Phase transition for the smallest eigenvalue of covariance matrices

Zhigang Bao[*], Jaehun Lee[†], Xiaocong Xu[‡]

[*]*mazgbao@ust.hk*, [†]*jaehun.lee@ust.hk*, [‡]*xxuay@connect.ust.hk*

[*]*Department of Mathematics, HKUST*

Abstract

In this paper, we study the smallest non-zero eigenvalue of the sample covariance matrices $\mathcal{S}(Y) = YY^*$, where $Y = (y_{ij})$ is an $M \times N$ matrix with iid mean 0 variance N^{-1} entries. We consider the regime $M = M(N)$ and $M/N \to c_\infty \in \mathbb{R} \setminus \{1\}$ as $N \to \infty$. It is known that for the extreme eigenvalues of Wigner matrices and the largest eigenvalue of $\mathcal{S}(Y)$, a weak 4th moment condition is necessary and sufficient for the Tracy-Widom law [21, 49]. In this paper, we show that the Tracy-Widom law is more robust for the smallest eigenvalue of $\mathcal{S}(Y)$, by discovering a phase transition induced by the fatness of the tail of y_{ij}'s. More specifically, we assume that y_{ij} is symmetrically distributed with tail probability $\mathbb{P}(|\sqrt{N}y_{ij}| \geq x) \sim x^{-\alpha}$ when $x \to \infty$, for some $\alpha \in (2, 4)$. We show the following conclusions: (i). When $\alpha > \frac{8}{3}$, the smallest eigenvalue follows the Tracy-Widom law on scale $N^{-\frac{2}{3}}$; (ii). When $2 < \alpha < \frac{8}{3}$, the smallest eigenvalue follows the Gaussian law on scale $N^{-\frac{\alpha}{4}}$; (iii). When $\alpha = \frac{8}{3}$, the distribution is given by an interpolation between Tracy-Widom and Gaussian; (iv). In case $\alpha \leq \frac{10}{3}$, in addition to the left edge of the MP law, a deterministic shift of order $N^{1-\frac{\alpha}{2}}$ shall be subtracted from the smallest eigenvalue, in both the Tracy-Widom law and the Gaussian law. Overall speaking, our proof strategy is inspired by [4] which is originally done for the bulk regime of the Lévy Wigner matrices. In addition to various technical complications arising from the bulk-to-edge extension, two ingredients are needed for our derivation: an intermediate left edge local law based on a simple but effective matrix minor argument, and a mesoscopic CLT for the linear spectral statistic with asymptotic expansion for its expectation.

Keywords: sample covariance matrix, smallest eigenvalue, Tracy-Widom law, heavy-tailed random matrix

1 Introduction

1.1 Main results

As one of the most classic models in random matrix theory, the sample covariance matrices have been widely studied. When considering the high-dimensional setting it is well-known that the empirical spectral distribution converges to Marchenko-Pastur law (MP law). Inspired by problems such as PCA, the extreme eigenvalue has also been extensively studied. Among the most well-known results in this direction are probably the Bai-Yin law [9] on the first order limit and the Tracy-Widom law [38, 39] on the second order fluctuation of the extreme eigenvalues. More specifically, let $Y = (y_{ij}) \in \mathbb{R}^{M \times N}$ be a random matrix with i.i.d. mean 0 and variance N^{-1} entries, and assume that $\sqrt{N}y_{ij}$'s are i.i.d. copies of an random variable Θ which is independent of N. The covariance matrix with the data matrix Y is defined as $\mathcal{S}(Y) = YY^*$. Let $\lambda_1(\mathcal{S}(Y)) \geq \ldots \geq \lambda_M(\mathcal{S}(Y))$ be the ordered eigenvalues of $\mathcal{S}(Y)$. We denote by $\mu_N = \frac{1}{M} \sum_{i=1}^{M} \delta_{\lambda_i}$ the empirical spectral distribution. In the regime $M = M(N)$, $c_N := M/N \to c_\infty \in$

$(0, \infty)$ as $N \to \infty$, it is well known since [52] that μ_N is weakly approximated by the MP law

$$\rho^{\mathsf{mp}}(\mathrm{d}x) = \frac{1}{2\pi c_N x} \sqrt{[(\lambda_+^{\mathsf{mp}} - x)(x - \lambda_-^{\mathsf{mp}})]_+} \mathrm{d}x + (1 - \frac{1}{c_N})_+ \delta_0(x), \qquad \lambda_{\pm}^{\mathsf{mp}} = (1 \pm \sqrt{c_N})^2. \qquad (1)$$

The Stieltjes transform of ρ^{mp} is denoted as $\mathsf{m}_{\mathsf{mp}}(z)$, which satisfies the following equation:

$$z c_N \mathsf{m}_{\mathsf{mp}}^2(z) + (z - (1 - c_N)) \mathsf{m}_{\mathsf{mp}}(z) + 1 = 0. \qquad (2)$$

Equivalently,

$$\mathsf{m}_{\mathsf{mp}}(z) = \frac{1 - c_N - z + \mathrm{i}\sqrt{(\lambda_+^{\mathsf{mp}} - z)(z - \lambda_-^{\mathsf{mp}})}}{2z c_N}, \qquad (3)$$

where the square root is taken with a branch cut on the negative real axis.

Throughout the paper, we will be interested in the regime $c_\infty \neq 1$. In this case, both $\lambda_{\pm}^{\mathsf{mp}}$ are called soft edges of the spectrum. Regarding the extreme eigenvalues, Bai-Yin law [9] states that

$$\lambda_1(\mathcal{S}(Y)) - \lambda_+^{\mathsf{mp}} \xrightarrow{a.s.} 0, \qquad \lambda_{M \wedge N}(\mathcal{S}(Y)) - \lambda_-^{\mathsf{mp}} \xrightarrow{a.s.} 0,$$

as long as $\mathbb{E}|\sqrt{N} y_{ij}|^4 < \infty$ is additionally assumed. It is also shown in [9] that $\mathbb{E}|\sqrt{N} y_{ij}|^4 < \infty$ is necessary and sufficient for the convergence of $\lambda_1(\mathcal{S}(Y))$ to λ_+^{mp}. It had been widely believed that the convergence of the smallest eigenvalue $\lambda_{M \wedge N}(\mathcal{S}(Y))$ to λ_-^{mp} requires a weaker moment condition, and indeed it was shown in [60] that the condition of mean 0 and variance 1 for $\sqrt{N} y_{ij}$'s is already sufficient. On the level of the second order fluctuation, as an extension of the seminal work on Wigner matrix [49], it was shown in [21] that the sufficient and necessary condition for the Tracy-Widom law of $\lambda_1(\mathcal{S}(Y))$ is the existence of a weak 4-th moment

$$\lim_{s \to \infty} s^4 \mathbb{P}(|\sqrt{N} y_{11}| \geq s) = 0. \qquad (4)$$

Similarly to the first order result in [60], it has been believed that the Tracy-Widom law shall hold for the smallest eigenvalue $\lambda_{M \wedge N}(\mathcal{S}(Y))$ under a weaker condition. In this work, we are going to show that the smallest eigenvalue counterpart of (4) is

$$\lim_{s \to \infty} s^{\frac{8}{3}} \mathbb{P}(|\sqrt{N} y_{11}| \geq s) = 0,$$

under Assumption 1 below. Moreover, when the tail $\mathbb{P}(|\sqrt{N} y_{11}| \geq s)$ becomes heavier, the distribution of $\lambda_{M \wedge N}(\mathcal{S}(Y))$ exhibits a phase transition from Tracy-Widom to Gaussian. For technical reason, we make the following assumptions on $\mathcal{S}(Y)$.

Assumption 1. *We make the following assumptions on the covariance matrix $\mathcal{S}(Y)$.*

(i). (On matrix entries) We suppose that $\sqrt{N} y_{ij}$'s are all iid copies of a random variable Θ which is independent of N. Suppose that $\mathbb{E}\Theta = 0$ and $\mathbb{E}\Theta^2 = 1$. We further assume that Θ is symmetrically distributed, absolutely continuous with a positive density at 0 and as $s \to \infty$,

$$\left| \mathbb{P}(\Theta > s) + \frac{\mathsf{c}}{\Gamma(1 - \alpha/2)} s^{-\alpha} \right| \lesssim s^{-(\alpha + \varrho)}$$

for some $\alpha \in (2, 4)$, some constant $\mathsf{c} > 0$ and some small $\varrho > 0$,

(ii). (On dimension) We assume that $M := M(N)$ and as $N \to \infty$

$$c_N := \frac{M}{N} \to c_\infty \in (0, \infty) \setminus \{1\}.$$

Our results are collected in the following main theorem. For brevity, we assume $M < N$ throughout this paper. Analogous results can be easily obtained by switching the role of M and N when $M > N$.

Theorem 1.1. *Suppose that Assumption 1 holds. There exists a random variable $\mathcal{X}_\alpha$, such that the following statements hold when $N \to \infty$.*

(i):

$$\frac{-M^{\frac{2}{3}}}{\sqrt{c_N}(1-\sqrt{c_N})^{4/3}}\big(\lambda_{M\wedge N}(\mathcal{S}(Y)) - \lambda_-^{\mathsf{mp}} - \mathcal{X}_\alpha\big) \Rightarrow \mathrm{TW}_1.$$

(ii):

$$\frac{N^{\frac{\alpha}{4}}\big(\mathcal{X}_\alpha - \mathbb{E}\mathcal{X}_\alpha\big)}{\sigma_\alpha} \Rightarrow N(0,1), \qquad \sigma_\alpha^2 = \frac{\mathsf{c}c_N^{(4-\alpha)/4}(1-\sqrt{c_N})^4(\alpha-2)}{2}\Gamma\Big(\frac{\alpha}{2}+1\Big).$$

(iii):

$$\mathbb{E}\mathcal{X}_\alpha = -N^{1-\frac{\alpha}{2}}\frac{\mathsf{c}(1-\sqrt{c_N})^2}{c_N^{(\alpha-2)/4}}\Gamma\Big(\frac{\alpha}{2}+1\Big) + \mathfrak{o}(N^{1-\frac{\alpha}{2}}),$$

(iv): In case $\alpha = 8/3$, the following convergence holds:

$$\frac{-M^{\frac{2}{3}}}{\sqrt{c_N}(1-\sqrt{c_N})^{\frac{4}{3}}}\big(\lambda_{M\wedge N}(\mathcal{S}(Y)) - \lambda_-^{\mathsf{mp}} - \mathbb{E}\mathcal{X}_\alpha\big) \Rightarrow \mathrm{TW}_1 + \mathcal{N}(0,\tilde{\sigma}^2), \quad \tilde{\sigma}^2 = \frac{\mathsf{c}c_\infty^{\frac{2}{3}}(1-\sqrt{c_\infty})^{\frac{4}{3}}}{3}\Gamma\left(\frac{7}{3}\right).$$

where TW_1 and $\mathcal{N}(0,\tilde{\sigma})$ in the RHS of the above convergence are independent.

Remark 1. *From the above theorem, we can see that a phase transition occurs at $\alpha = 8/3$. When $\alpha > 8/3$, the fluctuation of $\lambda_{M\wedge N}(\mathcal{S}(Y))$ is governed by TW_1 on scale $N^{-2/3}$. When $2 < \alpha < 8/3$, the fluctuation is dominated by that of $\mathcal{X}_\alpha$, and thus it is Gaussian on scale $N^{-\alpha/4}$. In the case $\alpha = 8/3$, the limiting distribution is given by the convolution of a Tracy-Widom and Gaussian. When $\alpha \leq 10/3$, a shift of order $N^{1-\alpha/2}$ is created by $\mathbb{E}\mathcal{X}_\alpha$. We remark here that a natural further direction is to exploit the expansion of $\mathbb{E}\mathcal{X}_\alpha$ up to an order smaller than the fluctuation. But due to technical reason, we do not pursue this direction in the current paper.*

1.2 Related References

The Tracy-Widom distribution in random matrices was first obtained for GOE and GUE in [62, 63] and was later extended to Wishart matrices in [38] and [39]. In the past few decades, the universality of the Tracy-Widom law has been extensively studied. The extreme eigenvalues of many random matrices with general distributions and structures have been proven to follow the Tracy-Widom distribution. We refer to the following literature [6, 10, 23, 28, 30, 41, 43, 46, 47, 49, 53–57, 59] for related developments. Although the Tracy-Widom distribution is very robust, some phase transitions may occur when considering heavy-tailed matrices or sparse matrices. For example, for sparse Erdős-Rényi graphs $G(N, p)$, it is known from [36] that a phase transition from Tracy-Widom to Gaussian will occur when p crosses $N^{-2/3}$. We also refer to [27, 31, 35, 45, 48] for related study. For heavy-tailed Wigner matrices or sample covariance matrices, as we mentioned, according to [49] and [21], the largest eigenvalue follows the Tracy Widom distribution if and only if a weak 4-th moment condition is satisfied. From [7, 20, 58], we also know the distribution of the largest eigenvalue when the matrix entries have heavier tail. We would also like to mention the recent research on the mobility edge of Lévy matrix with $\alpha < 1$ in [5]. On the other hand, if we focus on bulk statistics, universality will be very robust. For any $\alpha > 0$, it is proved in [2, 4] that the bulk universality is valid. An extension of [4] to the hard edge of the covariance matrix in case $M = N$ is considered in [50]. In our current work, we focus on the regime $\alpha \in (2, 4)$ for the left edge of the covariance matrices. According to [12], even the global law will no longer be MP law in case $\alpha < 2$, and thus we expect a significantly different analysis is needed in this regime. Regarding other works on the behaviour of the spectrum for heavy-tailed matrices, we refer to [13, 14, 16–18, 32, 33, 40] for instance.

