bigcup \left\{ z = E + \mathrm{i}\eta : \lambda_{+,t} \leqslant E \leqslant \frac{3}{4}c_V, n^{-2/3+\vartheta} \leqslant \eta \leqslant 10 \right\}, \tag{B.32}$$

where recall that $\lambda_{+,t}$ is the right-edge of $\rho_{w,t}$. The local law on the domain $\mathcal{D}_\vartheta$ is stated as follows.

Theorem B.10 (Theorem 2.2 of [23]). *Suppose $V = WW^\top$ is η_*-regular, and t satisfies* (B.10). *For any constant $\vartheta > 0$, the following estimates hold uniformly in $z \in \mathcal{D}_\vartheta$:*

- *for $E \leqslant \lambda_{+,t}$, we have*

$$|m_{1,t}(z) - m_{w,t}(z)| \prec \frac{1}{n\eta}; \tag{B.33}$$

- *for $E \geqslant \lambda_{+,t}$, we have*

$$|m_{1,t}(z) - m_{w,t}(z)| \prec \frac{1}{n(\kappa + \eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa + \eta}}. \tag{B.34}$$

As a consequence of this theorem, we can obtain the following rigidity estimate for the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$ of $\mathcal{Q}_{1,t}$ near the right edge $\lambda_{+,t}$. We define the quantiles of $\rho_{w,t}$ as in (A.14):

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} \rho_{w,t}(x)\mathrm{d}x > \frac{j-1}{p} \right\}, \quad 1 \leqslant j \leqslant p. \tag{B.35}$$

Lemma B.11. *For any j such that $\lambda_{+,t} - c_V/2 < \gamma_j \leqslant \lambda_{+,t}$, we have*

$$|\lambda_j - \gamma_j| \prec j^{-1/3} n^{-2/3}. \tag{B.36}$$

Proof. The estimate (B.36) follows from the estimates (B.33) and (B.34) combined with a standard argument using Helffer-Sjöstrand calculus. The details are already given in [35], [42] and [88]. □

Then we present the local laws for the case where W already satisfies a local law. In this case we can deal with any $t > 0$.

Assumption B.12. *We assume that $m_V(z) \equiv m_{w,0}(z)$ satisfies that for any constant $\vartheta > 0$,*

$$|m_{w,0}(z) - m_c(z)| \prec \frac{1}{n\eta}, \quad \lambda_+ - c_V \leqslant E \leqslant \lambda_+, \quad \frac{n^\vartheta}{n\eta} \leqslant \sqrt{|E - \lambda_+| + \eta} \leqslant 10,$$

and

$$|m_{w,0}(z) - m_c(z)| \prec \frac{1}{n(|E - \lambda_+| + \eta)} + \frac{1}{(n\eta)^2\sqrt{|E - \lambda_+| + \eta}}, \quad \lambda_+ \leqslant E \leqslant \lambda_+ + c_V, \quad n^{-2/3+\vartheta} \leqslant \eta \leqslant 10.$$

Here $m_c(z)$ is the Stieltjes transform of a deterministic law $\rho_c(x)$ that is compactly supported on $[0, \lambda_+]$, and satisfies $\|\rho_c\|_\infty = \mathrm{O}(1)$ and $\rho_c(x) \sim \sqrt{x}$ for $\lambda_+ - c_V \leqslant x \leqslant \lambda_+$.

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

$$\begin{aligned}\mathcal{D}_{\vartheta,c} := &\left\{z : \lambda_{c,t} - \frac{3}{4}c_V \leqslant E \leqslant \lambda_{c,t}, \frac{n^{\vartheta}}{n\eta} \leqslant \sqrt{\kappa_c + \eta} \leqslant 10\right\} \\ &\bigcup \left\{z : \lambda_{c,t} \leqslant E \leqslant \lambda_{c,t} + \frac{3}{4}c_V, n^{-2/3+\vartheta} \leqslant \eta \leqslant 10\right\}.\end{aligned} \tag{B.37}$$

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

Theorem B.13 (Theorem 2.2 of [23])**.** *Suppose Assumption B.12 holds. For any fixed constants $\vartheta, \delta > 0$, the following estimates hold uniformly in $z \in \mathcal{D}_{\vartheta,c}$ and $0 \leqslant t \leqslant n^{-\delta}$:*

- *for $E \leqslant \lambda_{c,t}$, we have*

$$|m_{1,t}(z) - m_{c,t}(z)| \prec \frac{1}{n\eta}; \tag{B.38}$$

- *for $E \geqslant \lambda_{c,t}$, we have*

$$|m_{1,t}(z) - m_{c,t}(z)| \prec \frac{1}{n(\kappa_c + \eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa_c + \eta}}. \tag{B.39}$$

Again using Theorem B.13, we can prove the following rigidity estimates for the eigenvalues of $\mathcal{Q}_{1,t}$ near the right edge $\lambda_{c,t}$. We define the quantiles γ_j as in (B.35) but with $\rho_{w,t}$ replaced by $\rho_{c,t}$.

Lemma B.14. *For any j such that $\lambda_{c,t} - c_V/2 < \gamma_j \leqslant \lambda_{c,t}$, we have*

$$|\lambda_j - \gamma_j| \prec j^{-1/3} n^{-2/3}. \tag{B.40}$$

Proof. The estimate (B.40) follows from the estimates (B.38) and (B.39) combined with a standard argument using Helffer-Sjöstrand calculus. The details are already given in [35], [42] and [88]. □

C Analysis of Dyson Brownian motion

This section is devoted to the proof of Theorem 2.3. In this section, we fix two time scales

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

for some constants ω_0 and ω_1 satisfying $1/3 - \phi_*/2 + \varepsilon/2 \leqslant \omega_0 \leqslant 1/3 - \varepsilon/2$ and $0 < \omega_1 < \omega_0/100$. The reason for choosing these two scales is the same as the one in [67]. 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 $t_0 \leqslant t \leqslant t_0 + t_1$ the sizes of the eigenvalue gaps 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 so called rectangular Dyson Brownian motion (DBM). Let $B_i(t)$, $i = 1, \cdots, p$, be independent standard Brownian motions. For $t \geqslant 0$, we define the process $\{\lambda_i(t) : 1 \leqslant i \leqslant p\}$ as the unique strong solution to the following system of SDEs [17, Appendix C]:

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

with initial data

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

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 eigenvalues gaps of $\mathcal{Q}_{t_0}$ with those of the edge eigenvalues gaps of Wigner matrices. (Recall that the asymptotic density $\rho_{w,t}$ is given by (B.14), while the Wigner semicircle law has density roughly $\pi^{-1}\sqrt{2-x}$ around the right edge at 2.) The 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 free convolution of $\sqrt{\gamma_w}V$ with MP law at $\gamma_w t_0 + t$ 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 DBM $\{\lambda_i(t)\}$ with those of the Wishart matrices. For this purpose, we define another DBM with initial data $\{u_i\}$ being the eigenvalues of a $p \times n$ Wishart matrix $\mathcal{U}\mathcal{U}^\top$, where the entries of $\mathcal{U}$ are i.i.d. Gaussian random variables with mean zero and variance n^{-1}. For $t \geqslant 0$, we define the process $\{\mu_i(t) : 1 \leqslant i \leqslant p\}$ as the unique strong solution to the following system of SDEs:

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

with initial data

$$\mu_i(0) := \mu_i(\gamma_\mu \mathcal{U}\mathcal{U}^\top + a_\mu), \quad \gamma_\mu := c_n^{-1/2}(1+\sqrt{c_n})^{-4/3}, \quad a_\mu := E_\lambda(0) - \gamma_\mu(\sqrt{1+c_n})^2.$$

Here γ_μ is chosen to match the edge eigenvalues gaps of the Marchenko-Pastur law with those of the Wigner semicircle law (recall that the MP law has density approximately $\pi^{-1}c_n^{-3/4}(1+\sqrt{c_n})^{-2}\sqrt{(1+\sqrt{c_n})^2 - x}$ around the right edge at $(1+\sqrt{c_n})^2$), and the shift a_μ is chosen to match the right edge of $E_\lambda(0)$ of $\rho_{\lambda,0}$ at $t = 0$. Without loss of generality, we shall assume that $a_\mu \geqslant 0$, since otherwise we can consider a positive shift of $\mathcal{Q}_t$ instead. For any $t > 0$, the process $\{\mu_i(t)\}$ has the same joint distribution as the eigenvalues of the matrix $(\gamma_\mu + t)XX^\top + a_\mu$ for a Wishart matrix $XX^\top$. In particular, by [56] we know that the edge eigenvalues of $\{\mu_i(t)\}$ obey the Tracy-Widom distribution asymptotically. In the following proof, we shall denote the MP law for $(\gamma_\mu + t)XX^\top + a_\mu$ as

$$\rho_{\mu,t} := \frac{\sqrt{[x - a_\mu - (\gamma_\mu + t)(1-\sqrt{c_n})^2][(\gamma_\mu + t)(1+\sqrt{c_n})^2 + a_\mu - x]}}{2\pi c_n (x - a_\mu)(\gamma_\mu + t)}, \tag{C.2}$$

whose right edge is given by

$$E_\mu(t) := (\gamma_\mu + t)(1+\sqrt{c_n})^2 + a_\mu = E_\lambda(0) + t(1+\sqrt{c_n})^2.$$

Finally, we denote the Stieltjes transform of $\rho_{\mu,t}$ by $m_{\mu,t}$.

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

Theorem C.1. *Fix any integer $\mathfrak{a} \in \mathbb{N}$. There exists a constant $\varepsilon > 0$ such that*

$$\max_{1\leqslant i\leqslant \mathfrak{a}} |(\lambda_i(t_1) - E_\lambda(t)) - (\mu_i(t_1) - E_\mu(t_1))| \leqslant n^{-2/3-\varepsilon}$$

with high probability.

It is easy to see that Theorem C.1 implies Theorem 2.3 immediately.