1.3 Proof strategy

Our starting point is a decomposition of Y, or more precisely a resampling of Y, from the work [4]. Consider the Bernoulli $0 - 1$ random variables ψ_{ij} and χ_{ij} defined by

$$\mathbb{P}[\psi_{ij} = 1] = \mathbb{P}[|y_{ij}| \geq N^{-\epsilon_b}], \quad \mathbb{P}[\chi_{ij} = 1] = \frac{\mathbb{P}[|y_{ij}| \in [N^{-1/2-\epsilon_a}, N^{-\epsilon_b}]]}{\mathbb{P}[|y_{ij}| < N^{-\epsilon_b}]} \tag{5}$$

for some small positive constants ϵ_a, ϵ_b. In the sequel, we shall first choose ϵ_b and then choose $\epsilon_a = \epsilon_a(\epsilon_b, \alpha)$ to be sufficiently small. Specifically, throughout the discussion, we can make the following choice

$$0 < \epsilon_b < (\alpha - 2)/10\alpha, \qquad 0 < \epsilon_a < \min\{\epsilon_b, 4 - \alpha\}/10000. \tag{6}$$

Let a, b, and c be random variables such that

$$\mathbb{P}[a_{ij} \in I] = \frac{\mathbb{P}[y_{ij} \in (-N^{-1/2-\epsilon_a}, N^{-1/2-\epsilon_a}) \cap I]}{\mathbb{P}[|y_{ij}| \leq N^{-1/2-\epsilon_a}]},$$

$$\mathbb{P}[b_{ij} \in I] = \frac{\mathbb{P}[y_{ij} \in \left((-N^{-\epsilon_b}, -N^{-1/2-\epsilon_a}] \cup [N^{-1/2-\epsilon_a}, N^{-\epsilon_b})\right) \cap I]}{\mathbb{P}[|y_{ij}| \in [N^{-1/2-\epsilon_a}, N^{-\epsilon_b})]},$$

$$\mathbb{P}[c_{ij} \in I] = \frac{\mathbb{P}[y_{ij} \in \left((-\infty, -N^{-\epsilon_b}) \cup (N^{-\epsilon_b}, \infty)\right) \cap I]}{\mathbb{P}[|y_{ij}| \geq N^{-\epsilon_b}]}.$$

For each $(i, j) \in [M] \times [N]$, we set

$$\mathsf{A}_{ij} = (1 - \psi_{ij})(1 - \chi_{ij})a_{ij}, \quad \mathsf{B}_{ij} = (1 - \psi_{ij})\chi_{ij}b_{ij}, \quad \mathsf{C}_{ij} = \psi_{ij}c_{ij}$$

where a, b, c, ψ, χ-variables are all mutually independent. Sample Y and X by setting

$$Y = \mathsf{A} + \mathsf{B} + \mathsf{C}, \qquad X = \mathsf{B} + \mathsf{C}. \tag{7}$$

The dependence among A, B and C is then governed by the ψ and χ variables.

The purpose of the above decomposition, especially the separation of part A, is to view our model as a deformed model. We hope that the light-tailed part A can regularize the spectrum of the heavy-tailed part $X = \mathsf{B} + \mathsf{C}$, leading to the emergence of the edge universality. This idea is rooted in the dynamic approach developed in the last decade. We refer to the monograph [26] for a detailed introduction of this powerful approach, and also refer to [1, 3, 19, 24, 25, 34, 42–44] for instance. On a more specific level, our proof strategy is inspired by [4] where the authors consider the bulk statistics of the Lévy Wigner matrices in the regime $\alpha \in (0, 2)$, which we will denote by H in the sequel. In [4], the main idea to prove the bulk universality of the local statistics is to compare the Lévy Wigner matrix $H = \mathsf{A}_H + \mathsf{B}_H + \mathsf{C}_H$ with the Gaussian divisible model $H_t = \sqrt{t}W_H + \mathsf{B}_H + \mathsf{C}_H$, where $\mathsf{A}_H, \mathsf{B}_H$ and C_H are defined similarly to $\mathsf{A}, \mathsf{B}, \mathsf{C}$ above, and W_H is a GOE independent of H. Here t is chosen in such a way that $\sqrt{t}(W_H)_{ij}$ matches $(\mathsf{A}_H)_{ij}$ up to the third moment, conditioning on $(\psi_H)_{ij} = 0$, where ψ_H is defined similarly to ψ. Roughly speaking, the proof strategy of [4] is as follows. First, one needs to prove that the spectrum of $\mathsf{B}_H + \mathsf{C}_H$ satisfies an intermediate local law, which shows that the spectral density of $\mathsf{B}_H + \mathsf{C}_H$ is bounded below and above at a scale $\eta_* \leq N^{-\delta}t$. This control of the spectral density is also called η_*-regularity. Next, with the η_*-regularity established, one can use the results from [43] to prove that the $\sqrt{t}W_H$ component can improve the spectral regularity to the optimal (bulk) scale $\eta \geq N^{-1+\delta}$, and further obtain the bulk universality of H_t. Finally, one can prove that the bulk local eigenvalue statistics of H and H_t have the same asymptotic distribution by comparing the Green functions of H and H_t. However, the main difficulty here is that, unlike in H_t, the small part A_H and the major part $\mathsf{B}_H + \mathsf{C}_H$ in H are not independent. They are coupled by the ψ and χ variables. Despite this dependence being explicit, great effort has been made to carry out the comparison in [4].

At a high level, our proof strategy involves adapting the approach from [4] for the bulk regime to the left edge of the covariance matrices. However, this adaptation is far from being straightforward. We summarize some major ideas as follows.

1. (*Intermediate local law*) Similar to many previous DBM works, if we want to initiate the analysis, we need an intermediate local law for the $X = \mathsf{B} + \mathsf{C}$ part. More precisely, we require an η_*-regularity of the eigenvalue density for $\mathcal{S}(X) = XX^*$ at the left edge of the MP law, for some $\eta_* \ll 1$. According to [7], such a regularity cannot be true at the right edge of the spectrum. In order to explain heuristically the difference between the largest and smallest eigenvalues under the heavy-tailed assumption, we recall the variational definition of the smallest and largest singular values of X, which are also the square roots of the corresponding eigenvalues of $\mathcal{S}(X)$,

$$\sigma_M(X) = \inf_{v \in S^{M-1}} \|X^* v\|_2, \qquad \sigma_1(X) = \sup_{v \in S^{M-1}} \|X^* v\|_2. \tag{8}$$

Denote by v_M and v_1 the right singular vectors of X^* corresponding to $\sigma_M(X)$ and $\sigma_1(X)$, respectively. From the variational representation, it is clear that v_1 favors the large entry of X^*, and thus $\sigma_1(X)$ will be large as long as there is a big entry in X. This is indeed the case when the weak 4-th moment condition is not satisfied. In contrast, in (8), since v_M is the minimizer, it tries to avoid the big entries of X^*, i.e., it tends to live in the null space of C^*. Hence, heuristically, we can believe that removing the C entries will not significantly change the smallest singular value, as long as the null space of C is sufficiently big. This will be true if $\mathrm{rank}(\mathsf{C}) = o(N)$, which indeed holds when $\alpha > 2$. This simple heuristic explains why the first order behaviour of the smallest singular value of X, is more robust under the weak moment condition, in contrast to the largest singular value. It also indicates the following strategy for obtaining an intermediate local law for X. Let $\Psi = (\psi_{ij})$. We define the index sets

$$\mathcal{D}_r := \mathcal{D}_r(\Psi) := \left\{ i \in [M] : \sum_{j=1}^{N} \psi_{ij} \geq 1 \right\}, \qquad \mathcal{D}_c := \mathcal{D}_c(\Psi) := \left\{ j \in [N] : \sum_{i=1}^{M} \psi_{ij} \geq 1 \right\} \tag{9}$$

which are the index set of rows/columns in which one can find at least one nonzero ψ_{ij}. For any matrix $A \in \mathbb{C}^{M \times N}$, let $A^{(\mathcal{D}_r)}$ and $A^{[\mathcal{D}_c]}$ be the minors of A with the $\mathcal{D}_r$ rows and $\mathcal{D}_c$ columns removed, respectively, and we also use $\mathcal{S}(\mathcal{B}) = \mathcal{B}\mathcal{B}^*$ for any rectangle matrix $\mathcal{B}$ in the sequel. By Cauchy interlacing, we can easily see that

$$\lambda_M(\mathcal{S}(X^{[\mathcal{D}_c]})) \leq \lambda_M(\mathcal{S}(X)) \leq \lambda_{M-|\mathcal{D}_r|}(\mathcal{S}(X^{(\mathcal{D}_r)}))$$

Further notice that $X^{(\mathcal{D}_r)} = \mathsf{B}^{(\mathcal{D}_r)}$ and $X^{[\mathcal{D}_c]} = \mathsf{B}^{[\mathcal{D}_c]}$, and thus we have

$$\lambda_M(\mathcal{S}(\mathsf{B}^{[\mathcal{D}_c]})) \leq \lambda_M(\mathcal{S}(X)) \leq \lambda_{M-|\mathcal{D}_r|}(\mathcal{S}(\mathsf{B}^{(\mathcal{D}_r)})). \tag{10}$$

Conditioning on the matrix Ψ, we notice that both $\mathcal{S}(\mathsf{B}^{[\mathcal{D}_c]})$ and $\mathcal{S}(\mathsf{B}^{(\mathcal{D}_r)})$ are random matrices with bounded support, since $|b_{ij}| \leq N^{-\epsilon_b}$. For such matrices, one has a local law with precision $N^{-2\epsilon_b}$; see [37]. This local law together with (10) will give a rigidity estimate of $\lambda_M(\mathcal{S}(X))$ on scale $\eta_* = N^{-\epsilon_b}$ according to our choice in (6). Similarly applying the above row and column minor argument, one can derive an intermediate local law for X, which implies that X satisfies the η_*-regularity at the left edge. We remark here that in our regime $\alpha \in (2, 4)$, a weak intermediate local law, or alternatively, a weak regularity with $\eta_* \sim N^{-\varepsilon}$ for some small $\varepsilon > 0$ would be sufficient. This is always possible if we choose a suitable ϵ_b. In contrast, in the work [4], in the regime $\alpha \in (0, 2)$, a stronger regularity with a more carefully chosen η_* is actually needed.

2. (*Gaussian divisible ensemble*) We then consider the Gaussian divisible model

$$V_t := \sqrt{t}W + \mathsf{B} + \mathsf{C} = \sqrt{t}W + X, \qquad \mathcal{S}(V_t) = V_t V_t^*, \tag{11}$$

where $W = (w_{ij}) \in \mathbb{R}^{M \times N}$ is a Gaussian matrix with iid $N(0, N^{-1})$ entries, and $t = N\mathbb{E}|\mathsf{A}_{ij}|^2$ (slightly different from the choice in [4] for convenience). With the η_*-regularity of $\mathcal{S}(X)$, we then choose $1 \gg t \gg \sqrt{\eta_*}$. Actually, our t would be order $N^{-2\epsilon_a}$. By choosing ϵ_a sufficiently small in light of (6), our t can be sufficiently close to 1. By conditioning on the matrix X, the following edge universality can be achieved for the Gaussian divisible model $\mathcal{S}(V_t)$ by extending the result in [43] and [23] to the left edge

of the sample covariance matrices

$$N^{\frac{2}{3}}\gamma\big((\lambda_M(\mathcal{S}(V_t))-\lambda_{-,t}\big)\Rightarrow \mathrm{TW}_1, \tag{12}$$

for some constant γ, where $\lambda_{-,t}$ can be approximated by a mesoscopic statistic of the spectrum of $\mathcal{S}(X)$. Specifically,

$$\lambda_{-,t}=(1-c_N t m_X(\zeta_{-,t}))^2\zeta_{-,t}+(1-c_N)t(1+c_N t m_X(\zeta_{-,t})), \tag{13}$$

where m_X is the Stieltjes transform of the spectral distribution of $\mathcal{S}(X)$, and $\zeta_{-,t}$ is a random parameter defined through (26). We remark here that even though $\zeta_{-,t}$ is random, it can be proven that with a high probability, $\lambda_M(\mathcal{S}(X))-\zeta_{-,t}\sim t^2$. Hence, regarding the Stieltjes transform $m_X(\zeta_{-,t})$, we are at a (random) mesoscopic energy scale of order t^2. From the work [15, 51], one already knows that the global statistic $m_X(z)-\mathbb{E}m_X(z)$ follows a CLT on scale $N^{-\alpha/4}$ for a fixed z with $\mathrm{Im}\,z>0$. Due to the randomness of our parameter $\zeta_{-,t}$, a further expansion of it around a deterministic parameter ζ_{e} will be needed to adapt the argument in [15, 51]. Consequently, after the expansion, we will need to control the fluctuations of $m_X^{(k)}(\zeta_{\mathsf{e}})$ for $k=0,\dots,K$ with a sufficiently large K. Studying the fluctuations of these mesoscopic statistics eventually leads to a CLT

$$N^{\frac{\alpha}{4}}(\lambda_{-,t}-\mathbb{E}\lambda_{-,t})\Rightarrow N(0,\sigma_\alpha^2).$$

In addition to the above CLT, we need one more step to study the expansion of $\mathbb{E}\lambda_{-,t}$. It turns out that

$$\mathbb{E}\lambda_{-,t}=\lambda_-^{\mathsf{mp}}-N^{1-\frac{\alpha}{2}}s_\alpha+\mathfrak{o}(N^{1-\frac{\alpha}{2}}).$$

3. (*Green function comparison*)

Finally, we shall extend the result (12) from the Gaussian divisible model to our original matrix $\mathcal{S}(Y)$, using a Green function comparison inspired by [4]. It is now well-understood that one can compare certain functionals of the Green functions of two matrices instead of their eigenvalue distributions. Recall Y_t from (11), and we define the interpolations

$$Y^\gamma=\gamma\mathsf{A}+t^{1/2}(1-\gamma^2)^{1/2}W+\mathsf{B}+\mathsf{C},\qquad S^\gamma=Y^\gamma(Y^\gamma)^*,$$
$$G^\gamma(z)=(S^\gamma-z)^{-1},\qquad \mathcal{G}^\gamma(z)=((Y^\gamma)^*Y^\gamma-z)^{-1}\qquad m^\gamma(z)=\frac{1}{M}\mathrm{Tr}G^\gamma(z), \tag{14}$$

In order to extend (12) from $S^0=\mathcal{S}(V_t)$ to $S^1=\mathcal{S}(Y)$, from [54] for instance, we know that it suffices to establish the following result for some smooth bounded $F:\mathbb{R}\to\mathbb{R}$ with bounded derivatives

$$\left|\mathbb{E}F\Big(N\int_{E_1}^{E_2}\mathrm{d}E\,\mathrm{Im}\,m^1(\lambda_{-,t}+E+\mathrm{i}\eta_0)\Big)-\mathbb{E}F\Big(N\int_{E_1}^{E_2}\mathrm{d}E\,\mathrm{Im}\,m^0(\lambda_{-,t}+E+\mathrm{i}\eta_0)\Big)\right|\le N^{-\delta}, \tag{15}$$

where $E_1<E_2$, and $|E_i|\le N^{-\frac{2}{3}+\varepsilon}$ for $i=1,2$, and $\eta_0=N^{-\frac{2}{3}-\varepsilon}$, if we have the rigidity estimate

$$|\lambda_M(S^a)-\lambda_{-,t}|\prec N^{-\frac{2}{3}},\qquad a=0,1 \tag{16}$$

The estimate is easily available for the case $a=0$ (Gaussian divisible model) by a straightforward extension of [43] and [23]. This rigidity estimate for case $a=0$ is actually a technical input of getting (12). Hence, before the comparison in (15), we shall first prove (16) for $a=1$, again by a Green function comparison. We claim that it suffices to show for all $z_{-,t}=\lambda_{-,t}+\kappa+\mathrm{i}\eta$, with $|\kappa|\in N^{-\epsilon_b/2}$, and $\eta\in[N^{-\frac{2}{3}-\varepsilon},N^{-\varepsilon}]$ with some small $\varepsilon>0$,

$$\mathbb{E}\Big|N\eta(\mathrm{Im}\,m^1(z_{-,t})-\mathrm{Im}\,\tilde{m}^0(z_{-,t}))\Big|^{2k}\le(1+o(1))\mathbb{E}\Big|N\eta(\mathrm{Im}\,m^0(z_{-,t})-\mathrm{Im}\,\tilde{m}^0(z_{-,t}))\Big|^{2k}+N^{-\delta k}. \tag{17}$$

Similar estimate also holds when one replaces Im to Re. Here we introduced a copy of $m^0(z)$

$$\tilde{m}^0(z) = \frac{1}{M} \mathrm{Tr}(\sqrt{t}\tilde{W} + X - z)^{-1},$$

and $\tilde{W}$ is an iid copy of W. Actually, for the Gaussian divisible model, conditioning on X and extending the Theorem 3 in [22] on the deformed rectangle matrices from the right edge to the left edge, one can actually get the estimate

$$\left|\mathrm{Im}\, m^0(z_{-,t}) - \mathrm{Im}\, m_t(z_{-,t})\right| \prec \begin{cases} \frac{1}{N\eta}, & \text{if } \kappa \geq 0, \\ \frac{1}{N(|\kappa|+\eta)} + \frac{1}{(N\eta)^2\sqrt{|\kappa|+\eta}}, & \text{if } \kappa \leq 0, \end{cases} \tag{18}$$

where m_t is defined in (25). Apparently, the above estimates also hold with m^0 replaced by $\tilde{m}^0$. Combining these estimates with (17) leads to the bounds $|\mathrm{Im}\, m^1(z_{-,t}) - \mathrm{Im}\, m_t(z_{-,t})| \prec 1/(N\eta)$ when $\kappa > -N^{-\frac{2}{3}+\varepsilon}$ and $|\mathrm{Im}\, m^1(z_{-,t}) - \mathrm{Im}\, m_t(z_{-,t})| \ll 1/(N\eta)$ (w.h.p) when $\kappa \leq -N^{-\frac{2}{3}+\varepsilon}$. Such estimates together with the real part analogue of the former will finally lead to the rigidity estimate in (16).

The proofs of (15) and (17) are similar. We can turn to bound

$$\mathrm{d}\,\mathbb{E}F\left(N \int_{E_1}^{E_2} \mathrm{Im}\, m^\gamma(z_{-,t}^0) \mathrm{d}E\right)/\mathrm{d}\gamma \tag{19}$$

for $z_{-,t}^0 := \lambda_{-,t} + E + \mathrm{i}\eta_0$ with $\eta_0 = N^{-\frac{2}{3}-\varepsilon}$, and

$$\mathrm{d}\,\mathbb{E}|N\eta(\mathrm{Im}\, m^\gamma(z_{-,t}) - \mathrm{Im}\, \tilde{m}^0(z_{-,t}))|^{2k}/\mathrm{d}\gamma \tag{20}$$

for $z_{-,t} = \lambda_{-,t} + E + \mathrm{i}\eta$, where $E \in (-N^{-\epsilon_b/2}, N^{-\frac{2}{3}+\epsilon})$ and $\eta = N^{-\frac{2}{3}}$. Actually, we shall first condition on Ψ, and then first estimate $\mathbb{E}_\Psi$ and then use a law of total expectation to estimate the full expectation. When one try to take the derivatives in (19)-(20) and estimate the resulting terms, we will need a priori bounds for the Green function entries

$$G_{ij}^\gamma(z), \quad \mathcal{G}_{uv}^\gamma(z), \quad ((Y^\gamma)^*G^\gamma(z))_{ui} \tag{21}$$

in the domain

$$\mathsf{D} = \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3) := \{z = \lambda_-^{\mathsf{mp}} + E + \mathrm{i}\eta : |E| \leq N^{-\varepsilon_1}, \eta \in [N^{-\frac{2}{3}-\varepsilon_2}, \varepsilon_3]\} \tag{22}$$

with appropriately chosen small constants $\varepsilon_1, \varepsilon_2, \varepsilon_3$. We shall show that most of these entries are stochastically dominated by 1 while a small amount of them are stochastically dominated by $1/t^2$. These bounds are not even known for the Gaussian divisible case, i.e., $\gamma = 0$, at the edge. The idea is to first prove the desired bounds of the quantities in (21) for $\gamma = 0$, and then prove another comparison result for the Green functions

$$\left|\mathbb{E}|G_{ij}^\gamma(z)|^{2k} - \mathbb{E}|G_{ij}^0(z)|^{2k}\right| \leq N^{-\delta} \tag{23}$$

for all $z \in \mathsf{D}$. Here we refer to [4] and [41] for similar strategy of using comparison to prove Green function bounds on local scale. Hence, based on the above discussion, the proof route is as following

$$\boxed{\text{bounds of (21) for } \gamma = 0} \longrightarrow \boxed{(23)} \longrightarrow \boxed{(17)} \longrightarrow \boxed{(15)}$$

which requires a three steps of Green function comparison with different observables. In contrast, in [4], one Green function comparison for the observable $F(\mathrm{Im}\, G_{a_1b_1}(z), \cdots, \mathrm{Im}\, G_{a_mb_m}(z))$ (and its real part analogue) with a deterministic parameter z in the bulk regime would be sufficient. Also notice that our parameter $z_{-,t}$ in (15) is random, which further complicates the comparison. Specifically, when we do expansions of the Green function entries w.r.t. the matrix entries, we shall also keep tracking the

derivatives of $\lambda_{-,t}$ w.r.t to these entries. The estimates of these derivatives involve delicate analysis of the subordination equations.

Regarding the bounds of (21) for $\gamma = 0$, here we shall explain the argument for G_{ij} only for simplicity. The other two kinds of entries in (21) can be handled similarly. For the Gaussian divisible model, conditioning on X, by extending the Theorem 3 in [22] on the deformed rectangle matrices from the right edge to the left edge, with the η_*-regularity of the spectrum of $\mathcal{S}(X)$, we have for $z \in \mathsf{D}$

$$|G_{ij}^0(z) - \Pi_{ij}(z)| \prec \Big(t\Big(\sqrt{\frac{\operatorname{Im} m_t(z)}{N\eta}} + \frac{1}{N\eta}\Big) + \frac{t^{1/2}}{N^{1/2}}\Big)\varpi^{-2}(z),$$

where ϖ is of order $t^2 + \eta$, and

$$\Pi_{ij} = (1 + c_N t m_t)\big(XX^* - \zeta_t(z))\big)_{ij}^{-1} =: (1 + c_N t m_t)G_{ij}(X, \zeta_t(z)).$$

which is simply a multiple of a Green function entries of $\mathcal{S}(X)$, but evaluated at a random parameter $\zeta_t(z)$. By the facts $t \sim N^{-2\epsilon_a}$, $\eta \gtrsim N^{-\frac{2}{3}-\varepsilon_2}$ and $|m_t(z)| \leq (ct|z|)^{-1/2}$ (cf. Lemma 2.1 (iv)), one can easily get $|G_{ij}^0(z) - \Pi_{ij}(z)| \prec 1$. Hence, what remains is to bound Π_{ij}, i.e., to bound $G_{ij}(X, \zeta_t(z))$, the Green function entry of the heavy-tailed covariance matrix $\mathcal{S}(X)$, in the regime $2 < \alpha < 4$. We notice that such a bound has been obtained in [2] for the heavy-tailed Wigner matrices in the same regime of α, but in the bulk. Extending such a bound to edge could be difficult due to the deterioration of the stability of self-consistent equation of the Stieltjes transform. However, we notice that with the η_*-regularity of the left edge of $\mathcal{S}(X)$ spectrum, one can show that the parameter $\zeta_t(z)$ is away from the left edge of the $\mathcal{S}(X)$ spectrum by a distance of order t^2. Hence, we are away from the edge by a mesoscopic distance, which allow us the conduct the argument similarly to the bulk case in [2] to get the desired bound for $\Pi_{ij}(z)$.

1.4 Organization

The rest of the paper will be organized as follows. In Section 2, we will state the main results for the Gaussian divisible model, whose proofs will be stated in Section 3. Section 4 is devoted to the statements of the Green function comparisons and prove our main theorem based on the comparisons. In Section 5, we prove these comparison results. Some technical estimates are stated in the supplementary material [11].

1.5 Notation

Throughout this paper, we regard N as our fundamental large parameter. Any quantities that are not explicit constant or fixed may depend on N; we almost always omit the argument N from our notation. We use $\|u\|_\alpha$ to denote the ℓ^α-norm of a vector u. We further use $\|A\|$ for the operator norm of a matrix A. We use C to denote some generic (large) positive constant. The notation $a \sim b$ means $C^{-1}b \leq |a| \leq Cb$ for some positive constant C. Similarly, we use $a \lesssim b$ to denote the relation $|a| \leq Cb$ for some positive constant C. $\mathcal{O}$ and $\mathfrak{o}$ denote the usual big and small O notation, and $\mathcal{O}_p$ and $\mathfrak{o}_p$ denote the big and small O notation in probability. When we write $a \ll b$ and $a \gg b$ for possibly N-dependent quantities $a = a(N)$ and $b = b(N)$, we mean $|a|/b \to 0$ and $|a|/b \to \infty$ when $N \to \infty$, respectively. For any positive integer n, let $[n] = [1:n]$ denote the set $\{1,\ldots,n\}$. For $a, b \in \mathbb{R}$, $a \vee b = \max\{a,b\}$ and $a \wedge b = \min\{a,b\}$. For a square matrix $A \in \mathbb{R}^{n \times n}$, we let $A_{\mathrm{diag}} = (A_{ij}\delta_{ij}) \in \mathbb{R}^{n \times n}$. We adopt the following Green function notation for any rectangle matrix $A \in \mathbb{R}^{m \times n}$, $G(A, z) = (AA^\top - z)^{-1}$.

2 Gaussian divisible model

In this section, we state the main results for a Gaussian divisible model, and leave the detailed proofs to the next section.

2.1 Some definitions

Recall that $W = (w_{ij}) \in \mathbb{R}^{M \times N}$ is a Gaussian matrix with iid $N(0, N^{-1})$ entries, and $t = N\mathbb{E}|\mathsf{A}_{ij}|^2$. Consider the standard signal-plus-noise model

$$V_t := X + \sqrt{t}W. \tag{24}$$

In this section, we will establish several spectral properties of $\mathcal{S}(V_t)$ that will be extended to $\mathcal{S}(Y)$ later. For most of the discussion in this part, we will condition on X and regard it as given, and work with the randomness of W. In light of this, we introduce the asymptotic eigenvalue density of $\mathcal{S}(V_t)$, denoted by ρ_t, through its corresponding Stieltjes transform $m_t := m_t(z)$. For any $t > 0$, m_t is known to be the unique solution to the following equation:

$$m_t = \frac{1}{M} \sum_{i=1}^{M} \frac{1 + c_N t m_t}{\lambda_i(\mathcal{S}(X)) - \zeta_t}, \tag{25}$$

subject to the condition that $\operatorname{Im} m_t > 0$ for any $z \in \mathbb{C}_+$. Here

$$\zeta_t := \zeta_t(z) := (1 + c_N t m_t(z))^2 z - t(1 - c_N t)(1 - c_N t m_t(z)). \tag{26}$$

In the context of free probability theory, ρ_t corresponds to the rectangular free convolution of the spectral distribution of $\mathcal{S}(X)$ with the MP law on scale t, and ζ_t is the so-called subordination function for the rectangular free convolution. The following lemma provides a precise description of the existence and uniqueness of the asymptotic density. The following result holds for any realization of X.

Lemma 2.1 (Existence and uniqueness of asymptotic density, Lemma 2 of [22]). *For any $t > 0$, the following properties hold.*

(i) There exists a unique solution m_t to equation (25) satisfying that $\operatorname{Im} m_t(z) > 0$ and $\operatorname{Im} z m_t(z) > 0$ if $z \in \mathbb{C}^+$.

(ii) For all $E \in \mathbb{R} \setminus \{0\}$, $\lim_{\eta \downarrow 0} m_t(E + i\eta)$ exits, and we denote it as $m_t(E)$. The function m_t is continuous on $\mathbb{R} \setminus \{0\}$, and $\rho_t(E) := \pi^{-1} \operatorname{Im} m_t(E)$ is a continuous probability density function on $\mathbb{R}^+ := \{E \in \mathbb{R} : E > 0\}$. Moreover, m_t is the Stieltjes transform of ρ_t. Finally, $m_t(E)$ is a solution to (25) for $z = E$.

(iii) For all $E \in \mathbb{R} \setminus \{0\}$, $\lim_{\eta \downarrow 0} \zeta_t(E + i\eta)$ exits, and we denote it as $\zeta_t(E)$. Moreover, we have $\operatorname{Im} \zeta_t(z) > 0$ if $z \in \mathbb{C}^+$.

(iv) We have $\operatorname{Re}(1 + c_N t m_t(z)) > 0$ for all $z \in \mathbb{C}^+$ and $|m_t(z)| \leq (c_N t |z|)^{-1/2}$.