Proof of Theorem 2.3. We take $t_0 = t - t_1$ for a small enough constant ω_1. Then Theorem C.1 gives (2.8) together with the fact that $\mu_i(t_1) - E_\mu(t_1)$ satisfies the Tracy-Widom fluctuation by [56]. ☐

C.1 Interpolating processes

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

$$dz_i(t,\alpha) = 2\sqrt{z_i(t,\alpha)}\frac{\mathrm{d}B_i}{\sqrt{n}} + \left(\frac{1}{n}\sum_{j\neq i}\frac{z_i(t,\alpha)+z_j(t,\alpha)}{z_i(t,\alpha)-z_j(t,\alpha)} + 1\right)\mathrm{d}t, \quad 1 \leqslant i \leqslant p, \tag{C.3}$$

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

$$\widetilde{m}_t(z,\alpha) := \frac{1}{p}\sum_{i=1}^{p}\frac{1}{z_i(t,\alpha) - z}. \tag{C.4}$$

Note that by Lemma B.5, due to the choices of γ_μ and γ_w we have that

$$\rho_{\lambda,0}(E_\lambda(0) - E) = \rho_{\mu,0}(E_\mu(0) - E)\left[1 + \mathrm{O}\left(\frac{|E|}{t_0^2}\right)\right], \quad 0 \leqslant E \leqslant \tau t_0^2, \tag{C.5}$$

for a sufficiently small constant $\tau > 0$. Let $\gamma_{\mu,i}(t)$ and $\gamma_{\lambda,i}(t)$ be the quantiles of $\rho_{\mu,t}$ and $\rho_{\lambda,t}$ defined through:

$$\gamma_{\mu,i}(t) := \sup_x\left\{\int_x^{+\infty}\rho_{\mu,t}(x)\mathrm{d}x > \frac{i-1}{p}\right\}, \quad \gamma_{\lambda,i}(t) := \sup_x\left\{\int_x^{+\infty}\rho_{\lambda,t}(x)\mathrm{d}x > \frac{i-1}{p}\right\}. \tag{C.6}$$

By Theorem B.10, we know that $|\widetilde{m}_0(z,0) - m_{\mu,0}(z)|$ and $|\widetilde{m}_0(z,1) - m_{\lambda,0}(z)|$ satisfy Assumption B.12. Hence using Lemma B.14, we obtain that for $k_* := c_* n$ for some constant $c_* > 0$ depending on c_V, we have

$$\sup_{0\leqslant t\leqslant 10t_1}(|z_i(t,0) - \gamma_{\mu,i}(t)| + |z_i(t,1) - \gamma_{\lambda,i}(t)|) \prec i^{-1/3}n^{-2/3}, \quad 1 \leqslant i \leqslant k_*. \tag{C.7}$$

Here 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 some large constant $C > 0$, by Lemma B.14 and a simple union bound we get that

$$\sup_{0\leqslant k\leqslant n^C}(|z_i(t_k,0) - \gamma_{\mu,i}(t_k)| + |z_i(t_k,1) - \gamma_{\lambda,i}(t_k)|) \prec i^{-1/3}n^{-2/3}. \tag{C.8}$$

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 $t_k \leqslant t \leqslant t_{k+1}$ using a simple continuity estimate. We refer the reader to Appendix B of [65] for more details.

Combining (C.5) and (C.6), we can get the following simple control on the quantiles near the edge.

Lemma C.2. *For $i = \mathrm{O}(n^{6\omega_0/5})$, we have that*

$$|\gamma_{\mu,i}(0) - \gamma_{\lambda,i}(0))| \lesssim \frac{i^{4/3}}{n^{2\omega_0} n^{2/3}}. \tag{C.9}$$

Proof. For simplicity, we denote $x := \gamma_{\mu,i}(0) - E_\mu(0)$ and $y := \gamma_{\lambda,i}(0) - E_\lambda(0)$. Without loss of generality, we assume that $x \leqslant y$. Note that by the square root behaviors of $\rho_{\mu,0}$ and $\rho_{\lambda,0}$ near the right edges, it is easy to see that $x \sim y \sim i^{2/3} n^{-2/3}$. Now using (C.5) and (C.6), we get that

$$\int_0^x \left[\rho_{\mu,0}(E_\mu(0) - E) - \rho_{\lambda,0}(E_\lambda(0) - E)\right] \mathrm{d}E = \int_x^y \rho_{\lambda,0}(E_\lambda(0) - E) \mathrm{d}E \ \Rightarrow \ |y^{3/2} - x^{3/2}| \lesssim \frac{x^{5/2}}{t_0^2}.$$

This implies $|y - x| \lesssim x^2/t_0^2$, which concludes the proof using $x \sim i^{2/3} n^{-2/3}$ 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 edges. Our main goal is to construct a density for each $0 \leqslant \alpha \leqslant 1$, which matches the distribution of $\{z_i(0,\alpha)\}$, and with which we can take a free rectangular convolution for any $0 \leqslant t \leqslant t_1$.

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 $E_\mu(0) - \tau \leqslant y < E_\mu(0)$ and $\rho_{\lambda,0}(y) > 0$ for $E_\lambda(0) - \tau \leqslant y < E_\lambda(0)$ for 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. continuous versions of quantiles) $\varphi_\mu(s)$ and $\varphi_\lambda(s)$ as

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

for some small constant $c_* > 0$. 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

$$\mathtt{D}_\alpha := [\alpha\varphi_\mu(c_*) + (1-\alpha)\varphi_\lambda(c_*), E_\lambda(0)]. \tag{C.10}$$

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

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

with which we define the asymptotic density as

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

By inverse function theorem, we have that

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

Together with (C.5), we immediately find that

$$\rho(E_+(0,\alpha)-E,\alpha)=\rho_{\mu,0}(E_\mu(0)-E)\left[1+\mathrm{O}\left(\frac{|E|}{t_0^2}\right)\right],\quad 0\leqslant E\leqslant \tau t_0^2, \tag{C.11}$$

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

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

This measure is defined in a way such that its Stieltjes transform is close to $\widetilde{m}_0(z,\alpha)$ defined in (C.4). 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 away from 0, so 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.

We let $\rho_t(E,\alpha)$ be the free rectangular convolution with the MP law at time t and define its Stieltjes transform as $m_t(z,\alpha)$. The properties of $\rho_t(E,\alpha)$ and $m_t(z,\alpha)$ have been studied in Section B. In particular, $\rho_t(E,\alpha)$ has a square root behavior near the right edge, which we will denote as $E_+(t,\alpha)$. Although $\rho_t(E,\alpha)$ is random, with the results in Section B we can provide a deterministic control on it.

Lemma C.3. *Let $\varepsilon,\tau>0$ be sufficiently small constants. For $0\leqslant E\leqslant \tau n^{-2\varepsilon}t_0^2$, we have that for any constant $D>0$,*

$$\rho_t(E_+(t,\alpha)-E,\alpha)=\rho_{\mu,t}(E_\mu(t)-E)(1+\mathrm{O}(n^\varepsilon t/t_0+n^{-D})). \tag{C.12}$$

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

$$\max_{1\leqslant i\leqslant c_\tau n^{1-3\varepsilon_0}t_0^3}|\widetilde{\gamma}_i(t,\alpha)-\widetilde{\gamma}_i(t,0)|\leqslant\left(n^\varepsilon\frac{t}{t_0}+n^{-D}\right)\frac{i^{2/3}}{n^{2/3}}, \tag{C.13}$$

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

Proof. The estimates (C.12) follows directly from (B.25). The estimate (C.13) follows from (C.12) using the same arguments as in the proof of (C.9). □

By the eigenvalues rigidity (C.7) and the construction of $\mathrm{d}\mu(E,\alpha)$, we can verify that $|m_0(z,\alpha)-\widetilde{m}_0(z,\alpha)|$ satisfies the Assumption B.12. Then by Lemma B.14, we have the following rigidity of eigenvalues for $\{z_i(t,\alpha)\}$. As before, 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 C.4. *There exists a constant $c_*>0$ so that*

$$\sup_{0\leqslant\alpha\leqslant1}\sup_{0\leqslant t\leqslant 10t_1}|z_i(t,\alpha)-\gamma_i(t,\alpha)|\prec i^{-1/3}n^{-2/3},\quad 1\leqslant i\leqslant c_* n. \tag{C.14}$$

Proof. This estimate follows from Lemma B.14 combined with a standard stochastic continuity argument. □

Recalling (B.9) and using the estimates in Section B, we can calculate that

$$\begin{aligned}\frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t} &= \left[\frac{(1-c_n)}{2\zeta_t(E_+(t,\alpha),\alpha)} - \frac{c_n m_t(E_+(t,\alpha),\alpha)}{b_t(E_+(t,\alpha),\alpha)}\right]\sqrt{\zeta_t(E_+(t,\alpha),\alpha)} - \frac{(1-c_n)^2 t}{4\zeta_t(E_+(t,\alpha),\alpha)\sqrt{E_+(t,\alpha)}} + \mathrm{O}(t^2) \\ &= \frac{(1-c_n)}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}} - c_n m_t(E_+(t,\alpha),\alpha)\sqrt{E_+(t,\alpha)-t(1-c_n)} - \frac{(1-c_n)^2 t}{4E_+^{3/2}(t,\alpha)} + \mathrm{O}(t^2), \end{aligned} \tag{C.15}$$

where $b_t(z,\alpha) := 1 + c_n t m_t(z,\alpha)$, $\zeta_t(z,\alpha) := z b_t^2(z,\alpha) - t(1-c_n)b_t(z,\alpha)$. In the proof, we will need to use the following function

$$\begin{aligned}\Psi_t(x,\alpha) := {} & \frac{(1-c_n)}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}} - \frac{(1-c_n)}{2\sqrt{E_+(t,\alpha)-x}} - \frac{(1-c_n)^2 t}{4E_+^{3/2}(t,\alpha)} \\ & - \operatorname{Re}\left[c_n m_t(E_+(t,\alpha),\alpha)\sqrt{E_+(t,\alpha)-t(1-c_n)} - c_n m_t(E_+(t,\alpha)-x,\alpha)\sqrt{E_+(t,\alpha)-x}\right],\end{aligned} \tag{C.16}$$

for $-\tau \leqslant x \leqslant \tau$ for some small constant $\tau > 0$. Next we prove the following matching estimate for the function $\Psi_t(x,\alpha)$, which is crucial for our proof. We remark that the proof explores a rather delicate cancellation in $\Psi_t(x,\alpha)$, and hence is much more nontrivial than the one for the corresponding results, Lemma 3.4, of [67].