For a realization of X, we can check if it satisfies the following regularity condition on $m_X(z)$. Such a condition is crucial for the edge universality of DBM; see [22, 43] for instance.

Definition 1 (η_*- regularity). *Let η_* be a parameter satisfying $\eta_* := N^{-\tau_*}$ for some constant $0 < \tau_* \leq 2/3$. For an $M \times N$ matrix H, we say $\mathcal{S}(H)$ is η_*-regular around the left edge $\lambda_- = \lambda_M(\mathcal{S}(H))$ if there exist constants $c_H > 0$ and $C_H > 1$ such that the following properties hold:*

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

$$\frac{1}{C_H} \sqrt{|E - \lambda_-| + \eta} \leq \operatorname{Im} m_H(E + i\eta) \leq C_H \sqrt{|E - \lambda_-| + \eta}. \tag{27}$$

For $z = E + i\eta$ with $\lambda_- - c_H \leq E \leq \lambda_-$ and $\eta_* \leq \eta \leq 10$, we have

$$\frac{1}{C_H} \frac{\eta}{\sqrt{|E - \lambda_-| + \eta}} \leq \operatorname{Im} m_H(E + i\eta) \leq C_H \frac{\eta}{\sqrt{|E - \lambda_-| + \eta}}. \tag{28}$$

(ii) We have $c_H/2 \leq \lambda_- \leq 2C_H$.
(iii) We have $\|\mathcal{S}(H)\| \leq N^{C_H}$.

The following lemma is a direct implication of η_*- regularity.

Lemma 2.2 (Lemma 6 of [22]). *Suppose (a realization of) $\mathcal{S}(X)$ is η_*-regular in the sense of Definition 1. Let μ_X be the measure associated with $m_X(z)$. For any fixed integer $k \geq 2$, and any $z \in \mathcal{D}$ with*

$$\mathcal{D} := \left\{ z = E + \mathrm{i}\eta : \lambda_- - \frac{3}{4}\tilde{c} \leq E \leq \lambda_-, 2\eta_* \leq \eta \leq 10 \right\}$$

$$\cup \left\{ z = E + \mathrm{i}\eta : \lambda_- \leq E \leq \lambda_- + \frac{3}{4}\tilde{c}, \eta_* + \sqrt{\eta_*(E - \lambda_-)} \leq \eta \leq 10 \right\}$$

$$\cup \left\{ z = E + \mathrm{i}\eta : \lambda_- - \frac{3}{4}\tilde{c} \leq E \leq \lambda_- - 2\eta_*, 0 \leq \eta \leq 10 \right\}.$$

Then we have

$$\int \frac{\mathrm{d}\mu_X(x)}{|x - E - \mathrm{i}\eta|^k} \sim \frac{\sqrt{|E - \lambda_-| + \eta}}{\eta^{k-1}}\mathbf{1}_{E \geq \lambda_-} + \frac{1}{(|E - \lambda_-| + \eta)^{k-3/2}}\mathbf{1}_{E < \lambda_-}.$$

The following notion of stochastic domination which originated from [29] will be used throughout the paper.

Definition 2 (Stochastic domination). *Let $\mathsf{X} = (\mathsf{X}^{(N)}(u) : N \in \mathbb{N}, u \in \mathrm{U}^{(N)}), \mathsf{Y} = (\mathsf{Y}^{(N)}(u) : N \in \mathbb{N}, u \in \mathrm{U}^{(N)})$ be two families of random variables, where Y is nonnegative, and $\mathrm{U}^{(N)}$ is a possibly N-dependent parameter set. We say that X is stochastically dominated by Y, uniformly in u, if for all small $\epsilon > 0$ and large $D > 0$,*

$$\sup_{u \in \mathrm{U}^{(N)}} \mathbb{P}\left(\left|\mathsf{X}^{(N)}(u)\right| > N^\varepsilon \mathsf{Y}^{(N)}(u) \right) \leqslant N^{-D}$$

for large enough $N > N_0(\epsilon, D)$. If X is stochastically dominated by Y, uniformly in u, we use the notation $\mathsf{X} \prec \mathsf{Y}$, or equivalently $\mathsf{X} = O_\prec(\mathsf{Y})$. Note that in the special case when X and Y are deterministic, $\mathsf{X} \prec \mathsf{Y}$ means that for any given $\epsilon > 0$, $|\mathsf{X}^{(N)}(u)| \leq N^\epsilon \mathsf{Y}^{(N)}(u)$ uniformly in u, for all sufficiently large $N \geq N_0(\epsilon)$.

2.2 η_*- regularity of $\mathcal{S}(X)$: A matrix minor argument

In this subsection, we state that with high probability η^*-regularity holds for $\mathcal{S}(X)$ with $\eta^* = N^{-\epsilon_b}$. Recall that X defined in (7). Let us recall $\Psi = (\psi_{ij}) \in \mathbb{R}^{M \times N}$, a random matrix with entries ψ_{ij} as defined in (5). By setting

$$\epsilon_\alpha = (\alpha - 2)/5\alpha, \tag{29}$$

we call a Ψ *good* if it has at most $N^{1-\epsilon_\alpha}$ entries equal to 1. The following lemma indicates that Ψ is, indeed, good with high probability.

Lemma 2.3. *For any large $D > 0$, we have $\mathbb{P}(\Omega_\Psi = \{\Psi \text{ is good}\}) \geq 1 - N^{-D}$.*

Proof. Observe that $\mathbb{P}(\Omega_\Psi = \{\Psi \text{ is good}\}) = 1 - \mathbb{P}(\#\{(i, j) : x_{ij} > N^{-\epsilon_b}\} > N^{1-\epsilon_\alpha})$. By Assumption 1 (i), we have

$$\mathbb{P}(\#\{(i, j) : x_{ij} > N^{-\epsilon_b}\} > N^{1-\epsilon_\alpha}) \lesssim \sum_{j=N^{1-\epsilon_\alpha}}^{N^2} \binom{N^2}{j} N^{-\alpha(1/2-\epsilon_b)j} \lesssim \sum_{j=N^{1-\epsilon_\alpha}}^{N^2} N^{-(\alpha-2)j/2} \lesssim N^{-D}.$$

The claim now follows by possibly adjusting the constants. □

Given any Ψ is good, the following proposition shows that $\mathcal{S}(X)$ is η_*-regular with $\eta_* = N^{-\tau_*}$ for some $\tau_* > 0$. Actually, we shall work with a truncation of X, $X^\mathcal{C} := (x_{ij}\mathbf{1}_{|x_{ij}| \leq N^{100}})_{i \in [M], j \in [N]}$, in order to guarantee Definition 1 (iii). Apparently, $\|\mathcal{S}(X^\mathcal{C})\| \leq N^{102}$ and $\mathbb{P}(X = X^\mathcal{C}) = 1 - \mathfrak{o}(1)$.

Proposition 2.4 (η_*- regularity of $\mathcal{S}(X)$). *Suppose that Ψ is good. Let $\eta_* = N^{-\epsilon_b}$. Then $\mathcal{S}(X)$ is η_*-regular around its smallest eigenvalue $\lambda_M(\mathcal{S}(X))$ in the sense of Definition 1 with high probability.*

The proof of Proposition 2.4 is based on the following two lemmas. For notational simplicity, we define $\mathsf{m}_{\mathsf{mp}}^{(t)}(z) := (1 - t)^{-1}\mathsf{m}_{\mathsf{mp}}(z/(1 - t))$ for any $t > 0$.

Lemma 2.5. *Fix $C > 0$. Let us consider $z \in \{E + i\eta : C^{-1}\lambda_-^{\mathsf{mp}} \leq E \leq \lambda_+^{\mathsf{mp}} + 1, 0 < \eta < 3\}$. We have*

$$|m_{\mathsf{B}}(z) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z)| \prec N^{-\epsilon_b} + (N\eta)^{-1}. \tag{30}$$

In addition,

$$|\lambda_M(\mathcal{S}(\mathsf{B})) - (1 - t)\lambda_-^{\mathsf{mp}}| \prec N^{-2\epsilon_b} + N^{-2/3}. \tag{31}$$

Proof. We further denote by $\tilde{t} := 1 - N\mathbb{E}|\mathsf{B}_{ij}|^2$. It is easy to show that $|\tilde{t} - t| = \mathfrak{o}(N^{-1})$, and thus we have $|\mathsf{m}_{\mathsf{mp}}^{(t)}(z) - \mathsf{m}_{\mathsf{mp}}^{(\tilde{t})}(z)| \leq (N\eta)^{-1}$. Hence, it suffices to show the following estimates

$$|m_{\mathsf{B}}(z) - \mathsf{m}_{\mathsf{mp}}^{(\tilde{t})}(z)| \prec N^{-\epsilon_b} + (N\eta)^{-1}, \qquad |\lambda_M(\mathcal{S}(\mathsf{B})) - (1 - \tilde{t})\lambda_-^{\mathsf{mp}}| \prec N^{-2\epsilon_b} + N^{-2/3}. \tag{32}$$

Notice that B is a so-called random matrix with bounded support. The first estimate in (32) is given by [37, Theorem 2.7]. We can show the second estimate in (32) adapting the proof of [37, Theorem 2.9] from the right edge to the left edge, in a straightforward way, given a crude lower bound of $\lambda_M(\mathcal{S}(\mathsf{B}))$ which is guaranteed by [61]. We omit the details. $\qquad\square$

Lemma 2.6. *Suppose Ψ is good. Then, we have $|\lambda_M(\mathcal{S}(X)) - (1 - t)\lambda_-^{\mathsf{mp}}| \prec N^{-2\epsilon_b}$.*

Proof. Denote by $\mathfrak{N}(\mathsf{C})$ the number of nonzero columns of C. Since Ψ is good, $|\mathfrak{N}(\mathsf{C})| \leq N^{1-\epsilon_\alpha}$, with high probability. By Cauchy interlacing, we can easily see that

$$\lambda_M(\mathcal{S}(X^{[\mathcal{D}_c]})) \leq \lambda_M(\mathcal{S}(X)) \leq \lambda_{M-|\mathcal{D}_r|}(\mathcal{S}(X^{(\mathcal{D}_r)}))$$

Further notice that $X^{(\mathcal{D}_r)} = \mathsf{B}^{(\mathcal{D}_r)}$ and $X^{[\mathcal{D}_c]} = \mathsf{B}^{[\mathcal{D}_c]}$, and thus we have

$$\lambda_M(\mathcal{S}(\mathsf{B}^{[\mathcal{D}_c]})) \leq \lambda_M(\mathcal{S}(X)) \leq \lambda_{M-|\mathcal{D}_r|}(\mathcal{S}(\mathsf{B}^{(\mathcal{D}_r)})).$$

Applying (31) to $\mathcal{S}(\mathsf{B}^{[\mathcal{D}_c]})$ and $\mathcal{S}(\mathsf{B}^{(\mathcal{D}_r)})$ with the modified parameter c_N, i.e., $M/(N - |\mathcal{D}_c|)$ and $(M - |\mathcal{D}_r|)/N$ respectively, in the definition of λ_-^{mp}, we can prove the conclusion with the fact $\epsilon_\alpha > 2\epsilon_b$. $\qquad\square$

Now we show the proof of Proposition 2.4.

Proof of Proposition 2.4. We shall show three properties (i), (ii) and (iii) (as in Definition 1) holds with high probability. Suppose that Ψ is good.

(i). Let μ_X and μ_{B} be the empirical spectral distributions of $\mathcal{S}(X)$ and $\mathcal{S}(\mathsf{B})$, respectively. By the rank inequality [8, Theorem A.44], $|\mu_X - \mu_{\mathsf{B}}| \leq 2\mathrm{rank}(\mathsf{C})/N$. Then,

$$|\mathrm{Im}\, m_X(z) - \mathrm{Im}\, m_{\mathsf{B}}(z)| \leq \int \left| \frac{\eta}{(\lambda - E)^2 + \eta^2}(\mu_X - \mu_{\mathsf{B}})(d\lambda) \right|.$$

It follows from $\eta((\lambda - E)^2 + \eta^2)^{-1} \leq \eta^{-1}$ that $|\mathrm{Im}\, m_X(z) - \mathrm{Im}\, m_{\mathsf{B}}(z)| \lesssim \mathrm{rank}(\mathsf{C})/(N\eta) = N^{-\epsilon_\alpha}\eta^{-1}$, where we use the assumption that Ψ is good. This together with Lemma 2.5 give

$$|\mathrm{Im}\, m_X(z) - \mathrm{Im}\, \mathsf{m}_{\mathsf{mp}}^{(t)}(z)| \prec N^{-\epsilon_\alpha}\eta^{-1} + N^{-\epsilon_b} + (N\eta)^{-1}.$$

For $E \in [\lambda_M(\mathcal{S}(X)), \lambda_M(\mathcal{S}(X)) + \eta_*]$, by Lemma 2.6, we have with high probability that,

$$|E - (1 - t)\lambda_-^{\mathsf{mp}}| \leq |E - \lambda_M(\mathcal{S}(X))| + |\lambda_M(\mathcal{S}(X)) - (1 - t)\lambda_-^{\mathsf{mp}}| \leq 2\eta_*.$$

Thus, for $\eta \geq \eta_*$, we have $\mathrm{Im}\, \mathsf{m}_{\mathsf{mp}}^{(t)}(z) \sim \sqrt{\eta}$, which implies that $\mathrm{Im}\, m_X(z) \sim \sqrt{|E - \lambda_M(\mathcal{S}(X))| + \eta}$. Similarly, for $E \in [\lambda_M(\mathcal{S}(X)) - \eta_*, \lambda_M(\mathcal{S}(X))]$ and $\eta \geq \eta_*$, we can show that $\mathrm{Im}\, m_X(z) \sim \eta/\sqrt{|E - \lambda_M(\mathcal{S}(X))| + \eta}$. If $E \geq \lambda_M(\mathcal{S}(X)) + \eta_*$, we can use the fact $|\lambda_M(\mathcal{S}(X)) - (1 - t)\lambda_-^{\mathsf{mp}}| \ll \eta_*$ to obtain that $E \geq (1 - t)\lambda_-^{\mathsf{mp}}$ and

$$\sqrt{|E - \lambda_M(\mathcal{S}(X))| + \eta} \sim \sqrt{|E - (1 - t)\lambda_-^{\mathsf{mp}}| + \eta}.$$

11

Similarly, if $E \leq \lambda_M(\mathcal{S}(X)) - \eta_*$, we obatin $E \leq (1-t)\lambda_-^{\mathsf{mp}}$ and

$$\frac{\eta}{\sqrt{|E - \lambda_M(\mathcal{S}(X))| + \eta}} \sim \frac{\eta}{\sqrt{|E - (1-t)\lambda_-^{\mathsf{mp}}| + \eta}}.$$

(ii). It holds with high probability by Lemma 2.5. (iii). See Remark 2 below. Therefore, we conclude the proof. $\qquad\square$

Remark 2. *Rigorously speaking, in order to have the above proposition, we shall work with $X^{\mathcal{C}}$ instead of X. Since these two matrices are identical with probability $1 - \mathfrak{o}(1)$, any spectral statistics of these two matrices are identical with probability $1 - \mathfrak{o}(1)$. For our main theorem, it would be sufficient to work with $X^{\mathcal{C}}$ instead of X in the sequel. However, for convenience, we will still work with X as if the above proposition is also true for $\mathcal{S}(X)$. In this case, the reader may simply assume that the entries of X are bounded by N^{100} (say). We can anyway recover the result without this additional boundedness assumption by comparing the matrix with its truncated version.*

Let $\lambda_{-,t}$ be the left edge of ρ_t. The Gaussian part in model (24) can further improve the scale of the square root behavior of ρ_t around $\lambda_{-,t}$ on the event that $\mathcal{S}(X)$ satisfies certain η_*- regularity. The following theorem makes this precise.