Lemma C.5. *Let $\varepsilon, \tau > 0$ be a sufficiently small constants. For $0 \leqslant E \leqslant \tau n^{-2\varepsilon} t_0^2$, we have that for any constant $D > 0$,*

$$|\Psi_t(E,\alpha) - \Psi_t(E,0)| \lesssim \left(\frac{n^\varepsilon}{t_0} + \frac{n^{-D}}{t}\right) E + t^2. \tag{C.17}$$

For $0 \leqslant E \leqslant \tau n^{-2\varepsilon} t t_0$, we have

$$|\Psi_t(-E,\alpha) - \Psi_t(-E,0)| \lesssim \left(n^\varepsilon \frac{t^{1/2}}{t_0^{1/2}} + n^{-D}\frac{t_0^{1/2}}{t^{1/2}}\right) E^{1/2} + t^2. \tag{C.18}$$

Proof. First, we claim that

$$\Psi_t(E,\alpha) = \widetilde{\Psi}_t(E,\alpha) + \mathrm{O}(t^2), \tag{C.19}$$

where

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

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

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

$$= \frac{(1-c_n)[E_+(t,\alpha) - b_t^2(E_+(t,\alpha),\alpha)\cdot E_+(t,\alpha) + (1-c_n)tb_t(E_+(t,\alpha),\alpha)]}{2\sqrt{\zeta_t(E_+(t,\alpha),\alpha)}\sqrt{E_+(t,\alpha)}\left(\sqrt{E_+(t,\alpha)} + \sqrt{\zeta_t(E_+(t,\alpha),\alpha)}\right)}$$
$$+ \frac{(1-c_n)c_n t m_t(E_+(t,\alpha),\alpha)}{\sqrt{E_+(t,\alpha)} + \sqrt{E_+(t,\alpha) - t(1-c_n)}} - \frac{(1-c_n)^2 t}{4E_+^{3/2}(t,\alpha)}$$
$$= \frac{-2(1-c_n)c_n t m_t(E_+(t,\alpha),\alpha)\cdot E_+(t,\alpha) + (1-c_n)^2 t}{4E_+^{3/2}(t,\alpha)} + \frac{(1-c_n)c_n t m_t(E_+(t,\alpha),\alpha)}{2\sqrt{E_+(t,\alpha)}} - \frac{(1-c_n)^2 t}{4E_+^{3/2}(t,\alpha)} + \mathrm{O}(t^2)$$
$$= \mathrm{O}(t^2).$$

On the other hand, using (B.24) we get

$$\widetilde{\Psi}_t(E,\alpha) - \widetilde{\Psi}_t(E,0) = -c_n \operatorname{Re}\left[m_t(E_+(t,\alpha),\alpha) - m_t(E_+(t,\alpha)-E,\alpha)\right]\sqrt{E_+(t,\alpha)}$$
$$+ c_n \operatorname{Re}\left[m_t(E_+(t,0),0) - m_t(E_+(t,0)-E,0)\right]\sqrt{E_+(t,0)} + \mathrm{O}(E)$$
$$= c_n \operatorname{Re}\left[(m_t(E_+(t,0),0) - m_t(E_+(t,0)-E,0)) - (m_t(E_+(t,\alpha),\alpha) - m_t(E_+(t,\alpha)-E,\alpha))\right]\sqrt{E_+(t,\alpha)}$$
$$+ \mathrm{O}\left(t\left|m_t(E_+(t,0),0) - m_t(E_+(t,0)-E,0)\right| + E\right).$$

By (B.19), we have

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

and by (B.17) and (B.18), we have

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

Together with Lemma B.8, we can conclude (C.17) and (C.18). □

Remark C.6. Later we will consider the evolution after $t = N^{-C}$ for some large constant $C > 0$, so that the n^{-D} terms in (C.17) and (C.18) are negligible as long as D is taken 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.7), 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 B.12. Moreover by Theorem B.10, 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 B.12. Hence we obtain that

$$|m_0(z,0) - m_{\mu,0}(z)| \prec \frac{1}{n\eta}, \quad |E - E_\mu(0)| \leqslant \frac{3}{4}c_V,\ n^{-2/3+\vartheta} \leqslant \eta \leqslant 10. \tag{C.20}$$

and

$$|m_0(z,1) - m_{\lambda,0}(z)| \prec \frac{1}{n\eta}, \quad |E - E_\lambda(0)| \leqslant \frac{3}{4}c_V,\ n^{-2/3+\vartheta} \leqslant \eta \leqslant 10. \tag{C.21}$$

With the above estimates, we can control $|E_+(t,1) - E_\lambda(t)|$ and $|E_+(t,0) - E_\mu(t)|$ for $0 \leqslant t \leqslant 10t_1$.

Lemma C.7. *We have that*

$$\max_{0\leqslant t\leqslant 10t_1} |E_+(t,1) - E_\lambda(t)| \prec t^3 + \frac{t}{\sqrt{n}}, \tag{C.22}$$

and

$$\max_{0\leqslant t\leqslant 10t_1} |E_+(t,0) - E_\mu(t)| \prec t^3 + \frac{t}{\sqrt{n}}. \tag{C.23}$$

Proof. First repeating the proof of Lemma B.7 (which was given in Lemma 3.20 of [23]) but with t_0 replaced by 1, we get that

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

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

$$|E_+(t,1) - E_\lambda(t)| \lesssim |\zeta_{+,1} - \zeta_{+,\lambda}| + t|m_0(\zeta_{+,1},1) - m_{\lambda,0}(\zeta_{+,\lambda})|. \tag{C.24}$$

Recall that $\zeta_{+,1} - E_\lambda(0) \sim t^2$ and $\zeta_{+,\lambda} - E_\lambda(0) \sim t^2$ by (B.12). Then using (B.17) and (B.18), we get $|m'_{\lambda,0}(\zeta)| \lesssim t^{-1}$ for ζ between $\zeta_{+,1}$ and $\zeta_{+,\lambda}$. Thus we can bound (C.24) by

$$\begin{aligned}|E_+(t,1) - E_\lambda(t)| &\lesssim t^3 + t|m_0(\zeta_{+,1},1) - m_{\lambda,0}(\zeta_{+,1})| + t|m_{\lambda,0}(\zeta_{+,1}) - m_{\lambda,0}(\zeta_{+,\lambda})| \\ &\lesssim t^3 + t|m_0(\zeta_{+,1},1) - m_{\lambda,0}(\zeta_{+,1})|. \end{aligned}\tag{C.25}$$

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

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

with high probability. Together with (C.21) for $[m_0(\zeta_{+,1} + \mathrm{i}n^{-1/2},1) - m_{\lambda,0}(\zeta_{+,1} + \mathrm{i}n^{-1/2})]$, we get

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

Plugging it into (C.25), we conclude (C.22). The estimate (C.23) 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 densities for $y_i(t,\alpha)$ are given by

$$f_t(E,\alpha) := 2E\rho_t(E^2,\alpha).$$

Similarly we can define $f_{\lambda,t}$ and $f_{\mu,t}$. Moreover, the quantiles of $f_t(E,\alpha)$ are exactly $\sqrt{\gamma_i(t,\alpha)}$. Now with Lemma C.3 and Lemma C.4, we can easily conclude the following lemma.

Lemma C.8. *We have the following rigidity estimate of singular values:*

$$\sup_{0\leqslant\alpha\leqslant1}\ \sup_{0\leqslant t\leqslant 10t_1}\left|y_i(t,\alpha) - \sqrt{\gamma_i(t,\alpha)}\right| \prec n^{-2/3} i^{-1/3}, \quad 1 \leqslant i \leqslant c_* n. \tag{C.26}$$

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

$$f_t\left(\sqrt{E_+(t,\alpha)} - E, \alpha\right) = f_{\mu,t}\left(\sqrt{E_\mu(t)} - E\right)(1 + \mathrm{O}(n^\varepsilon t/t_0 + n^{-D})), \tag{C.27}$$

and

$$\max_{1\leqslant i\leqslant c_\tau n^{1-3\varepsilon_0} t_0^3} |\widehat{\gamma}_i(t,\alpha) - \widehat{\gamma}_i(t,0)| \leqslant \left(n^\varepsilon \frac{t}{t_0} + n^{-D}\right)\frac{i^{2/3}}{n^{2/3}}, \tag{C.28}$$

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.26) follows directly from Lemma C.4, (C.27) follows from (C.12) together with $|E_+(t,\alpha) - E_\lambda(0)| = \mathrm{O}(t)$ by (B.24), and (C.28) can be proved easily using (C.27). □

C.2 Short-range approximation

As in [67], 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 SDE for the singular values $y_i(t,\alpha) = \sqrt{z_i(t,\alpha)}$. By Ito's formula, we have that for $1 \leqslant i \leqslant p$,

$$\begin{aligned} \mathrm{d}y_i(t,\alpha) &= \frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2y_i(t,\alpha)}\left(\frac{1}{n}\sum_{j\neq i}\frac{y_i^2(t,\alpha)+y_j^2(t,\alpha)}{y_i^2(t,\alpha)-y_j^2(t,\alpha)} + \frac{n-1}{n}\right)\mathrm{d}t \\ &= \frac{\mathrm{d}B_i}{\sqrt{n}} + \left(\frac{1}{2n}\sum_{j\neq i}\frac{1}{y_i(t,\alpha)-y_j(t,\alpha)} + \frac{1}{2n}\sum_{j\neq i}\frac{1}{y_i(t,\alpha)+y_j(t,\alpha)} + \frac{n-p}{2ny_i(t,\alpha)}\right)\mathrm{d}t. \end{aligned} \tag{C.29}$$

Note that the diffusion term now has a constant coefficient. We remark that for Dyson Brownian motion for the eigenvalues of Wigner type random matrices [16, 17, 65, 66, 67], the diffusion coefficients are already constant and hence they do not need to use this extra step. For convenience, we introduce the shifted processes

$$\widetilde{z}_i(t,\alpha) := E_+(t,\alpha) - z_i(t,\alpha), \quad \widetilde{y}_i(t,\alpha) := \sqrt{E_+(t,\alpha)} - y_i(t,\alpha). \tag{C.30}$$

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

$$\begin{aligned} \mathrm{d}\widetilde{y}_i(t,\alpha) = &-\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n}\sum_{j\neq i}\frac{1}{\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}\mathrm{d}t - \frac{1}{2n}\sum_{j\neq i}\frac{1}{2\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}\mathrm{d}t \\ &- \frac{n-p}{2n(\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha))}\mathrm{d}t + \frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t}\mathrm{d}t, \end{aligned} \tag{C.31}$$

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

$$\mathcal{A} := \left\{(i,j) : |i-j| \leqslant \ell(10\ell^2 + i^{2/3} + j^{2/3})\right\} \bigcup \left\{(i,j) : i,j > i_*/2\right\}, \quad i_* := c_* n, \tag{C.32}$$

where c_* is the constant as appeared in Lemma C.4. It is easy to see 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

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

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

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

where we recall that $\widetilde{\gamma}_i(t,\alpha)$ and $\widehat{\gamma}_i(t,\alpha)$ are defined below (C.13) and (C.28), respectively. Finally, we denote

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

where $\widetilde{c}_V > 0$ is some 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 of the following SDEs for $t \geqslant n^{-C_0}$ with the same initial data (recall Remark C.6)

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

where C_0 is an absolute constant (for example $C_0 = 100$ will be more than enough). Corresponding to (C.30), we denote

$$\widehat{z}_i(t, \alpha) := E_+(t, \alpha) - (\sqrt{E_+(t, \alpha)} - \widehat{y}_i(t, \alpha))^2. \tag{C.33}$$

For $1 \leqslant i \leqslant n^{\omega_a}$, the SDEs are

$$\begin{aligned} \mathrm{d}\widehat{y}_i(t, \alpha) = & -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n} \sum_j^{\mathcal{A},(i)} \frac{1}{\widehat{y}_i(t, \alpha) - \widehat{y}_j(t, \alpha)} \mathrm{d}t - \frac{n-p}{2n\sqrt{E_+(t, 0)}} \mathrm{d}t + \frac{\mathrm{d}\sqrt{E_+(t, 0)}}{\mathrm{d}t} \mathrm{d}t \\ & - \left[c_n \int_{\mathcal{I}_i^c(t,0)} \frac{\sqrt{E_+(t, 0)}\rho_t(E_+(t, 0) - E, 0)}{E - E_+(t, 0) + (\sqrt{E_+(t, 0)} - \widehat{y}_i(t, \alpha))^2} \mathrm{d}E \right] \mathrm{d}t; \end{aligned} \tag{C.34}$$