Theorem 2.7 (Lemma 1 of [22]). *On Ω_Ψ, we have*

$$\rho_t \sim \sqrt{(E - \lambda_{-,t})_+} \quad for \quad \lambda_{-,t} - \frac{3}{4}\tilde{c} \leq E \leq \lambda_{-,t} + \frac{3}{4}\tilde{c},$$

and for $z = E + \mathrm{i}\eta \in \mathbb{C}^+$,

$$\mathrm{Im}\, m_t(z) \sim \begin{cases} \sqrt{|E - \lambda_{-,t}| + \eta}, & \lambda_{-,t} \leq E \leq \lambda_{-,t} + \frac{3}{4}\tilde{c} \\ \dfrac{\eta}{\sqrt{|E - \lambda_{-,t}| + \eta}}, & \lambda_{-,t} - \frac{3}{4}\tilde{c} \leq E \leq \lambda_{-,t} \end{cases}. \tag{33}$$

Next, we recall the definition in (22). The following theorem provide bounds on the Green function entries for the Gaussian divisible model. Further recall the notation in (9), we set $\mathcal{T}_r := [M] \setminus \mathcal{D}_r$, $\mathcal{T}_c := [N] \setminus \mathcal{D}_c$.

Theorem 2.8. *Suppose that Ψ is good. Let $z \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ with $10\epsilon_a \leq \varepsilon_1 \leq \epsilon_b/500$ and sufficiently small $\varepsilon_2, \varepsilon_3$. The following estimates hold w.r.t. the probability measure $\mathbb{P}_\Psi$.*

(i)
$$|G_{ij}(V_t, z)| \prec \mathbf{1}_{i \in \mathcal{T}_r \text{ or } j \in \mathcal{T}_r} + t^{-2}(1 - \mathbf{1}_{i \in \mathcal{T}_r \text{ or } j \in \mathcal{T}_r}),$$

(ii)
$$|G_{uv}(V_t^\top, z)| \prec \mathbf{1}_{u \in \mathcal{T}_c \text{ or } v \in \mathcal{T}_c} + t^{-2}(1 - \mathbf{1}_{u \in \mathcal{T}_c \text{ or } v \in \mathcal{T}_c}),$$

(iii)
$$|[G(V_t, z)V_t]_{iu}| \prec N^{-\epsilon_b/2}\mathbf{1}_{i \in \mathcal{T}_r \text{ or } u \in \mathcal{T}_c} + t^{-2}(1 - \mathbf{1}_{i \in \mathcal{T}_r \text{ or } u \in \mathcal{T}_c}).$$

The proof of Theorem 2.8 is based on the following results.

Lemma 2.9. *Suppose that the assumptions in Theorem 2.8 hold. There exist constants $c, C > 0$ such that for the domain $\mathsf{D}_\zeta = \mathsf{D}_\zeta(c, C) \subset \mathbb{C}_+$ defined by*

$$\mathsf{D}_\zeta := \mathsf{D}_1 \cup \mathsf{D}_2, \tag{34}$$

where

$$\mathsf{D}_1 := \{\zeta = E + \mathrm{i}\eta : E \leq (1-t)\lambda_-^{\mathsf{mp}} - ct^2, \eta \geq ctN^{-2/3-\varepsilon_2}\}$$
$$\mathsf{D}_2 := \{\zeta = E + \mathrm{i}\eta : \eta \geq c(\log N)^{-C}t^2\}$$

we have $\zeta_t(z) \in \mathsf{D}_\zeta$ with high probability.

Proof. The proof relies on the definition of $\zeta_t(z)$ as well as the square root behaviour of ρ_t as stated in Theorem 2.7; see supplementary material [11] for the detailed proof. □

Proposition 2.10. *Let D_ζ be as in (34). Consider $\zeta \in \mathsf{D}_\zeta$. Suppose that Ψ is good. The following estimates hold w.r.t. the probability measure $\mathbb{P}_\Psi$. There exists a constant $c = c(\epsilon_a, \epsilon_\alpha, \epsilon_b) > 0$ such that*

$$|G_{ij}(X, \zeta) - \delta_{ij}\mathsf{m}_{\mathsf{mp}}^{(t)}(\zeta)| \prec N^{-c}\mathbf{1}_{i,j\in\mathcal{T}_r} + t^{-2}(1 - \mathbf{1}_{i,j\in\mathcal{T}_r}),$$
$$|G_{uv}(X^\top, \zeta) - \delta_{uv}\underline{\mathsf{m}}_{\mathsf{mp}}^{(t)}(\zeta)| \prec N^{-c}\mathbf{1}_{u,v\in\mathcal{T}_c} + t^{-2}(1 - \mathbf{1}_{u,v\in\mathcal{T}_c}),$$

where $\underline{\mathsf{m}}_{\mathsf{mp}}^{(t)}(\zeta) = c_N\mathsf{m}_{\mathsf{mp}}^{(t)}(\zeta) - (1 - c_N)/\zeta$.

Proof. The proof of Proposition 2.10 is similar to the light-tailed case proved in [54], but here we shall apply large deviation formula for heavy-tailed random variables; see supplementary material [11] for the detailed proof. □

Proof of Theorem 2.8. Given the previous results, the proof strategy for this theorem is briefly introduced in the last paragraph of the Introduction, Section 1, with the detailed proof found in the supplementary materials [11]. □

The above theorems provide strong evidence supporting the validity of the Tracy-Widom law for $\lambda_M(\mathcal{S}(V_t))$ around $\lambda_{-,t}$. In fact, we are able to establish the following theorem regarding the convergence of the distribution. Before stating the result, we define the function

$$\Phi_t(\zeta) := (1 - c_N t m_X(\zeta))^2 \zeta + (1 - c_N)t(1 - c_N t m_X(\zeta)), \tag{35}$$

and the scaling parameter

$$\gamma_N := \gamma_N(t) := -\left(\frac{1}{2}\left[4\lambda_{-,t}\zeta_t(\lambda_{-,t}) + (1 - c_N)^2 t^2\right]c_N^2 t^2 \Phi_t''(\zeta_t(\lambda_{-,t}))\right)^{-1/3}. \tag{36}$$

Theorem 2.11. *Let $f : \mathbb{R} \to \mathbb{R}$ be a test function satisfying $\|f\|_\infty \leq C$ and $\|\nabla f\|_\infty \leq C$ for a constant C. Then we have for any X whose corresponding Ψ is good,*

$$\lim_{N\to\infty} \mathbb{E}\left[f\left(\gamma_N M^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t})\right)|X\right] = \lim_{N\to\infty} \mathbb{E}\left[f\left(M^{2/3}(\mu_M^{\mathsf{GOE}} + 2)\right)\right]. \tag{37}$$

This further implies that if Ψ is good,

$$\lim_{N\to\infty} \mathbb{E}_\Psi\left[f\left(\gamma_N M^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t})\right)\right] = \lim_{N\to\infty} \mathbb{E}\left[f\left(M^{2/3}(\mu_M^{\mathsf{GOE}} + 2)\right)\right], \tag{38}$$

where μ_M^{GOE} denotes the least eigenvalue of a M by M Gaussian Orthogonal Ensemble (GOE) with $N(0, M^{-1})$ off-diagonal entries.

Remark 3. *The proof of the above theorem is essentially an adapt of the edge universality for the DBM in [43] and the analogue for the rectangle DBM in [22, 23]. More specifically, we shall extend the analysis in [22, 23] from the right edge of the covariance type matrix to the left edge. Based on the η_*-regularity, the proof is nearly the same as [22, 23], and thus we do not reproduce the details and only provide some remarks in the supplementary material [11].*

2.3 Distribution of $\lambda_{-,t}$

Theorem 2.12. *There exists a deterministic quantity $\lambda_{\mathsf{shift}} > 0$ depending on N such that the following two properties hold.*

(i)

$$\frac{N^{\alpha/4}(\lambda_{-,t} - \lambda_{\mathsf{shift}})}{\sigma_\alpha} \Rightarrow \mathcal{N}(0,1), \quad \sigma_\alpha^2 = \frac{\mathsf{c}c_N^{(4-\alpha)/4}(1 - \sqrt{c_N})^4(\alpha - 2)}{2}\Gamma\left(\frac{\alpha}{2} + 1\right).$$

(ii)
$$\lambda_{\mathsf{shift}} = \lambda_-^{\mathsf{mp}} - \frac{\mathsf{c}N^{1-\alpha/2}(1-\sqrt{c_N})^2}{c_N^{(\alpha-2)/4}}\Gamma\Big(\frac{\alpha}{2}+1\Big) + \mathfrak{o}(N^{1-\alpha/2}).$$

Remark 4. *Note that the leading order of λ_{shift} only depends on α. The size of the fluctuation of $\lambda_{-,t}$ is also determined by α.*

The proof of Theorem 2.12 is given in the next section.

3 Proofs for Gaussian divisible model

3.1 Preliminary estimates

Before providing the preliminary estimates for the expansion of the least eigenvalue of $\mathcal{S}(V_t)$, we first state the following lemma, which characterizes the support of ρ_t and its edges using the local extrema of $\Phi_t(\zeta)$ on $\mathbb{R}$.

Lemma 3.1 (Proposition 3 of [64]). *Fix any $t > 0$. The function $\Phi_t(x)$ on $\mathbb{R} \setminus \{0\}$ admits $2q$ positive local extrema counting multiplicities for some integer $q \geq 1$. The preimages of these extrema are denoted by $0 < \zeta_{1,-}(t) \leq \zeta_{1,+}(t) \leq \zeta_{2,-}(t) \leq \zeta_{2,+}(t) \leq \cdots \leq \zeta_{q,-}(t) \leq \zeta_{q,+}(t)$, and they belong to the set $\{\zeta \in \mathbb{R} : 1 - c_N t m_X(\zeta_t) > 0\}$. Moreover, $\lambda_{-,t} = \Phi_t(\zeta_{1,-}(t))$, and $\zeta_{1,-}(t) < \lambda_M(\mathcal{S}(X)) < \zeta_{1,+}(t)$.*

Remark 5. *Here we remark that the model considered in [64] is slightly different in the sense that the model therein contains many 0 eigenvalues, which will force $\zeta_{1,-}(t)$ to be negative. In our case, going through the same analysis as [64] will simply give $0 < \zeta_{1,-}(t)$.*

Next, we shall introduce the deterministic counterpart of $\zeta_{-,t}$ (to be denoted by $\bar{\zeta}_{-,t}$). First, we notice that the MP law holds for both the matrix V_t and X, but with slightly different scaling factors. Specifically, we have $m_{V_t}(z) - \mathsf{m}_{\mathsf{mp}}(z) = \mathfrak{o}_p(1)$ and $m_X(z) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z) = \mathfrak{o}_p(1)$. Recall the definitions of $\zeta_t(z)$ in (26) and $\Phi_t(\zeta)$ in (35). It is important to note that these two quantities are random, and we can also define their deterministic counterparts using the Stieltjes transform of the MP Law. We denote them as follows:

$$\bar{\zeta}_t(z) \coloneqq (1 + c_N t \mathsf{m}_{\mathsf{mp}}(z))^2 z - t(1 - c_N t)(1 - c_N t \mathsf{m}_{\mathsf{mp}}(z)), \tag{39}$$

$$\bar{\Phi}_t(\zeta) \coloneqq (1 - c_N t \mathsf{m}_{\mathsf{mp}}^{(t)}(\zeta))^2 \zeta + (1 - c_N) t (1 - c_N t \mathsf{m}_{\mathsf{mp}}^{(t)}(\zeta)). \tag{40}$$

To further simplify the notation, we let $\zeta_{-,t} = \zeta_t(\lambda_{-,t})$ and $\bar{\zeta}_{-,t} = \bar{\zeta}_t(\lambda_-^{\mathsf{mp}})$. Let $\beta = (\alpha - 2)/24$.

Lemma 3.2. *The following preliminary estimates hold:*

(i) $\zeta_{-,t} - \lambda_M(\mathcal{S}(X)) \leq 0$, and $\lambda_M(\mathcal{S}(X)) - \zeta_{-,t} \sim t^2$ holds on Ω_Ψ.

(ii) There exist some sufficiently small constant $\tau > 0$, such that for any $z \in \mathbb{C}^+$ satisfying $|z - \zeta_{-,t}| \leq \tau t^2$, we have on Ω_Ψ that

$$m_X(z) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z) \prec N^{-\beta}, \quad |m_X^{(k)}(\zeta)| \lesssim t^{-2k+1}, \quad m_X^{(k)}(\zeta_{-,t}) \sim t^{-2k+1}, \quad k \geq 1.$$

(iii) $\bar{\zeta}_{-,t} - \zeta_{-,t} \prec N^{-\beta} t$.

Proof. See the supplementary material [11]. $\square$

We also compute the following limits.