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

$$\begin{aligned} \mathrm{d}\widehat{y}_i(t, \alpha) = & -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n} \sum_j^{\mathcal{A},(i)} \frac{\mathrm{d}t}{\widehat{y}_i(t, \alpha) - \widehat{y}_j(t, \alpha)} + \left[\frac{c_n}{2} \int_{\widehat{\mathcal{I}}_i^c(t,\alpha) \cap \mathcal{J}(t,\alpha)} \frac{f_t(\sqrt{E_+(t, \alpha)} - E, \alpha)}{\widehat{y}_i(t, \alpha) - E} \mathrm{d}E \right] \mathrm{d}t \\ & + \frac{1}{2n} \sum_{j \geqslant 3i_* 4} \frac{\mathrm{d}t}{\widetilde{y}_i(t, \alpha) - \widetilde{y}_j(t, \alpha)} - \frac{n-p}{2n(\sqrt{E_+(t, \alpha)} - \widetilde{y}_i(t, \alpha))} \mathrm{d}t \\ & - \frac{1}{2n} \sum_{j \neq i} \frac{1}{2\sqrt{E_+(t, \alpha)} - \widetilde{y}_i(t, \alpha) - \widetilde{y}_j(t, \alpha)} \mathrm{d}t + \frac{\mathrm{d}\sqrt{E_+(t, \alpha)}}{\mathrm{d}t} \mathrm{d}t; \end{aligned} \tag{C.35}$$

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

$$\begin{aligned} \mathrm{d}\widehat{y}_i(t, \alpha) = & -\frac{\mathrm{d}B_i}{\sqrt{n}} + \frac{1}{2n} \sum_j^{\mathcal{A},(i)} \frac{\mathrm{d}t}{\widehat{y}_i(t, \alpha) - \widehat{y}_j(t, \alpha)} + \frac{1}{2n} \sum_j^{\mathcal{A}^c,(i)} \frac{\mathrm{d}t}{\widetilde{y}_i(t, \alpha) - \widetilde{y}_j(t, \alpha)} + \frac{\mathrm{d}\sqrt{E_+(t, \alpha)}}{\mathrm{d}t} \mathrm{d}t \\ & - \frac{n-p}{2n(\sqrt{E_+(t, \alpha)} - \widetilde{y}_i(t, \alpha))} \mathrm{d}t - \frac{1}{2n} \sum_{j \neq i} \frac{1}{2\sqrt{E_+(t, \alpha)} - \widetilde{y}_i(t, \alpha) - \widetilde{y}_j(t, \alpha)} \mathrm{d}t. \end{aligned} \tag{C.36}$$

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

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

Then we choose the constants ω_0, ω_1, ω_ℓ and ω_a such that

$$0 < \mathfrak{C}^{-1}\omega_1 \leqslant \omega_\ell \leqslant \mathfrak{C}\omega_a \leqslant \mathfrak{C}^2\omega_0 \leqslant \mathfrak{C}^{-1} \tag{C.37}$$

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. Second, by choosing scale $\omega_a \ll \omega_0$, the density $\rho_t(E_+(t, \alpha) - E, \alpha)$ is approximately α independent (after taking into account of other drift terms) for the SDEs with $1 \leqslant i \leqslant \omega_a$, so that it can be replaced with an α-independent quantity.

Next we show that $\widetilde{y}_i(t,\alpha)$ are good approximations for $\widehat{y}_i(t,\alpha)$. Before that, we recall the semigroup approach for first order 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 $\{T(t) : t \geqslant 0\}$ in $\mathcal{L}(\Omega)$ is a semigroup if

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

For detailed discussion for semigroups of operators, we refer the readers to [11].

Definition C.9. *For any operator $\mathcal{W}$ in $\mathbb{R}^p$, 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, i.e.,*

$$\partial_t \mathcal{U}^{\mathcal{W}}(s,t) = -\mathcal{W}(t)\mathcal{U}^{\mathcal{W}}(s,t),$$

for any $t \geqslant s$ and $\mathcal{U}^{\mathcal{W}}(s,s) = \mathrm{id}$.

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

Lemma C.10. *With high probability, we have that for any constant $\varepsilon > 0$,*

$$\sup_{0 \leqslant \alpha \leqslant 1} \sup_{n^{-C_0} \leqslant t \leqslant 10 t_1} \max_{1 \leqslant i \leqslant p} |\widetilde{y}_i(t,\alpha) - \widehat{y}_i(t,\alpha)| \leqslant n^{-2/3+\varepsilon+\omega_1-2\omega_\ell}. \tag{C.38}$$

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

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

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

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

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

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

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

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

The term $\mathcal{E}$ contains the remaining errors, and we will need to control its l^∞ norm.

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

$$\sup_{0\leqslant\alpha\leqslant 1}\max_{1\leqslant i\leqslant i_*/2}|\widetilde{y}_i(t,\alpha)-\widehat{y}_i(t,\alpha)|\leqslant n^{-2/3},\quad \text{for } n^{-C_0}\leqslant t\leqslant 10t_1. \tag{C.39}$$

Later we will remove it with a simple continuity argument. Note that $\mathcal{V}_1(i)\leqslant 0$, hence the operator $\mathcal{V}_1$ is negative. Then the semigroup of $\mathcal{B}_1+\mathcal{V}_1$ is a contraction on every l^p space. To see this, for $u(s)=\mathcal{U}^{\mathcal{B}_1+\mathcal{V}_1}(0,s)u_0$, we can calculate that for $p\geqslant 1$,

$$\begin{aligned}\partial_t\sum_i|u_i(s)|^p &= \sum_i |u_i(s)|^{p-1}\operatorname{sgn}(u_i(s))\left[(\mathcal{B}_1u(s))_i+\mathcal{V}_1(i)u_i(s)\right]\leqslant\sum_i|u_i(s)|^{p-1}\operatorname{sgn}(u_i(s))\,(\mathcal{B}_1u(s))_i\\ &= -\frac{1}{4n}\sum_{i,j\in\mathcal{A}}\frac{\left[|u_i(s)|^{p-1}\operatorname{sgn}(u_i(s))-|u_j(s)|^{p-1}\operatorname{sgn}(u_j(s))\right](u_i(s)-u_j(s))}{(\widetilde{z}_i-\widetilde{z}_j)(\widehat{z}_i-\widehat{z}_j)}\leqslant 0.\end{aligned}$$

For the l^∞ space, we just need to use $\|u\|_\infty=\lim_{p\to\infty}\|u\|_p$ on $[\![1,p]\!]$. Then 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

$$\|v(t)\|_\infty\leqslant\int_{n^{-C_0}}^t\|\zeta(s)\|_\infty\mathrm{d}s. \tag{C.40}$$

Next we provide some bounds on $\|\zeta(s)\|_\infty$. Fist we have that $\zeta_i(t)=0$ for $i\geqslant i_*/2$. Then under (C.39), for $n^{\omega_a}<i\leqslant i_*/2$, we have

$$\zeta_i(t)=\frac{1}{2n}\sum_{j\leqslant 3i_*/4}^{\mathcal{A}^c,(i)}\frac{1}{\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)}-\frac{c_n}{2}\int_{\widehat{\mathcal{I}}_i^c(t,\alpha)\cap\mathcal{J}(t,\alpha)}\frac{f_t(\sqrt{E_+(t,\alpha)}-E,\alpha)}{\widetilde{y}_i(t,\alpha)-E}\mathrm{d}E. \tag{C.41}$$

We now bound ζ_i for $n^{\omega_a}\leqslant i\leqslant i_*/2$ using the rigidity of singular values (C.26). Decomposing the integral in (C.41) according to the quantiles of f_t as $\sum_j\int_{\gamma_{j+1}}^{\gamma_j}$ and using (C.26), we obtain that for any constant $\varepsilon>0$,

$$|\zeta_i|\leqslant\frac{n^\varepsilon}{n^{5/3}}\sum_{j\leqslant 3i_*/4}^{\mathcal{A}^c,(i)}\frac{1}{(\widehat{\gamma}_i-\widehat{\gamma}_j)^2j^{1/3}}\leqslant\frac{Cn^\varepsilon}{n^{1/3}}\sum_{j\leqslant 3i_*/4}^{\mathcal{A}^c,(i)}\frac{i^{2/3}+j^{2/3}}{(i-j)^2j^{1/3}} \tag{C.42}$$

with high probability, where for the second inequality we used

$$|\widehat{\gamma}_i-\widehat{\gamma}_j|\sim|i^{2/3}-j^{2/3}|n^{-2/3}\gtrsim|i-j|(i+j)^{-1/3}n^{-2/3},\quad\text{for } (i,j)\notin\mathcal{A}.$$

Using the inequalities (3.67) and (3.68) of [67], we can bound (C.42) by

$$|\zeta_i|\leqslant C\frac{n^\varepsilon}{n^{1/3}n^{2\omega_\ell}}\quad\text{with high probability.} \tag{C.43}$$

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

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

$$
\begin{aligned}
&- c_n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i^c(t,\alpha)}\frac{\rho_t(E_+(t,\alpha)-E,\alpha)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E + c_n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}\int_{\mathcal{I}_i^c(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&+ c_n\left(\sqrt{E_+(t,0)}-\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}\right)\int_{\mathcal{I}_i^c(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E \\
&+ c_n\sqrt{E_+(t,0)}\left[\int_{\mathcal{I}_i^c(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widetilde{y}_i(t,\alpha))^2}\mathrm{d}E - \int_{\mathcal{I}_i^c(t,0)}\frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right] \\
&- \frac{n-p}{2n\sqrt{E_+(t,\alpha)-\widetilde{z}_i(t,\alpha)}} + \frac{n-p}{2n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}} + \frac{\mathrm{d}\sqrt{E_+(t,\alpha)}}{\mathrm{d}t} - \frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t} \\
&- \frac{n-p}{2n\sqrt{E_+(t,0)-\widetilde{z}_i(t,\alpha)}} + \frac{n-p}{2n\sqrt{E_+(t,0)}} - \frac{1}{2n}\sum_j^{\mathcal{A},(i)}\frac{1}{2\sqrt{E_+(t,\alpha)}-\widetilde{y}_i(t,\alpha)-\widetilde{y}_j(t,\alpha)} \\
&=: A_1 + A_2 + A_3 + A_4 + A_5 + A_6,
\end{aligned}
$$

where we recall (C.30) and (C.33).

First, for term A_6 we notice that for $1 \leqslant i \leqslant n^{\omega_a}$ there are at most $n^{2\omega_a/3+\omega_\ell}$ indices j such that $(i,j) \in \mathcal{A}$. On the other hand, by the rigidity estimate (C.14) and (C.39), we have $|\widetilde{z}_i(t,\alpha)| \leqslant n^{-2/3+2\omega_a/3+\varepsilon}$ with high probability. Hence we can bound

$$
|A_6| \leqslant n^{-2/3+\omega_a}, \quad \text{with high probability.} \tag{C.44}
$$

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

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

which immediately gives that

$$
|A_3| \lesssim |\widetilde{z}_i(t,\alpha)| \leqslant n^{-2/3+\omega_a}, \quad \text{with high probability.} \tag{C.45}
$$

For A_4, we have

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

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

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

Thus we can bound the integral 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 (B.24) and the rigidity (C.26) for $|\widetilde{y}_i(t,\alpha)|$, we get that for any constant $\varepsilon > 0$,

$$
|A_4| \lesssim t|\widetilde{y}_i(t,\alpha)|n^{1/3-\omega_\ell} \lesssim n^{-1/3+\omega_1}\left(i^{2/3}n^{-2/3} + i^{-1/3}n^{-2/3+\varepsilon}\right)n^{1/3-\omega_\ell} \leqslant n^{-2/3+\omega_a}, \tag{C.46}
$$

with high probability. The term A_1 can be handled in exactly the same way as B_1 in (3.71) of [67] and we get that for any constant $\varepsilon > 0$,

$$|A_1| \leqslant n^{-1/3-2\omega_\ell+\varepsilon} + n^{-1/2+\varepsilon}, \quad \text{with high probability.} \tag{C.47}$$

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

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

where we used (C.15), (C.16) and $|E_+(t,\alpha) - E_+(t,0)| = \mathrm{O}(t)$ in the second step. Using Lemma C.5, we can bound that for any constant $\varepsilon > 0$,

$$|B_2| \leqslant \frac{n^\varepsilon t^{1/2}}{t_0^{1/2}}|\widetilde{z}_i(t,\alpha)|^{1/2} + \frac{n^\varepsilon|\widetilde{z}_i(t,\alpha)|}{t_0} + \mathrm{O}(t^2) \lesssim n^{-1/3+\varepsilon+\omega_a/3+\omega_1/2-\omega_0/2}, \tag{C.48}$$

with high probability, where we used that $|\widetilde{z}_i(t,\alpha)| \lesssim n^{-2/3+2\omega_a/3}$ by rigidity (C.14) because the largest index i_+ is at most $\mathrm{O}(n^{\omega_a})$.

It remains to bound B_1:

$$\begin{aligned}
B_1 &= c_n\sqrt{E_+(t,\alpha) - \widetilde{z}_i(t,\alpha)}\left[\int_{\mathcal{I}_i(t,\alpha)} \frac{\rho_t(E_+(t,\alpha)-E,\alpha)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E - \int_{\mathcal{I}_i(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right] \\
&+ c_n\left[\sqrt{E_+(t,\alpha) - \widetilde{z}_i(t,\alpha)} - \sqrt{E_+(t,0) - \widetilde{z}_i(t,\alpha)}\right]\int_{\mathcal{I}_i(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E =: B_{11} + B_{12}.
\end{aligned}$$

The term B_{11} can be bounded as B_3 in (3.82) of [67] and we get that for any constant $\varepsilon > 0$,

$$|B_{11}| \leqslant n^{-1/3+\varepsilon+2\omega_a/3+\omega_1-\omega_0}, \quad \text{with high probability.} \tag{C.49}$$

For B_{22}, we need to provide a bound on

$$\int_{\mathcal{I}_i(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E. \tag{C.50}$$

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 $i \geqslant n^\delta$ for some $\delta < \omega_\ell/10$. Then it is easy to check that $\widetilde{z}_i(t,\alpha)$ is away from the boundary of $\mathcal{I}_i(t,\alpha)$ at least by a distance of order n^{-2}, and we also have $|\widetilde{z}_i(t,\alpha)| \geqslant n^{-2}$. Then we can bound

$$\left|\int_{\mathcal{I}_i(t,0)} \frac{\rho_t(E_+(t,0)-E,0)}{E-\widetilde{z}_i(t,\alpha)}\mathrm{d}E\right| \leqslant \int_{\mathcal{I}_i(t,0),|E-\widetilde{z}_i(\alpha,t)|>n^{-50}} \frac{\sqrt{|E|}}{|\widetilde{z}_i(t,\alpha)-E|}\mathrm{d}E$$

$$+\left|\int_{|E-\widetilde{z}_i(t,\alpha)|\leqslant n^{-50}} \frac{\rho_t(E_+(t,0)-E,0)-\rho_t(E_+(t,0)-\widetilde{z}_i(t,\alpha),0)}{\widetilde{z}_i(t,\alpha)-E}\mathrm{d}E\right| =: D_1+D_2.$$

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

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

For the term D_2, by Lemma B.6 we have

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

which gives that

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

Next we consider the case with $i<n^{\delta}$. We only need to consider the case where $|\widetilde{z}_i(t,\alpha)|\leqslant n^{-100}$, because otherwise we can obtain an estimate in the same way as the case $i\geqslant n^{\delta}$. Then we decompose

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

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 $\widetilde{z}_i(t,\alpha)\leqslant 0$, then we have $F_3=F_4=0$. Otherwise, for F_4 we have

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

and for F_3 we have

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

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

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

which further implies that

$$|B_{12}| \lesssim n^{-1/3+\omega_a+\varepsilon}t = n^{-2/3+\varepsilon+\omega_a+\omega_1}. \tag{C.51}$$

In sum, combining (C.44), (C.45), (C.46), (C.47), (C.48), (C.49) and (C.51) and using the hierarchy of parameters (C.37), we obtain that for $1\leqslant i\leqslant n^{\omega_a}$, for any constant $\varepsilon>0$,

$$|\zeta_i(t)| \leqslant n^{-1/3-2\omega_\ell+\varepsilon} \quad \text{with high probability.} \tag{C.52}$$

Then combining (C.43) and (C.52), we obtain that for any constant $\varepsilon > 0$,

$$\|\zeta(t)\|_\infty \leqslant n^{-1/3-2\omega_\ell+\varepsilon} \quad \text{with high probability,} \tag{C.53}$$

uniformly in all $n^{-C_0} \leqslant t \leqslant t_1$ under the assumption (C.39). Plugging it into (C.40), we get

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

which concludes (C.38) under (C.39). Note that the right hand side of (C.38) is much smaller than $n^{-2/3}$ as in (C.39). Then using a simple continuity argument we can remove the assumption (C.39). Note that the continuity argument is actually 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.14) and local law, Theorem B.13, hold for all $n^{-C_0} \leqslant t \leqslant t_1$. Then we can perform the continuity argument on Ξ. □

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

Corollary C.11. *Let $i \leqslant n^{3\omega_\ell+\delta}$ for a constant $0 < \delta < \omega_\ell - \omega_1$. Then we have*

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

Proof. This is an immediate consequence of Lemma C.10 and (C.26). □

C.3 Proof of Theorem C.1

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 $u_i(t,\alpha) := \partial_\alpha \widehat{y}_i(t,\alpha)$. With (C.34)-(C.36), we find that $u = (u_i(t,\alpha) : 1 \leqslant i \leqslant p)$ satisfies the PDE

$$\partial_t u = \mathcal{L}u + \zeta^{(0)}. \tag{C.54}$$

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

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

and $\mathcal{V}$ is a diagonal operator, $(\mathcal{V}u)_i = \mathcal{V}_i u_i$, such that for $1 \leqslant i \leqslant n^{\omega_a}$,

$$\mathcal{V}_i := -2c_n\sqrt{E_+(t,0)} \int_{\mathcal{I}_i^c(t,0)} \frac{(\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))\rho_t(E_+(t,0)-E,0)}{[E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2]^2} \mathrm{d}E, \tag{C.55}$$

and for $n^{\omega_a} < i \leqslant i_*/2$

$$\mathcal{V}_i := -\frac{c_n}{2} \int_{\widehat{\mathcal{I}}_i^c(t,\alpha)\cap\mathcal{J}(t,\alpha)} \frac{f_t(\sqrt{E_+(t,\alpha)} - E,\alpha)}{[\widehat{y}_i(t,\alpha) - E]^2} \mathrm{d}E, \tag{C.56}$$

and $\mathcal{V}_i = 0$ when $i_*/2 < i \leqslant p$. As discussed below (C.39), we know that the semigroup of $\mathcal{L}$ is a contraction on every l^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 $1 \leqslant i \leqslant n^{\omega_a}$. For $i > n^{\omega_a}$, it is easy to check that for some constant $C > 0$,

$$|\zeta_i^{(0)}| \leqslant \mathbf{1}_{\{i\geqslant n^{\omega_a}\}} n^C \quad \text{with high probability.} \tag{C.57}$$

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

$$\partial_t v = \mathcal{L}v, \quad v_i(n^{-C_0}) = u_i(n^{-C_0})\mathbf{1}_{\{1\leqslant i\leqslant \ell^3 n^{\delta_v}\}}. \tag{C.58}$$

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

Proposition C.12. *With high probability, we have*

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

One can see that Proposition C.12 is an immediate consequence of the following finite speed of propagation estimate, whose proof is postponed to Section C.4.

Lemma C.13. *For any small constant $\delta > 0$, we have that for $a \leqslant \ell^3 n^{\delta}$ and $b \geqslant \ell^3 n^{2\delta}$,*

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

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

Remark C.14. Notice that we have $\mathcal{U}^{\mathcal{L}}_{ab}(s,t)\geqslant 0$ and $\mathcal{U}^{\mathcal{L}}_{ba}(s,t)\geqslant 0$ by the maximum principle. More precisely, let $v_i(t)=\exp(-\int_0^t \mathcal{V}_i(s)\mathrm{d}s)u_i(t)$, then $v=(v_i)$ satisfies the equation $\partial_t v=\mathcal{B}v$. Suppose $v_i(s)\geqslant 0$ for all i at time s, then we claim that $v_i(t)\geqslant 0$. In fact, suppose $v_j(t')=\min\{v_i(t'):1\leqslant i\leqslant p\}$ is the smallest entry of $v(t')$ at time $s\leqslant t'\leqslant t$, then we have $\partial_t v_j(t')=(\mathcal{B}v(t'))_j\geqslant 0$, i.e. the smallest entry of v will always increase. Hence the entries of v can never be negative.

Proof of Proposition C.12. Fix a $n^{-C_0}\leqslant t\leqslant 10t_1$, by Duhamel's principle we have

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

Then using Lemma C.13, $u_i(n^{-C_0},\alpha)-v_i(n^{-C_0},\alpha)=0$ for $i\leqslant \ell^3 n^{\delta_v}$, $\zeta_i^{(0)}(s)=0$ for $i\leqslant n^{\omega_a}$ and (C.57), we can conclude the proof. ☐

Another key ingredient is the following energy estimate. We postpone its proof until we complete the proof of Theorem C.1. Here we have fixed the starting time point to be n^{-C_0}, but the same conclusion holds for any starting time by the semigroup property.