Lemma 3.3. *For any $t = \mathfrak{o}(1)$, we have the following approximations:*

(i) $\mathsf{m}_{\mathsf{mp}}^{(t)}(\bar{\zeta}_{-,t}) = (\sqrt{c_N} - c_N)^{-1} - t c_N^{-1/2}(1 - \sqrt{c_N})^{-2} + \mathcal{O}(t^{3/2})$.

(ii) $t(\mathsf{m}_{\mathsf{mp}}^{(t)}(\bar{\zeta}_{-,t}))' = c_N^{-1}(1 - \sqrt{c_N})^{-2}/2 + \mathcal{O}(t^{1/2})$.

(iii) $t^3(\mathsf{m}_{\mathsf{mp}}^{(t)}(\bar{\zeta}_{-,t}))'' = c_N^{-3/2}(1 - \sqrt{c_N})^{-2}/4 + \mathcal{O}(t^{1/2})$.

(iv) $\gamma_N - c_N^{-1/2}(1 - \sqrt{c_N})^{-4/3} = \mathfrak{o}_p(1)$.

Proof. It is easy to solve $\bar{\zeta}_{-,t} = (1-t)\lambda_-^{\mathsf{mp}} - \sqrt{c_N} t^2$ from (39) and (3). The calculation is then elementary by the explicit formula (3). $\square$

3.2 Proof of Theorem 2.12

Before giving the proof, we need the following pre-process. First, note that we have the following deterministic upper bound when Ψ is good:

$$\zeta_{-,t} \cdot \mathbf{1}_{\Omega_\Psi} \le \lambda_M(\mathcal{S}(X)) \cdot \mathbf{1}_{\Omega_\Psi} \le \lambda_{M-|\mathcal{D}_r|}(\mathcal{S}(\mathsf{B}^{(\mathcal{D}r)})) \cdot \mathbf{1}_{\Omega_\Psi} \le N^{2-2\epsilon_b}.$$

This indicates that $\mathbb{E}(\zeta_{-,t} \cdot \mathbf{1}_{\Omega_\Psi})$ is well-defined. We define

$$\zeta_{\mathsf{e}} := \mathbb{E}(\zeta_{-,t} \cdot \mathbf{1}_{\Omega_\Psi}), \quad \Delta_\zeta := \zeta_{-,t} - \zeta_{\mathsf{e}}. \tag{41}$$

We also write for $z \in \mathbb{C}^+$ and an integer $k \ge 0$, $\Delta_m(z) := m_X(z) - \mathbb{E}m_X(z)$ and $\Delta_m^{(k)}(z) := m_X^{(k)}(z) - \mathbb{E}m_X^{(k)}(z)$, where we remark that $\Delta_m(z) = \Delta_m^{(0)}(z)$. It is noteworthy that $\mathbb{E}m_X(z)$ is well-defined when z possesses a non-zero imaginary part. To ensure that the expectation of $m_X(\zeta_{\mathsf{e}})$ exist, we add a small imaginary part to ζ_{e}, and define for any $K_\zeta > 0$, $\hat{\zeta}_{\mathsf{e}} = \hat{\zeta}_{\mathsf{e}}(K_\zeta) := \zeta_{\mathsf{e}} + \mathrm{i}N^{-100K_\zeta}$.

We will begin by stating some preliminary bounds useful to estimate $\mathbb{E}\lambda_{-,t}$.

Lemma 3.4. *Recall that $\beta = (\alpha - 2)/24$. There exists some small $\tau > 0$, such that for any $z \in \mathbb{C}^+$ satisfies $|z - \zeta_{\mathsf{e}}| \le \tau t^2$ and $\operatorname{Im} z \ge N^{-100K_\zeta}$, the following a priori high probability bounds:*

$$\Delta_m^{(k)}(z) \prec N^{-\beta}t^{-2k}, \quad and \quad \Delta_\zeta \prec N^{-\beta/2}t^2 \tag{42}$$

Furthermore, we have the following a priori variance bounds:

$$\operatorname{Var}(\Delta_m^{(k)}(z)) \le N^{-1+\epsilon}t^{-2k-4}, \quad and \quad \operatorname{Var}(\Delta_\zeta \mathbf{1}_{\Omega_\Psi}) \le N^{-1+\epsilon}. \tag{43}$$

We postpone the proof of Lemma 3.4 to the end of this subsection. Let us prove Theorem 2.12 equipped with Lemma 3.4.

Proof of Theorem 2.12. Recall the expression of $\lambda_{-,t}$ in (13). We shall switch $\zeta_{-,t}$ and $m_X(\zeta_{-,t})$ with ζ_{e} and $\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})$ respectively. First, expanding $m_X(\zeta_{-,t})$ around $m_X(\zeta_{\mathsf{e}})$, we have for sufficiently large $s > 0$,

$$\lambda_{-,t} = \zeta_{-,t}\Big(1 - \sum_{k=0}^s \frac{c_N t}{k!}m_X^{(k)}(\zeta_{\mathsf{e}})\Delta_\zeta^k\Big)^2 + (1 - c_N)t\Big(1 - \sum_{k=0}^s \frac{c_N t}{k!}m_X^{(k)}(\zeta_{\mathsf{e}})\Delta_\zeta^k\Big) + \mathcal{O}_\prec(N^{-\alpha/4-\epsilon}).$$

Note that for any integer $k \ge 0$, it can be easily verified that w.h.p., $|m_X^{(k)}(\zeta_{\mathsf{e}}) - m_X^{(k)}(\hat{\zeta}_{\mathsf{e}})| \le N^{-50s}$, by chooinsg $K_\zeta > 0$ large enough. This means that we can replace $m_X^{(k)}(\zeta_{\mathsf{e}})$ with $m_X^{(k)}(\hat{\zeta}_{\mathsf{e}})$. Through an elementary calculation, we have

$$\lambda_{-,t} = \lambda_{\mathsf{shift}} - \Big(2c_N t\big(1 - c_N t\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})\big)\zeta_{\mathsf{e}} - c_N t^2(1-c_N)\Big)\Delta_m(\hat{\zeta}_{\mathsf{e}}) + \mathsf{ZOT}_\zeta\Delta_\zeta + \mathcal{P}(\Delta_\zeta, \{\Delta_m^{(k)}(\hat{\zeta}_{\mathsf{e}})\}_{k\ge0}).$$

where $\lambda_{\mathsf{shift}} := \big(1 - c_N t\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})\big)^2\zeta_{\mathsf{e}} + (1 - c_N)t\big(1 - c_N t\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})\big)$, and we denote by ZOT_ζ the collection of zero-th order terms, i.e.,

$$\mathsf{ZOT}_\zeta := \big(1 - c_N t\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})\big)\big(1 - c_N t\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}}) - 2c_N t\zeta_{\mathsf{e}}\mathbb{E}m_X'(\hat{\zeta}_{\mathsf{e}})\big) - c_N(1-c_N)t^2\mathbb{E}m_X'(\hat{\zeta}_{\mathsf{e}}), \tag{44}$$

and $\mathcal{P}(\Delta_\zeta, \{\Delta_m^{(k)}(\hat{\zeta}_{\mathsf{e}})\}_{k\ge1})$ collects all the high order terms. We need to bound the last two terms. It can be easily obtained by prior bounds in Lemma 3.4 that $\mathcal{P}(\Delta_\zeta, \{\Delta_m^{(k)}(\hat{\zeta}_{\mathsf{e}})\}_{k\ge0}) = \mathcal{O}_p(N^{-\alpha/4-(4-\alpha)/8})$. Moreover, due to Remark 6 below, we find that $\mathsf{ZOT}_\zeta = \mathcal{O}(N^{-\alpha/4-(4-\alpha)/8})$.

The following two propositions complete the proof.

Proposition 3.5. *Let σ_α be as in Theorem 2.12. We have*

$$2c_N\big(1 - c_N t\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})\big)\zeta_{\mathsf{e}} \cdot \left(\frac{t\Delta_m(\hat{\zeta}_{\mathsf{e}})}{\sigma_\alpha}\right) \Rightarrow \mathcal{N}(0,1).$$

Proposition 3.6. *We have*

$$\lambda_{\mathsf{shift}} = \lambda_-^{\mathsf{mp}} - \frac{\mathsf{c}N^{1-\alpha/2}(1-\sqrt{c_N})^2}{c_N^{(\alpha-2)/4}}\Gamma\left(\frac{\alpha}{2}+1\right) + \mathfrak{o}(N^{1-\alpha/2}).$$

We shall prove the above propositions in the next subsections. $\qquad\qquad\qquad\square$

Proof of Lemma 3.4. Using Lemma 3.2 (ii), we can obtain that $\Delta_m(z) = m_X(z) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z) + \mathbb{E}(\mathsf{m}_{\mathsf{mp}}^{(t)}(z) - m_X(z)) \prec N^{-\beta}$. The bound for $\Delta_m^{(k)}(z)$ follows by a simple application of Cauchy integral formula.

In order to bound Δ_ζ, we first observe that

$$\zeta_{\mathsf{e}} - \bar{\zeta}_{-,t} = \mathbb{E}\big[(\zeta_{-,t} - \bar{\zeta}_{-,t}) \cdot \mathbf{1}_{\Omega_\Psi}\big] - \bar{\zeta}_{-,t} \cdot \mathbb{P}(\Omega_\Psi^c) \leq N^{-\beta/2}t^2. \tag{45}$$

where the last step follows from Lemmas 2.3 and 3.2 (iii). Therefore, by Lemma 3.2 (iii) again, we can get the desired bound for Δ_ζ.

Next we consider $\mathrm{Var}(\Delta_m(z))$. We first let $\mathcal{F}_k$ be the σ-field generated by the first k columns of X. Then we define $D_k^+ := \mathbb{E}\big[M^{-1}(\mathrm{Tr}G(X,z) - \mathrm{Tr}G(X^{(k)},z))\big|\mathcal{F}_k\big]$, $D_k^- := \mathbb{E}\big[M^{-1}(\mathrm{Tr}G(X^{(k)},z) - \mathrm{Tr}G(X,z))\big|\mathcal{F}_{k-1}\big]$, and $D_k := D_k^+ + D_k^-$. By the Efron-Stein inequality, we have

$$\mathrm{Var}(m_X(z)) = \sum_{i=1}^N \mathbb{E}(|D_i|^2) \leq 2\sum_{i=1}^N \mathbb{E}\big(|D_i^+|^2\big) + \mathbb{E}\big(|D_i^-|^2\big).$$

Using the resolvent expansion, we can obtain

$$\mathbb{E}\big(|D_k^+|^2\big) \leq \frac{1}{M^2}\mathbb{E}\Big[\Big|\frac{x_k^\top G^2(X^{(k)},z)x_k}{1+x_k^\top G(X^{(k)},z)x_k}\Big|^2 \cdot \mathbf{1}_{|z-\lambda_M(\mathcal{S}(X^{(k)}))|\geq ct^2}\Big] + N^{-D} \lesssim \frac{N^\epsilon}{N^2t^4},$$

where in the first step, we used Lemma 3.2 (i) to derive, with high probability, that for $|z - \zeta_{\mathsf{e}}| \leq \tau t^2$ with sufficiently small $\tau > 0$, there exists some sufficiently small $c > 0$,

$$|z - \lambda_M(\mathcal{S}(X^{(k)}))| \geq |\zeta_{-,t} - \lambda_M(\mathcal{S}(X))| - |z - \zeta_{\mathsf{e}}| - |\Delta_\zeta|$$
$$- |\lambda_M(\mathcal{S}(X)) - (1-t)\lambda_-^{\mathsf{mp}}| - |\lambda_M(\mathcal{S}(X^{(k)})) - (1-t)\lambda_-^{\mathsf{mp}}| \geq ct^2, \tag{46}$$

which gives $\mathbb{P}(|z - \lambda_M(\mathcal{S}(X^{(k)}))| \geq ct^2) < N^{-D}$ for arbitrary large $D > 0$, and z has non-zero imaginary part which yields deterministic upper bound for the random variable. Similarly, we have $\mathbb{E}\big(|D_k^-|^2\big) \lesssim N^\epsilon/(N^2t^4)$. This establishes the bound for $\mathrm{Var}(\Delta_m(z))$.

The bound for $\mathrm{Var}(\Delta_m^{(k)}(z))$ follows by an application of Cauchy integral formula. Note that, since the contour of the Cauchy integral will cross real line, the integrand may not be well defined deterministically due to the possible singularity (although with tiny probability) of the Green function. Hence, we will need to cut off the part of the integral when the imaginary part of the variable is small. To elucidate the procedure, we will outline how to do the cutoff for the Cauchy integral representation of $\mathbb{E}(m_X^{(k)}(z))$ only. The one for variance can be done similarly. Consider z that satisfies $|z - \zeta_{\mathsf{e}}| \leq \tau t^2/2$ and $\mathrm{Im}\, z \geq N^{-100K_\zeta}$, we first define $\Omega_z := \{|z - \lambda_M(\mathcal{S}(X))| \geq ct^2\}$. A similar argument as (46) leads to $\mathbb{P}(\Omega_z^c) \leq N^{-D}$ for arbitrary large $D > 0$. Then we may choose a contour $\omega_z := \{z' : |z' - z| = \tau t^2/10\}$ with sufficiently small τ, and set $\mathfrak{w} := \{z' : |\mathrm{Im}\, z'| \geq N^{-100K_\zeta}\}$. Then we obtain

$$\mathbb{E}(m_X^{(k)}(z)) = \mathbb{E}(m_X^{(k)}(z) \cdot \mathbf{1}_{\Omega_z}) + \mathbb{E}(m_X^{(k)}(z) \cdot \mathbf{1}_{\Omega_z^c}) = \frac{k!}{2\pi\mathrm{i}}\mathbb{E}\Big[\oint_\omega \frac{m_X(a)}{(a-z)^{k+1}}\mathrm{d}a \cdot \mathbf{1}_{\Omega_z}\Big] + N^{-D}$$

$$= \frac{k!}{2\pi i}\Big(\mathbb{E}\Big[\oint_{\omega\cap\mathfrak{w}}\frac{m_X(a)}{(a-z)^{k+1}}da\cdot\mathbf{1}_{\Omega_z}\Big]+\mathbb{E}\Big[\oint_{\omega\cap\mathfrak{w}^c}\frac{m_X(a)}{(a-z)^{k+1}}da\cdot\mathbf{1}_{\Omega_z}\Big]\Big)+N^{-D}$$

$$= \frac{k!}{2\pi i}\Big(\mathbb{E}\Big[\oint_{\omega\cap\mathfrak{w}}\frac{m_X(a)}{(a-z)^{k+1}}da\Big]+\mathbb{E}\Big[\oint_{\omega\cap\mathfrak{w}^c}\frac{m_X(a)}{(a-z)^{k+1}}da\cdot\mathbf{1}_{\Omega_z}\Big]\Big)+N^{-D}$$

$$= \frac{k!}{2\pi i}\mathbb{E}\Big[\oint_{\omega\cap\mathfrak{w}}\frac{m_X(a)}{(a-z)^{k+1}}da\Big]+\mathcal{O}(N^{-50K_\zeta})+N^{-D}. \tag{47}$$

For the remaining term, the effective imaginary part of a within $\omega\cap\mathfrak{w}$ allows us to interchange $\mathbb{E}$ with the contour integral. Then, the upper bound for $\mathbb{E}(m_X(a))$ can be directly applied to estimate this term. Using the same cutoff of the contours, the bound for $\mathrm{Var}(\Delta_m^{(k)}(z))$ is obtained through a double integral representation together with the Cauchy-Schwarz inequality. We omit further details for brevity.