Proposition C.15. *For any small constant $\delta_1 > 0$, consider $w\in\mathbb{R}^p$ with $w_i = 0$ for $i\geqslant \ell^3 n^{\delta_1}$. Then for any constants $\varepsilon,\eta > 0$ and fixed $q\geqslant 1$, there exists a constant $C(q,\eta) > 0$ independent of ε and η such that for all $2n^{-C_0}\leqslant t\leqslant 2t_1$,*

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

With all the above preparations, we are now ready to give the proof of Theorem C.1 .

Proof of Theorem C.1. Fix any $1 \leqslant i \leqslant \mathfrak{a}$, by (C.22) and (C.23) we have that with high probability,

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

for some constant $\tau > 0$. Recalling (C.30), we have that

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

where in the second step we used (B.24) and the rigidity estimate (C.26). Together with Lemma C.10, we obtain that

$$|(\lambda_i(t_1) - E_\lambda(t_1)) - (\mu_i(t_1) - E_\mu(t_1))| \lesssim |\widehat{y}_i(t_1,1) - \widehat{y}_i(t_1,0)| + n^{-2/3-\tau} \tag{C.59}$$

with high probability for some constant $\tau > 0$. Note that

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

Applying Proposition C.12 (together with a simple stochastic continuity argument to pass to all $0 \leqslant \alpha \leqslant 1$), we find that with high probability,

$$|\widehat{y}_i(t_1,0) - \widehat{y}_i(t_1,1)| \leqslant n^{-50} + \left|\int_0^1 v_i(t_1,\alpha)\mathrm{d}\alpha\right|. \tag{C.60}$$

By (C.7) and (C.9), we have that at $t = 0$,

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

for any small constant $\delta > 0$. Moreover, at $t = n^{-C_0}$ the eigenvalues are perturbed at most by $n^{-C_0/2}$, so we can calculate that

$$\|v(n^{-C_0},\alpha)\|_4 \prec n^{-2/3}, \quad 0 \leqslant \alpha \leqslant 1.$$

Finally, using Proposition C.15 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.60) and further into (C.59), we conclude the proof. □

The proof of Proposition C.15 is almost the same as the one for Lemma 3.11 in [67], so we only give an outline of the proof.

Proof of Proposition C.15. The proof relies on Lemma C.13 and the following Höder's continuity control.

Lemma C.16. *Fix a constant $0 < \delta_1 < \omega_\ell - \omega_1$. Let $w = (w_1, w_2, \cdots)$ be a vector such that $w_i = 0$ for $i \geqslant \ell^3 n^{\delta_1}$. For any constants $\eta, \varepsilon > 0$, there is a $C > 0$ independent of ε and η, and a constant $c_\eta > 0$ such that the following estimates hold with high probability for all $n^{-C_0} \leqslant s \leqslant t \leqslant 5t_1$,*

$$\|\mathcal{U}^{\mathcal{L}}(s,t)w\|_2 \leqslant \left(\frac{n^{C\eta+\varepsilon}}{c_\eta n^{1/3}(t-s)}\right)^{\frac{3}{2}(1-6\eta)} \|w\|_1, \tag{C.61}$$

and

$$\|(\mathcal{U}^{\mathcal{L}}(s,t))^\top w\|_2 \leqslant \left(\frac{n^{C\eta+\varepsilon}}{c_\eta n^{1/3}(t-s)}\right)^{\frac{3}{2}(1-6\eta)} \|w\|_1. \tag{C.62}$$

Proof. The proof is very similar to the ones for [67, Lemma 3.13], [15, Proposition 10.4] and [43, Section 10]. More precisely, our operator $\mathcal{L}$ is almost the same as the one in [67, Lemma 3.13], where the only difference is the operator $\mathcal{V}$. However, the $\mathcal{V}_i$ in (C.55) satisfies exactly the same estimate as the $\mathcal{V}_i$ in [67]. We omit the details of the proof. □

Now we complete the proof of Proposition C.15. Fix constants $0 < \delta_1 < \delta_2 < \omega_\ell - \omega_1$. We denote the indicator function $\mathcal{X}_2(i) = \mathbf{1}_{\{1 \leqslant i \leqslant \ell^3 n^{\delta_2}\}}$ and let $\mathcal{X}_2$ be the associated digonal operator. For any $v \in \mathbb{R}^p$ with $\|v\|_1 = 1$, we decompose that

$$\langle \mathcal{U}^{\mathcal{L}} w, v\rangle = \langle w, (\mathcal{U}^{\mathcal{L}})^\top v\rangle = \langle w, (\mathcal{U}^{\mathcal{L}})^\top \mathcal{X}_2 v\rangle + \langle w, (\mathcal{U}^{\mathcal{L}})^\top (1-\mathcal{X}_2) v\rangle.$$

where we abbreviated $\mathcal{U}^{\mathcal{L}} := \mathcal{U}^{\mathcal{L}}(n^{-C_0}, t)$. For the second term, by Lemma C.13 we obtain that

$$\left|\langle w, (\mathcal{U}^{\mathcal{L}})^\top (1-\mathcal{X}_2) v\rangle\right| \leqslant n^{-100}\|w\|_1 \leqslant n^{-99}\|w\|_2.$$

For the first term, by Lemma C.16 and Cauchy-Schwarz inequality, we obtain that for any constant $\eta > 0$,

$$\langle w, (\mathcal{U}^{\mathcal{L}})^\top \mathcal{X}_2 v\rangle \leqslant \|w\|_2 \|(\mathcal{U}^{\mathcal{L}})^\top \mathcal{X}_2 v\|_2 \leqslant \|w\|_2 \left(\frac{n^{C\eta+\varepsilon}}{c_\eta n^{1/3}(t-n^{-C_0})}\right)^{\frac{3}{2}(1-6\eta)} \|v\|_1.$$

By l^1–l^∞ duality and using $t \geqslant 2n^{-C_0}$, we find that

$$\|\mathcal{U}^{\mathcal{L}} w\|_\infty \leqslant C(\eta)\left(\frac{n^{C\eta+\varepsilon}}{n^{1/3}t}\right)^{\frac{3}{2}(1-6\eta)} \|w\|_2.$$

Consequently, by the semigroup property, we find that

$$\begin{aligned}\|\mathcal{U}^{\mathcal{L}}(n^{-C_0},t)w\|_\infty = \|\mathcal{U}^{\mathcal{L}}(2t/3,t)\mathcal{U}^{\mathcal{L}}(n^{-C_0},2t/3)w\|_\infty &\leqslant C(\eta)\left(\frac{n^{C\eta+\varepsilon}}{n^{1/3}t}\right)^{\frac{3}{2}(1-6\eta)} \|\mathcal{U}^{\mathcal{L}}(n^{-C_0},2t/3)w\|_2 \\ &\leqslant C(\eta)\left(\frac{n^{C\eta+\varepsilon}}{n^{1/3}t}\right)^{3(1-6\eta)} \|w\|_1,\end{aligned}$$

where we used Lemma C.16 again in the last step. Finally, the proof for the general l_q-norm case follows from the standard interpolation argument. □

C.4 Proof of Lemma C.13

Finally in this section, we prove the finite speed of propagation estimate, Lemma C.13. 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 C.17. *Fix a small constant $0 < \delta < \omega_\ell - \omega_1$. For any $a \geqslant \ell^3 n^\delta, b \leqslant \ell^3 n^\delta/2$ and fixed s such that $0 \leqslant s \leqslant 10t_1$, we have that with high probability,*

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

for any large constant $D > 0$.

We postpone its proof until we complete the proof of Lemma C.13. We will also need to use the following lemma in order to extend the result to all $0 \leqslant s \leqslant t \leqslant 10t_1$.

Lemma C.18. *Let υ_i be a solution of $\partial_t \upsilon = \mathcal{L}\upsilon$, with $\upsilon_i(0) \geqslant 0$. Then for $0 \leqslant t \leqslant 10t_1$, we have*

$$\frac{1}{2} \sum_i \upsilon_i(0) \leqslant \sum_i \upsilon_i(t) \leqslant \sum_i \upsilon_i(0).$$

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 first bound (C.55). Using the rigidity (C.26) and Lemma C.10, we have that with high probability,

$$E - E_+(t,0) + (\sqrt{E_+(t,0)} - \widehat{y}_i(t,\alpha))^2 \gtrsim E + n^{-2/3+2\omega_\ell}, \quad 1 \leqslant i \leqslant n^{\omega_a}, \quad E \in \mathcal{I}_i^c(t,0).$$

Together with $\rho_t(E_+(t,0) - E, 0) \sim \sqrt{E}$, we get that for $1 \leqslant i \leqslant n^{\omega_a}$,

$$0 \leqslant -\mathcal{V}_i \lesssim \int_{\mathcal{I}_i^c(t,0)} \frac{\sqrt{E}}{|E + n^{-2/3+2\omega_\ell}|^2} \mathrm{d}E \lesssim n^{1/3-\omega_\ell}.$$

We can get the same bound for (C.56). Then applying Gronwall's inequality to

$$-\left(C n^{1/3-\omega_\ell} \right) \sum_i u_i \leqslant \partial_t \sum_i u_i \leqslant 0,$$

we can conclude the proof. □

Now we can give the proof of Lemma C.13.

Proof of Lemma C.13. Fix any constant $0 < \varepsilon < \delta$. By the semigroup property, we have

$$\mathcal{U}^{\mathcal{L}}_{ai}(n^{-C_0}, t) = \sum_j \mathcal{U}^{\mathcal{L}}_{aj}(s,t) \mathcal{U}^{\mathcal{L}}_{ji}(n^{-C_0}, s) \geqslant \mathcal{U}^{\mathcal{L}}_{ab}(s,t) \mathcal{U}^{\mathcal{L}}_{bi}(n^{-C_0}, s). \tag{C.63}$$

By Lemma C.18, we find that $\sum_i \mathcal{U}^{\mathcal{L}}_{bi}(n^{-C_0}, s) \geqslant 1/2$. Moreover, by Lemma C.17 we have that $\mathcal{U}^{\mathcal{L}}_{bi}(n^{-C_0}, s) \leqslant n^{-100}$ for any $i \leqslant \ell^3 n^{\delta+\varepsilon}$. This implies that there exists an $i_* \geqslant \ell^3 n^{\delta+\varepsilon}$ such that $\mathcal{U}^{\mathcal{L}}_{bi_*}(n^{-C_0}, s) \geqslant (4n)^{-1}$. However, by Lemma C.17 we have that $\mathcal{U}^{\mathcal{L}}_{ai_*}(0,t) \leqslant n^{-D}$ for any large constant $D > 0$. Thus picking $i = i_*$ in (C.63), we get that $\mathcal{U}^{\mathcal{L}}_{ab}(s,t) \leqslant n^{-D+2}$. This finishes the proof for the estimate on $\mathcal{U}^{\mathcal{L}}_{ab}(s,t)$. The estimate on $\mathcal{U}^{\mathcal{L}}_{ba}(s,t)$ can be proved in a similar way. □

It remains to prove Lemma C.17. The strategy was first developed in [17], and later used in [65, 67] to study the symmetric DBM for Wigner type matrices. Our proof is similar to the ones for [65, Lemma 4.2] and [67, Lemma 4.1], so we will not write down all the details.

Proof of Lemma C.17. We focus on the case $s = 0$ and the general case can be dealt with similarly with a simple time shift. Let ψ be a smooth function such that (i) $\psi(x) = -x$ for $|x| \leqslant \ell^2 n^{-2/3+2\delta/3}$, (ii) $\psi'(x) = 0$ for $|x| > 2\ell^2 n^{-2/3+2\delta/3}$, (iii) ψ is decreasing, (iv) $|\psi(x) - \psi(y)| \leqslant |x-y|$ and $|\psi'(x)| \leqslant 1$, and (v) $|\psi''(x)| \leqslant C\ell^{-2} n^{2/3-2\delta/3}$ for some constant $C > 0$. Similar to [67, Lemma 4.1], we now consider a solution of

$$\partial_t f = \mathcal{L} f, \quad \text{with} \;\; f_i(0) = \delta_{q_*},$$

for any $q_* \geqslant q := \ell^3 n^\delta$. Let $\nu > 0$ be a fixed constant and define functions

$$\phi_k := \exp(\nu\psi(\widehat{y}_k(t,\alpha) - \widehat{\gamma}_q(t,\alpha))), \quad v_k := \phi_k f_k, \quad F(t) := \sum_k v_k^2.$$

We will choose a specific ν along the proof such that on one hand, ϕ_k is large enough, and on the other hand, $F(t)$ is bounded. By Ito's formula, we find that F satisfies the SDE

$$\mathrm{d}F = -\sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij}(v_i - v_j)^2 \mathrm{d}t + 2\sum_i \mathcal{V}_i v_i^2 \mathrm{d}t \tag{C.64}$$

$$+ \sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij} v_i v_j \left(\frac{\phi_i}{\phi_j} + \frac{\phi_j}{\phi_i} - 2\right) \mathrm{d}t \tag{C.65}$$

$$+ 2\nu \sum_i v_i^2 \psi'(\widehat{y}_i - \widehat{\gamma}_q) \mathrm{d}(\widehat{y}_i - \widehat{\gamma}_q) \tag{C.66}$$

$$+ \sum_i v_i^2 \left(\frac{2\nu^2}{n}[\psi'(\widehat{y}_i - \widehat{\gamma}_q)]^2 + \frac{\nu}{n}\psi''(\widehat{y}_i - \widehat{\gamma}_q)\right) \mathrm{d}t, \tag{C.67}$$

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 suitable stopping time. Let τ_1 be the stopping time such that for $t < \tau_1$, Lemmas C.4 and C.10 hold true for a sufficiently small constant $0 < \varepsilon < \delta/1000$. Note that with high probability, $\tau_1 \geqslant 10 t_1$. Let τ_2 be the first time such that $F \geqslant 10$. Then we let the stopping time τ be

$$\tau := \min\{\tau_1, \tau_2, 10t_1\},$$

and 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.64)-(C.67). First, (C.64) is a dissipative term, so it can only decrease the size of $F(t)$. By Corollary C.11, 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 $i \leqslant C\ell^3 n^\delta$ and $(i,j) \in \mathcal{A}$, then $j \leqslant C'\ell^3 n^\delta$ for some constant $C' > 0$. Thus the nonzero terms in (C.65) must satisfy that $i, j \leqslant C\ell^3 n^\delta$ for a large enough constant $C > 0$. Then by Corollary C.11, for $i, j \leqslant C\ell^3 n^\delta$ statisfying $(i,j) \in \mathcal{A}$, we have

$$|\widehat{y}_i - \widehat{y}_j| \lesssim \frac{\ell^2 n^{\delta/3}}{n^{2/3}}.$$

Now with the Taylor expansion of $e^{-x}+e^{x}-2$, we get that if

$$\frac{\nu\ell^2 n^{\delta/3}}{n^{2/3}} \leqslant C_1 \tag{C.68}$$

for some constant $C_1>0$, then

$$(C.65) \leqslant C\frac{\nu^2}{n}\sum_{(j,i)\in\mathcal{A}}(v_i^2+v_j^2)\mathbf{1}_{\{\phi_j\neq\phi_i\}}\mathrm{d}t \leqslant \frac{\nu^2\ell^3 n^{2\delta/3}}{n}F(t)\mathrm{d}t. \tag{C.69}$$

The term (C.67) can be easily bounded as

$$(C.67) \leqslant C\left(\frac{\nu^2}{n}+\frac{\nu\ell^{-2}}{n^{1/3+2\delta/3}}\right)F(t)\mathrm{d}t. \tag{C.70}$$

It remains to control (C.66). Since $\psi'(\widehat{y}_i-\zeta_q)\neq 0$ only when $i\leqslant C\ell^3 n^\delta \ll n^{\omega_a}$, thus $\hat{y}_i$ satisfies the SDE (C.34), which gives

$$\begin{aligned}\mathrm{d}(\widehat{y}_i(t,\alpha)-\hat{\gamma}_q(t,\alpha)) = &-\frac{\mathrm{d}B_i}{\sqrt{n}}-\frac{n-p}{2nE_+(t,0)}\mathrm{d}t+\frac{1}{2n}\sum_j^{\mathcal{A},(i)}\frac{1}{\widehat{y}_i(t,\alpha)-\widehat{y}_j(t,\alpha)}\mathrm{d}t+\left(\frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t}-\frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t}\right)\mathrm{d}t\\ &-\left[c_n\int_{\mathcal{I}_i^c(t,0)}\frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2}\mathrm{d}E\right]\mathrm{d}t.\end{aligned} \tag{C.71}$$

By Burkholder-Davis-Gundy inequality and Markov's inequality, we find that with high probability,

$$\sup_{0\leqslant t\leqslant\tau}\nu\left|\int_0^t\sum_i v_i^2\psi'(\widehat{y}_i-\widehat{\gamma}_q)\frac{\mathrm{d}B_i}{\sqrt{n}}\right|\leqslant n^{\varepsilon}\nu\left(\frac{n^{\omega_1}}{n^{4/3}}\right)^{1/2} \tag{C.72}$$

for any constant $\varepsilon>0$. Moreover, with the same arguments for (4.17) of [67] we can obtain that

$$\frac{\nu}{n}\sum_{(i,j)\in\mathcal{A}}\frac{v_i^2\psi'(\widehat{y}_i-\widehat{\gamma}_q)}{\widehat{y}_i-\widehat{y}_j}\leqslant\frac{1}{100}\sum_{(i,j)\in\mathcal{A}}\mathcal{B}_{ij}(v_i-v_j)^2+C\left(\frac{\nu n^{\omega_\ell}}{n^{1/3}}+\frac{\nu^2 n^{3\omega_\ell+2\delta/3}}{n}\right)F(t) \tag{C.73}$$

for large enough constant $C>0$. The main difference from the argument in [67] is about the term

$$\begin{aligned}&-\frac{n-p}{2n\sqrt{E_+(t,0)}}-c_n\int_{\mathcal{I}_i^c(t,0)}\frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2}\mathrm{d}E+\frac{\mathrm{d}\sqrt{E_+(t,0)}}{\mathrm{d}t}-\frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t}\\ &=c_n\sqrt{E_+(t,0)}\left[m_t\left((\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2,0\right)-m_t(E_+(t,0),0)\right]\\ &+c_n\int_{\mathcal{I}_i(t,0)}\frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2}\mathrm{d}E-\frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t}+\mathrm{O}(t),\end{aligned} \tag{C.74}$$

where we used (C.15) and $\zeta_t(E_+(t,0),0)=E_+(t,0)+\mathrm{O}(t)$ in the derivation. Recalling that $m_t(\cdot,0)$ is the Stieltjes transform of the MP law (C.2), so by the square root behavior around the right edge we have that

$$|m_t((\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2,0)-m_t(E_+(t,0),0)|\lesssim\sqrt{|\widehat{y}_i(t,\alpha)|}\lesssim n^{-1/3+\omega_\ell+\delta/3}$$

with high probability, where we used (C.26) in the last step. For the second term on the right-hand side of (C.74), it can be bounded in the same way as (C.50) and we can get

$$\left|\int_{\mathcal{I}_i(t,0)} \frac{\sqrt{E_+(t,0)}\rho_t(E_+(t,0)-E,0)}{E-E_+(t,0)+(\sqrt{E_+(t,0)}-\widehat{y}_i(t,\alpha))^2}\mathrm{d}E\right| \lesssim n^{-1/3+\omega_\ell+\delta/3},$$

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 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 (almost) 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.13). Thus we obtain from the above equation that

$$\left|\frac{\mathrm{d}\widehat{\gamma}_q(t,\alpha)}{\mathrm{d}t}\right| \lesssim n^{-1/3+\omega_\ell+\delta/3}.$$

Combining the above estimate, we get

$$|(C.74)| = \mathrm{O}(n^{-1/3+\omega_\ell+\delta/3}). \tag{C.75}$$

Together with (C.69), (C.70),(C.72) and (C.73), we find that if ν satisfies the condition of (C.68), then with high probability,

$$\partial_t F(t) \leqslant C\left(\frac{\nu^2 n^{3\omega_l+2\delta/3}}{n} + \frac{\nu n^{\omega_\ell+\delta/3}}{n^{1/3}}\right)F(t).$$

Then by Gronwall's inequality, we get

$$\sup_{0\leqslant s\leqslant\tau} F(s) \leqslant F(0) + C\left(\frac{\nu^2 n^{3\omega_l+2\delta/3+\omega_1}}{n^{4/3}} + \frac{\nu n^{\omega_\ell+\omega_1+\delta/3}}{n^{2/3}}\right)$$

with high probability. Hence choosing $\nu = n^{2/3-2\omega_l-\delta/3}$, we get by continuity that $\tau = 10t_1$ holds with high probability, i.e.

$$\sup_{0\leqslant s\leqslant 10t_1} F(s) \leqslant 10, \quad \text{with high probability.}$$

Now notice that if $i \leqslant \ell^3 n^\delta/2$, 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 $\mathcal{U}^{\mathcal{L}}_{iq_*}(0,t) \leqslant n^{-D}$ for any large constant D for any $i \leqslant \ell^3 n^\delta/2$ and $q_* \geqslant \ell^3 n^\delta$. The proof for $\mathcal{U}^{\mathcal{L}}_{q_* i}$ is the same by setting $\psi \to -\psi$. □

References

[1] E. Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[2] A. Adhikari and Z. Che. The edge universality of correlated matrices. *arXiv preprint arXiv 1712.04889*, 2017.

[3] O. Ajanki, L. Erdős, and T. Krüger. Quadratic vector equations on complex upper half-plane. *arXiv:1506.05095*, 2015.

[4] J. Alt. Singularities of the density of states of random Gram matrices. *Electron. Commun. Probab.*, 22:13 pp., 2017.