Lastly, we shall bound $\mathrm{Var}(\Delta_\zeta)$. Since $(\lambda_M(\mathcal{S}(X))-\zeta_{-,t})\cdot\mathbf{1}_{\Omega_\Psi}\sim t^2\cdot\mathbf{1}_{\Omega_\Psi}$ and $\Delta_\zeta\prec N^{-\beta/2}t^2$, on the event Ω_Ψ, $\lambda_M(\mathcal{S}(X))-\zeta_{\mathsf{e}}=\lambda_M(\mathcal{S}(X))-\zeta_{-,t}+\Delta_\zeta\sim t^2$ with high probability. Using Lemma 2.2, the bound in the above display also implies that on the event Ω_Ψ,

$$m_X^{(k)}(\zeta_{\mathsf{e}})\sim t^{-2k+1},\quad k\geq 1. \tag{48}$$

Recall that $\Phi_t'(\zeta_{-,t})=0$, which reads

$$(1-c_N t m_X(\zeta_{-,t}))^2-2c_N t m_X'(\zeta_{-,t})\cdot\zeta_{-,t}\,(1-c_N t m_X(\zeta_{-,t}))-c_N(1-c_N)t^2 m_X'(\zeta_{-,t})=0. \tag{49}$$

Replacing $\zeta_{-,t}$ and $m_X(\zeta_{-,t})$ with ζ_{e} and $\mathbb{E}[m_X(\hat{\zeta}_{\mathsf{e}})]$, as in the proof of Theorem 2.12, it follows from (49) that

$$\mathsf{ZOT}_\zeta+\mathsf{FOT}_\zeta+\mathcal{P}_\zeta(\Delta_\zeta,\{\Delta_m^{(k)}\}_{k\geq 0})=0, \tag{50}$$

where the term ZOT_ζ is defined as in (44),

$$\mathsf{FOT}_\zeta:=\big(2c_N^2 t^2\zeta_{\mathsf{e}}\mathbb{E}m_X'(\hat{\zeta}_{\mathsf{e}})-2\mathsf{f}_m\big)\Delta_m(\hat{\zeta}_{\mathsf{e}})-\big(c_N(1-c_N)t^2+2\mathsf{f}_m\zeta_{\mathsf{e}}\big)\Delta_m^{(1)}(\hat{\zeta}_{\mathsf{e}})$$
$$-\big(4\mathsf{f}_m\mathbb{E}m_X'(\hat{\zeta}_{\mathsf{e}})+c_N(1-c_N)t^2\mathbb{E}m_X^{(2)}(\hat{\zeta}_{\mathsf{e}})+2c_N^2 t^2\zeta_{\mathsf{e}}(\mathbb{E}m_X'(\hat{\zeta}_{\mathsf{e}}))^2+2\mathsf{f}_m\zeta_{\mathsf{e}}\mathbb{E}m_X^{(2)}(\hat{\zeta}_{\mathsf{e}})\big)\Delta_\zeta$$

with $\mathsf{f}_m:=c_N t\big(1-c_N t\mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})\big)$, and $\mathcal{P}_\zeta(\Delta_\zeta,\Delta_m^{(k)})$ is the collection of high order terms. Note that $\mathsf{f}_m\sim t$ and $\mathcal{P}_\zeta(\Delta_\zeta,\Delta_m^{(k)})$ is a polynomial in Δ_ζ and $\Delta_m^{(k)}$'s, containing monomials of order no smaller than 2.

Hence, by Cauchy Schwarz and bounds in (42), one can get the following bounds

$$\mathrm{Var}\Big(\mathcal{P}(\Delta_\zeta,\Delta_m^{(k)}(\hat{\zeta}_{\mathsf{e}}))\mathbf{1}_{\Omega_\Psi}\Big)\lesssim N^{-1/2-\beta/4}\sqrt{\mathrm{Var}(\Delta_\zeta\mathbf{1}_{\Omega_\Psi})}+N^{-\beta/4}\mathrm{Var}(\Delta_\zeta\mathbf{1}_{\Omega_\Psi})+N^{-D}, \tag{51}$$

$$\mathbb{E}\Big(\mathcal{P}(\Delta_\zeta,\Delta_m^{(k)}(\hat{\zeta}_{\mathsf{e}}))\mathbf{1}_{\Omega_\Psi}\Big)\lesssim N^{-1/2+\epsilon/2}t^{-3}\sqrt{\mathrm{Var}(\Delta_\zeta\mathbf{1}_{\Omega_\Psi})}+t^{-2}\mathrm{Var}(\Delta_\zeta\mathbf{1}_{\Omega_\Psi})+N^{-D}. \tag{52}$$

Using (48), we can see that the leading order term of the coefficient of Δ_ζ in FOT_ζ is $-2\mathsf{f}_m\zeta_{\mathsf{e}}\mathbb{E}(m_X^{(2)}(\hat{\zeta}_{\mathsf{e}}))\sim t^{-2}$. Therefore, we can derive from (50) that

$$C_1(t)\Delta_\zeta=C_2(t)\Delta_m(\hat{\zeta}_{\mathsf{e}})+C_3(t)\Delta_m^{(1)}(\hat{\zeta}_{\mathsf{e}})+\frac{\mathsf{ZOT}_\zeta+\mathcal{P}_\zeta(\Delta_\zeta,\Delta_m^{(k)}(\hat{\zeta}_{\mathsf{e}}))}{2\mathsf{f}_m\zeta_{\mathsf{e}}\mathbb{E}m_X^{(2)}(\hat{\zeta}_{\mathsf{e}})}, \tag{53}$$

where $C_i(t),i=1,2,3$ are deterministic quantities satisfying $C_1(t)=1+\mathcal{O}(t)$, $C_2(t)=\mathcal{O}(t^3)$, and $C_3(t)=\mathcal{O}(t^3)$. Multiplying $\mathbf{1}_{\Omega_\Psi}$ at both sides and then compute the variance:

$$\mathrm{Var}(\Delta_\zeta\mathbf{1}_{\Omega_\Psi})\lesssim t^6\mathrm{Var}(\Delta_m(\hat{\zeta}_{\mathsf{e}}))+t^6\mathrm{Var}(\Delta_m^{(1)}(\hat{\zeta}_{\mathsf{e}}))+t^4\mathrm{Var}\big(\mathcal{P}_\zeta(\Delta_\zeta,\Delta_m^{(k)}(\hat{\zeta}_{\mathsf{e}}))\big)$$
$$\lesssim N^{-1+\epsilon}+N^{-1/2-\beta/4}\sqrt{\mathrm{Var}(\Delta_\zeta\mathbf{1}_{\Omega_\Psi})}, \tag{54}$$

Solving the above inequality for $\mathrm{Var}(\Delta_\zeta \mathbf{1}_{\Omega_\Psi})$ gives $\mathrm{Var}(\Delta_\zeta \mathbf{1}_{\Omega_\Psi}) \lesssim N^{-1+\epsilon}$, which completes the proof of Lemma 3.4. $\qquad\square$

Remark 6 (Bound ZOT_ζ). *We start with (53). Multiplying $\mathbf{1}_{\Omega_\Psi}$ at both sides and then taking expectation, we have $(\mathsf{ZOT}_\zeta + \mathbb{E}[\mathcal{P}_\zeta(\Delta_\zeta, \Delta_m^{(k)}(\hat{\zeta}_\mathsf{e})) \cdot \mathbf{1}_{\Omega_\Psi}])/(2\mathsf{f}_m \zeta_\mathsf{e} \mathbb{E} m_X^{(2)}(\hat{\zeta}_\mathsf{e})) + \mathcal{O}(N^{-D}) = 0$. Using (52) together with the variance bound for $\Delta_\zeta \mathbf{1}_{\Omega_\Psi}$ in Lemma 3.4, we can obtain that $\mathsf{ZOT}_\zeta = \mathcal{O}(N^{-1+\epsilon} t^{-3})$. By the fact $t \gg N^{(\alpha-4)/32}$, it follows that $\mathsf{ZOT}_\zeta = \mathcal{O}(N^{-\alpha/4 - (4-\alpha)/8})$.*

3.3 Proof of Proposition 3.5

Proposition 3.5 follows from Lemma 3.3 and the following theorem together with some simple algebraic calculation. Recall that $\Delta_m(\hat{\zeta}_\mathsf{e}) = m_X(\hat{\zeta}_\mathsf{e}) - \mathbb{E}(m_X(\hat{\zeta}_\mathsf{e}))$.

Theorem 3.7 (CLT of the linear eigenvalue statistics of $\mathcal{S}(X)$). *For any $2 < \alpha < 4$,*

$$\frac{N^{\alpha/4} t \Delta_m(\hat{\zeta}_\mathsf{e})}{\sigma_m} \Rightarrow \mathcal{N}(0,1),$$

where

$$\sigma_m^2 := \mathsf{c} t^2 c_N \int_0^\infty \int_0^\infty \partial_z \partial_{z'} \Big\{ \frac{e^{-s-s'-sc_N \mathsf{m}_{\mathsf{mp}}^{(t)}(z) - s' c_N \mathsf{m}_{\mathsf{mp}}^{(t)}(z')}}{ss'}$$
$$\times \Big(\big(s\mathsf{m}_{\mathsf{mp}}^{(t)}(z) + s'\mathsf{m}_{\mathsf{mp}}^{(t)}(z')\big)^{\alpha/2} - \big(s\mathsf{m}_{\mathsf{mp}}^{(t)}(z)\big)^{\alpha/2} - \big(s'\mathsf{m}_{\mathsf{mp}}^{(t)}(z')\big)^{\alpha/2} \Big) \Big\} \Big|_{z=z'=\hat{\zeta}_\mathsf{e}} \mathrm{d}s \mathrm{d}s'.$$

To prove Theorem 3.7, we will work on the truncated matrix $\tilde{X} = (\tilde{x}_{ij})$ with $\tilde{x}_{ij} = x_{ij} \mathbf{1}_{\sqrt{N}|x_{ij}| \le N^\vartheta}$ and $\vartheta = 1/4 + 1/\alpha + \epsilon_\vartheta$ such that $N^{-\alpha\epsilon_\vartheta} \ll t$ and $\epsilon_\vartheta < (3\alpha-5)/(4\alpha)$. It will become clear from the following lemma that the fluctuations of m_X and $m_{\tilde{X}}$ are asymptotically the same.

Lemma 3.8. *We have $N^{\alpha/4} t \big(m_X(\hat{\zeta}_\mathsf{e}) - m_{\tilde{X}}(\hat{\zeta}_\mathsf{e})\big) = \mathfrak{o}_p(1)$.*

Proof. This lemma simply follows from the rank inequality and Bennett's inequality together with Lemma 3.2 (i). $\qquad\square$

Proof of Theorem 3.7. By Lemma 3.8, it is enough to consider the convergence (in distribution) of $\mathcal{M}_N(\tilde{X}) := N^{\alpha/4} t \big(m_{\tilde{X}}(\hat{\zeta}_\mathsf{e}) - \mathbb{E} m_{\tilde{X}}(\hat{\zeta}_\mathsf{e})\big)$. We will use the Martingale approach. To this end, we define $\mathcal{F}_k$ as the sigma-algebra generated by the first k columns of $\tilde{X}$. Denoting conditional expectation w.r.t. $\mathcal{F}_k$ by $\mathbb{E}_k$, we obtain the following martingale difference decomposition of $\mathcal{M}_N(\tilde{X})$

$$\mathcal{M}_N(\tilde{X}) = \sum_{k=1}^N N^{\alpha/4} t (\mathbb{E}_k - \mathbb{E}_{k-1}) \big(m_{\tilde{X}}(\hat{\zeta}_\mathsf{e}) - m_{\tilde{X}^{(k)}}(\hat{\zeta}_\mathsf{e})\big).$$

Our aim is to show that $\mathcal{M}_N(\tilde{X})$ converges in distribution to a Gaussian distribution $\mathcal{N}(0, \sigma_m^2)$ via the martingale CLT.