[5] J. Alt, L. Erdős, and T. Krüger. Local law for random Gram matrices. *Electron. J. Probab.*, 22:41 pp., 2017.

[6] 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 Transactions on Information Theory*, 63(6):3451–3467, 2017.

[7] Z. Bao, X. Ding, and K. Wang. Singular vector and singular subspace distribution for the matrix denoising model. *Annals of Statistics (In press)*, 2020.

[8] Z. Bao, G. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.*, 43(1):382–421, 2015.

[9] Z. G. Bao, G. M. Pan, and W. Zhou. Local density of the spectrum on the edge for sample covariance matrices with general population. *Preprint*, 2013.

[10] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau. Edge rigidity and universality of random regular graphs of intermediate degree. *Geometric and Functional Analysis*, 2020.

[11] 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*, pages 87–172. Birkhäuser, 2007.

[12] P. Bianchi, M. Debbah, M. Maida, and J. Najim. Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inf. Theor.*, 57(4):2400–2419, 2011.

[13] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19(33):1–53, 2014.

[14] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Prob. Theor. Rel. Fields*, 164(1):459–552, 2016.

[15] P. Bourgade, L. Erdös, and H. B. Yau. Edge universality of Beta ensembles. *Communications in Mathematical Physics*, 332:261–353, 2014.

[16] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin. Fixed energy universality for generalized Wigner matrices. *Communications on Pure and Applied Mathematics*, 69(10):1815–1881, 2016.

[17] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicity. *Communications in Mathematical Physics*, 350(1):231–278, 2017.

[18] P. Bourgade, H.-T. Yau, and J. Yin. Local circular law for random matrices. *Probab. Theory Relat. Fields*, 159:545–595, 2014.

[19] G. Cipolloni, L. Erdős, T. Krüger, and D. Schröder. Cusp universality for random matrices, II: The real symmetric case. *Pure Appl. Anal.*, 1(4):615–707, 2019.

[20] R. Couillet and M. Debbah. *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[21] X. Ding. High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli*, 26(1):387–417, 2020.

[22] X. Ding and F. Yang. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.*, 28(3):1679–1738, 2018.

[23] X. Ding and F. Yang. Edge statististics of large dimensional deformed rectangular matrices. *In preparation*, 2020.

[24] X. Ding and F. Yang. Spiked separable covariance matrices and principal components. *Annals of Statistics (to appear)*, 2020.

[25] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

[26] R. B. Dozier and J. W. Silverstein. Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices. *Journal of Multivariate Analysis*, 98(6):1099 – 1122, 2007.

[27] R. B. Dozier and J. W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *Journal of Multivariate Analysis*, 98(4):678 – 694, 2007.

[28] M. El Amine Seddik, C. Louart, M. Tamaazousti, and R. Couillet. Random Matrix Theory Proves that Deep Learning Representations of GAN-data Behave as Gaussian Mixtures. *arXiv preprint arXiv 2001.08370*, 2020.

[29] N. El Karoui. Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007.

[30] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, pages 2757–2790, 2008.

[31] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14:1837–1926, 2013.

[32] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.*, 314:587–640, 2012.

[33] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Delocalization and diffusion profile for random band matrices. *Commun. Math. Phys.*, 323:367–416, 2013.

[34] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab.*, 18:1–58, 2013.

[35] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

[36] L. Erdős, T. Krüger, and D. Schröder. Cusp universality for random matrices I: Local law and the complex hermitian case. *Communications in Mathematical Physics*, 2020.

[37] L. Erdős, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau. Bulk universality for Wigner matrices. *Communications on Pure and Applied Mathematics*, 63(7):895–925, 2010.

[38] L. Erdős, B. Schlein, and H.-T. Yau. Universality of random matrices and local relaxation flow. *Inventiones mathematicae*, 185(1):75–119, 2011.

[39] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 48(1):1–46, 2012.

[40] L. Erdős and H.-T. Yau. A dynamical approach to random matrix theory. *Courant Lecture Notes in Mathematics*, 28, 2017.

[41] L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Relat. Fields*, 154(1):341–407, 2012.

[42] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229:1435 – 1515, 2012.

[43] L. Erdős and H.-T. Yau. Gap universality of generalized Wigner and β-ensembles. *Journal of the European Mathematical Society*, 017(8):1927–2036, 2015.

[44] Z. Fan and I. M. Johnstone. Tracy-Widom at each edge of real covariance and MANOVA estimators. *arXiv preprint arXiv 1707.02352*, 2017.

[45] V. Feldman, W. Perkins, and S. Vempala. Subsampled power iteration: A unified algorithm for block models and planted csp's. In *Advances in Neural Information Processing Systems*, NIPS'15, page 2836–2844, 2015.

[46] L. Florescu and W. Perkins. Spectral thresholds in the bipartite stochastic block model. In *29th Annual Conference on Learning Theory*, pages 943–959, 2016.

[47] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. A survey of statistical network models. *Found. Trends Mach. Learn.*, 2(2):129–233, 2010.

[48] D. Granziol, T. Garipov, D. Vetrov, S. Zohren, S. Roberts, and A. G. Wilson. Towards understanding the true loss surface of deep neural networks using random matrix theory and iterative spectral methods, 2020.

[49] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.*, 17(3):875–930, 2007.

[50] X. Han, G. Pan, and B. Zhang. The Tracy–Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.*, 44(4):1564–1592, 2016.

[51] Y. He and A. Knowles. Fluctuations of extreme eigenvalues of sparse Erdős-Rényi graphs. *arXiv:2005.02254*, 2020.

[52] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[53] 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.*, 48(2):916–962, 2020.

[54] J. Y. Hwang, J. O. Lee, and K. Schnelli. Local law and Tracy–Widom limit for sparse sample covariance matrices. *Ann. Appl. Probab.*, 29(5):3006–3036, 2019.

[55] D. Jiang, C. Tang, and A. Zhang. Cluster analysis for gene expression data: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 16(11):1370–1386, 2004.

[56] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29:295–327, 2001.

[57] I. M. Johnstone and A. Onatski. Testing in high-dimensional spiked models. *Ann. Statist.*, 48(3):1231–1254, 2020.

[58] S. Kay. *Fundamentals of Statistical Signal Processing: Detection theory.* Prentice-Hall PTR, 1998.

[59] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, pages 1–96, 2016.

[60] M. Kolar and E. P. Xing. Estimating sparse precision matrices from data with missing values. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, page 635–642, 2012.

[61] W. Kong and G. Valiant. Spectrum estimation from samples. *Ann. Statist.*, 45(5):2218–2247, 2017.

[62] S. Kritchman and B. Nadler. Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, 57(10):3930–3941, 2009.

[63] D. Kundu. Estimating the number of signals in the presence of white noise. *Journal of Statistical Planning and Inference*, 90(1):57 – 68, 2000.

[64] P. C. Kyriakidis and A. G. Journel. Geostatistical space–time models: A review. *Mathematical Geology*, 31(6):651–684, 1999.

[65] B. Landon, P. Sosoe, and H.-T. Yau. Fixed energy universality of Dyson brownian motion. *Advances in Mathematics*, 346:1137 – 1332, 2019.

[66] B. Landon and H.-T. Yau. Convergence of local statistics of Dyson Brownian Motion. *Communications in Mathematical Physics*, 355(3):949–1000, 2017.

[67] B. Landon and H.-T. Yau. Edge statistics of Dyson Brownian motion. *arXiv preprint arXiv:1712.03881*, 2017.

[68] O. Ledoit and M. Wolf. Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139:360 – 384, 2015.

[69] O. Ledoit and M. Wolf. Numerical implementation of the QuEST function. *Computational Statistics & Data Analysis*, 115:199 – 223, 2017.

[70] J. O. Lee and K. Schnelli. Edge universality for deformed Wigner matrices. *Reviews in Mathematical Physics*, 27(08):1550018, 2015.

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

[72] J. O. Lee and K. Schnelli. Local law and Tracy–Widom limit for sparse random matrices. *Probability Theory and Related Fields*, 171(1):543–616, 2018.

[73] J. O. Lee and J. Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163:117–173, 2014.

[74] 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.*, 16:1934–1959, 2011.

[75] K. Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3):1029–1058, 2014.

[76] 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*, 285(5428):751–753, 1999.

[77] X. Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11):5113–5129, 2008.

[78] R. R. Nadakuditi. Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Transactions on Information Theory*, 60(5):3002–3018, 2014.

[79] R. R. Nadakuditi and A. Edelman. Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Transactions on Signal Processing*, 56(7):2625–2638, 2008.

[80] 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 Journal of Selected Topics in Signal Processing*, 4(3):468–480, 2010.

[81] M. Ndaoud, S. Sigalla, and A. B. Tsybakov. Improved clustering algorithms for the Bipartite Stochastic Block Model. *arXiv preprint arXiv 1911.07987*, 2019.

[82] S. Neumann. Bipartite stochastic block models with tiny clusters. In *Advances in Neural Information Processing Systems 31*, pages 3867–3877. 2018.

[83] A. Onatski. The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.*, 18(2):470–490, 2008.

[84] A. Onatski. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479, 2009.

[85] A. Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244 – 258, 2012.

[86] A. Onatski, M. J. Moreira, and M. Hallin. Signal detection in high dimension: The multispiked case. *Ann. Statist.*, 42(1):225–254, 2014.

[87] D. Paul and A. Aue. Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150:1 – 29, 2014.

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

[89] C. A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159:151–174, 1994.

[90] C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177:727–754, 1996.

[91] D. W. Tufts and A. A. Shah. Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix. *IEEE Transactions on Signal Processing*, 41(4):1716–1721, 1993.

[92] P. Vallet, P. Loubaton, and X. Mestre. Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case. *IEEE Transactions on Information Theory*, 58(2):1043–1068, 2012.

[93] J. Vinogradova, R. Couillet, and W. Hachem. Statistical inference in large antenna arrays under unknown noise pattern. *IEEE Transactions on Signal Processing*, 61(22):5633–5645, 2013.

[94] K. Werner, M. Jansson, and P. Stoica. On estimation of covariance matrices with Kronecker product structure. *IEEE Transactions on Signal Processing*, 56(2):478–491, 2008.

[95] H. Xi, F. Yang, and J. Yin. Local circular law for the product of a deterministic matrix with a random matrix. *Electron. J. Probab.*, 22:77 pp., 2017.

[96] H. Xi, F. Yang, and J. Yin. Convergence of eigenvector empirical spectral distribution of sample covariance matrices. *Ann. Statist.*, 48(2):953–982, 2020.

[97] D. Yang, Z. Ma, and A. Buja. Rate optimal denoising of simultaneously sparse and low rank matrices. *J. Mach. Learn. Res.*, 17(1):3163–3189, 2016.

[98] F. Yang. Edge universality of separable covariance matrices. *Electron. J. Probab.*, 24:57 pp., 2019.

[99] J. Yang, P. Chen, and T.-J. Wu. On estimating the number of signals. In *2007 IEEE Antennas and Propagation Society International Symposium*, pages 1132–1135, 2007.