Theorem 3.9 (Martingale CLT, Theorem A.3 of [15]). *Let $(\mathcal{F}_k)_{k \ge 0}$ be a filtration such that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and let $(\mathcal{W}_k)_{k \ge 0}$ be a square-integrable complex-valued martingale starting at zero w.r.t. this filtration. For $k \ge 1$, we define the random variables $Y_k := \mathcal{W}_k - \mathcal{W}_{k-1}$, $v_k := \mathbb{E}_k[|Y_k|^2]$, $\tau_k := \mathbb{E}_k[Y_k^2]$, and we also define $v(N) := \sum_{k \ge 1} v_k$, $\tau(N) := \sum_{k \ge 1} \tau_k$, $\sum_{k \ge 1} \mathbb{E}[|Y_k^2| \mathbf{1}_{|Y_k| \ge \varepsilon}]$. Suppose that for some constants $v \ge 0$, $\tau \in \mathbb{C}$, and for each $\varepsilon > 0$, $v(N) \xrightarrow{\mathbb{P}} v$, $\tau(N) \xrightarrow{\mathbb{P}} \tau$, $L(\varepsilon, N) \to 0$. Then, the martingale $\mathcal{W}_N$ converges in distribution to a centered complex Gaussian variable $\mathcal{Z}$ such that $\mathbb{E}(|\mathcal{Z}|^2) = v$ and $\mathbb{E}(\mathcal{Z}^2) = \tau$ as $N \to \infty$.*

We want to apply Theorem 3.9 with setting $\mathcal{W}_N = \mathcal{M}_N(\tilde{X})$. Using the resolvent identity,

$$\mathcal{M}_N(\tilde{X}) = \sum_{k=1}^N Y_k(\hat{\zeta}_\mathsf{e}) := \sum_{k=1}^N \frac{t}{N^{1-\alpha/4}} (\mathbb{E}_k - \mathbb{E}_{k-1}) \frac{\tilde{x}_k^\top (G(\tilde{X}^{(k)}, \hat{\zeta}_\mathsf{e}))^2 \tilde{x}_k}{1 + \tilde{x}_k^\top G(\tilde{X}^{(k)}, \hat{\zeta}_\mathsf{e}) \tilde{x}_k}.$$

First note that $|Y_k(\hat{\zeta}_{\mathsf{e}})| \lesssim N^{-1+\alpha/4}t^{-1}$ with high probability. We also have the deterministic upper bound for $Y_k(\hat{\zeta}_{\mathsf{e}})$ since $\hat{\zeta}_{\mathsf{e}}$ possesses effective imaginary part. Combining these two facts, we can verify that the $L(\varepsilon, N)$ goes to 0.

In order to conclude the proof via Theorem 3.9, we need to check convergences of $v(N)$ and $\tau(N)$. This follows from Propositions 3.10 and 3.11 below. $\qquad\square$

Proposition 3.10. *Let*

$$\tilde{Y}_k(\zeta) := \frac{t}{N^{1-\alpha/4}}(\mathbb{E}_k - \mathbb{E}_{k-1})f_k(\zeta) := \frac{t}{N^{1-\alpha/4}}(\mathbb{E}_k - \mathbb{E}_{k-1})\frac{\tilde{x}_k^\top(G(\tilde{X}^{(k)},\zeta))_{\mathrm{diag}}^2\tilde{x}_k}{1 + \tilde{x}_k^\top(G(\tilde{X}^{(k)},\zeta))_{\mathrm{diag}}\tilde{x}_k}.$$

Then there exists some constant τ, such that for any $\zeta, \zeta' \in \Xi(\tau) = \{\xi \in \mathbb{C} : |\xi - \hat{\zeta}_{\mathsf{e}}| \leq \tau t^2, |\mathrm{Im}\,\xi| \geq N^{-100}\}$, the summation $\sum_{k=1}^N \mathbb{E}_{k-1}[Y_k(\zeta)Y_k(\zeta')] - \mathbb{E}_{k-1}[\tilde{Y}_k(\zeta)\tilde{Y}_k(\zeta')]$ converges in probability to 0.

Proof. The proof is similar to the counterpart in [15]; see the supplementary material [11] for details. $\qquad\square$

Proposition 3.11. *For any $k \in [N]$, there exists some constant τ, such that for any $z, z' \in \{\zeta \in \mathbb{C} : |\zeta - \hat{\zeta}_{\mathsf{e}}| \leq \tau t^2, |\mathrm{Im}\,\zeta| \geq N^{-100}\}$,*

$$\frac{N^{-1+\alpha/2}t^2\mathbb{E}_{k-1}\big((\mathbb{E}_k - \mathbb{E}_{k-1})f_k(z)(\mathbb{E}_k - \mathbb{E}_{k-1})f_k(z')\big)}{\mathcal{K}(z,z')} \xrightarrow{\mathbb{P}} 1,$$

as $N \to \infty$. The kernel $\mathcal{K}(z,z')$ is defined as

$$\mathcal{K}(z,z') := \mathsf{c}N^{1-\alpha/2}t^2 c_N \int_0^\infty \int_0^\infty \partial_z\partial_{z'}\bigg\{\frac{e^{-s-s'-sc_N\mathsf{m}_{\mathsf{mp}}^{(t)}(z)-s'c_N\mathsf{m}_{\mathsf{mp}}^{(t)}(z')}}{ss'}$$
$$\times \left(\big(s\mathsf{m}_{\mathsf{mp}}^{(t)}(z) + s'\mathsf{m}_{\mathsf{mp}}^{(t)}(z')\big)^{\alpha/2} - \big(s\mathsf{m}_{\mathsf{mp}}^{(t)}(z)\big)^{\alpha/2} - \big(s'\mathsf{m}_{\mathsf{mp}}^{(t)}(z')\big)^{\alpha/2}\right)\bigg\}\mathrm{d}s\mathrm{d}s'.$$

Before giving the proof of Proposition 3.11, let us introduce the parameter $\sigma_N := \sqrt{N\mathbb{E}\tilde{x}_{ij}^2}$ and Lemma 3.12 below. Note that

$$\mathbb{E}\big(N\tilde{x}_{ij}^2\mathbf{1}_{\sqrt{N}x_{ij}>N^\vartheta}\big) = \int_{N^{2\vartheta}}^\infty \mathbb{P}(|\sqrt{N}x_{ij}|^2 > x)\mathrm{d}x \sim N^{\vartheta(2-\alpha)}, \tag{55}$$

which gives $\sigma_N^2 - (1-t) = \mathcal{O}(N^{\vartheta(2-\alpha)})$. The following lemma collects some useful properties of $\tilde{x}_{ij}$ and the expansion for the characteristic function of x_{ij}.

Lemma 3.12. *Then there exists constant $C > 0$, such that*

(i) $\tilde{x}_{ij}$'s are i.i.d. centered, with variance σ_N^2/N, third moment bound $N^{3/2}\mathbb{E}[|\tilde{x}_{ij}|^3] \leq CN^{\vartheta(3-\alpha)_+}$, and fourth moment bound $N^2\mathbb{E}[|\tilde{x}_{ij}|^4] \leq CN^{\vartheta(4-\alpha)}$,

(ii) for any $\lambda \in \mathbb{C}$ such that $\mathrm{Im}\,\lambda \leq 0$,

$$\phi_N(\lambda) := \mathbb{E}\big(e^{-\mathrm{i}\lambda|x_{ij}|^2}\big) = 1 - \frac{\mathrm{i}(1-t)\lambda}{N} + c\frac{(\mathrm{i}\lambda)^{\frac{\alpha}{2}}}{N^{\frac{\alpha}{2}}} + \varepsilon_N(\lambda), \quad and \quad \varepsilon_N(\lambda) = \mathcal{O}\Big(\frac{|\lambda|^{(\alpha+\varrho)/2}}{N^{(\alpha+\varrho)/2}} \vee \frac{|\lambda|^2}{N^2}\Big).$$

Proof. The proof of (i) is elementary. To prove (ii), we observe

$$1 - \phi_N(\lambda) = \int_0^\infty \big(\exp(-\mathrm{i}\lambda u/N) - 1\big)\mathrm{d}F^c(u) = \frac{\mathrm{i}\lambda}{N}\int_0^\infty \exp(-\mathrm{i}\lambda u/N)F^c(u)\mathrm{d}u,$$

where F be the distribution function of Nx_{ij}^2 and let $F^c = 1 - F$. Since $\int_0^\infty F^c(u)\mathrm{d}u = 1 - t$, we notice

$$1 - \phi_N(\lambda) = \frac{\mathrm{i}\lambda(1-t)}{N} + \frac{\mathrm{i}\lambda}{N}\int_0^\infty \big(\exp(-\mathrm{i}\lambda u/N) - 1\big)F^c(u)\mathrm{d}u.$$

The estimate (ii) can be obtained using the tail density assumption on $\sqrt{N}y_{ij}$ (cf. Assumption 1 (i)). $\qquad\square$

Proof of Proposition 3.11. Let $\hat{f}_k$ be defined as f_k, but with the matrix $\tilde{X}$ replaced by a matrix $\hat{X}$. The columns $\hat{X}_i$ of $\hat{X}$ are the same as those of $\tilde{X}$ if $i \leq k$, but are independent random vectors with the same distribution as the columns of $\tilde{X}$ if $i > k$. It is still valid to use the notation $\mathbb{E}_k$ since $\tilde{X}$ and $\hat{X}$ share the same first k columns. By the following elementary identity

$$\mathbb{E}_{k-1}\big((\mathbb{E}_k - \mathbb{E}_{k-1})(f_k(z))(\mathbb{E}_k - \mathbb{E}_{k-1})(f_k(z'))\big)$$
$$= \mathbb{E}_k\big(\mathbb{E}_{\tilde{x}_k}(f_k(z)\hat{f}_k(z'))\big) - \big(\mathbb{E}_k\mathbb{E}_{\tilde{x}_k}f_k(z)\big)\big(\mathbb{E}_k\mathbb{E}_{\tilde{x}_k}\hat{f}_k(z')\big), \tag{56}$$

it suffices to study the approximation for $\mathbb{E}_{\tilde{x}_k}f_k(z)$ and $\mathbb{E}_{\tilde{x}_k}f_k(z)\hat{f}_k(z')$. In the sequel, we write $f_k = f_k(z)$, $\hat{f}'_k = \hat{f}_k(z')$, $G_k = G(\tilde{X}^{(k)}, z)$ and $G'_k = G(\tilde{X}^{(k)}, z')$ for simplicity. By a minor process argument, for any $D > 0$, there exists constant $C_k > 0$ such that $|\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - \hat{\zeta}_{\mathsf{e}}| \geq C_k t^2$, with probability at least $1 - N^{-D}$. This implies that there exists some constant $C_k > 0$ such that for any arbitrary large $D > 0$,

$$\mathbb{P}(\tilde{\Omega}_k = \{\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - \bar{\zeta}_{-,t} \geq C_k t^2\}) \geq 1 - N^{-D}.$$

Then it is readily seen that $\operatorname{Re}[G(\tilde{X}^{(k)}, z)]_{jj} \cdot \mathbf{1}_{\tilde{\Omega}_k} \geq 0$ for any $z \in \{|z - \bar{\zeta}_{-,t}| \leq C_k t^2/10, |\operatorname{Im} z| \geq N^{-100}\}$. Since $\tilde{\Omega}_k$ is independent of $\tilde{x}_k$, we can write $\mathbb{E}_{\tilde{x}_k}(f_k) = \mathbb{E}_{\tilde{x}_k}(f_k)\mathbf{1}_{\tilde{\Omega}_k} + \mathbb{E}_{\tilde{x}_k}(f_k)\mathbf{1}_{\tilde{\Omega}_k^c}$. Using the facts that $|\tilde{x}_{jk}| \leq N^{1/\alpha+1/4+\epsilon_\vartheta}$ and $|[G_k]_{jj}| \leq |\operatorname{Im} z|^{-1} \leq N^{101}$, we have for some large constant $K > 0$ such that $|\mathbb{E}_{\tilde{x}_k}(f_k)\mathbf{1}_{\tilde{\Omega}_k^c}| \leq N^K \mathbf{1}_{\tilde{\Omega}_k^c}$.

Next, we will mainly focus on the estimation for $\mathbb{E}_{\tilde{x}_k}(f_k)\mathbf{1}_{\tilde{\Omega}_k}$. In the sequel, we omit the indicate function $\mathbf{1}_{\tilde{\Omega}_k}$ from the display for simplicity, and keep in mind that all the estimates are done on the event $\tilde{\Omega}_k$. Using the identity that for w with $\operatorname{Re} w > 0$, $w^{-1} = \int_0^\infty e^{-sw}\mathrm{d}s$, we have

$$\mathbb{E}_{\tilde{x}_k}f_k = \mathbb{E}_{\tilde{x}_k}\Big(\int_0^\infty \sum_j \tilde{x}_{jk}^2[G_k^2]_{jj} e^{-s\left(1+\sum_j \tilde{x}_{jk}^2[G_k]_{jj}\right)}\mathrm{d}s\Big) = -\int_0^\infty \frac{e^{-s}}{s}\partial_z\Big\{\mathbb{E}_{\tilde{x}_k}\Big(e^{-s\sum_j \tilde{x}_{jk}^2[G_k]_{jj}}\Big)\Big\}\mathrm{d}s.$$

Recall $\tilde{\phi}_N$ and ϕ_N in Lemma 3.12. We have

$$\mathbb{E}_{\tilde{x}_k}f_k = -\int_0^\infty \frac{e^{-s}}{s}\partial_z\Big\{\prod_j \phi_N\big(-\mathrm{i}s[G_k]_{jj}\big)\Big\}\mathrm{d}s + \mathsf{Diff},$$

where $\mathsf{Diff} := \int_0^\infty \frac{e^{-s}}{s}\partial_z\Big\{\prod_j \phi_N\big(-\mathrm{i}s[G_k]_{jj}\big) - \prod_j \tilde{\phi}_N\big(-\mathrm{i}s[G_k]_{jj}\big)\Big\}\mathrm{d}s$. Note by the definition of $\tilde{x}_{jk}$'s for any $j \in [N]$, the following estimate holds uniformly for all λ with $\operatorname{Im}\lambda \leq 0$,

$$\big|\phi_N(\lambda) - \tilde{\phi}_N(\lambda)\big| = \Big|\mathbb{E}\Big[(e^{-\mathrm{i}\lambda|x_{ij}|^2} - 1)\cdot \mathbf{1}_{\sqrt{N}|x_{ij}|>N^\vartheta}\Big]\Big| \leq 2\mathbb{P}(\sqrt{N}|x_{ij}| > N^\vartheta) \lesssim N^{-\alpha\vartheta}.$$

Therefore, by a Cauchy integral argument with contour radius equals to ct^2 for some sufficiently small $c > 0$, we have for sufficiently large K,

$$\Big|\int_{N^{-K}}^\infty \frac{e^{-s}}{s}\partial_z\Big\{\prod_j \phi_N\big(-\mathrm{i}s[G_k]_{jj}\big) - \prod_j \tilde{\phi}_N\big(-\mathrm{i}s[G_k]_{jj}\big)\Big\}\mathrm{d}s\Big| \lesssim t^{-2}N^{-\alpha\vartheta}\int_{N^{-K}}^\infty \frac{e^{-s}}{s}\mathrm{d}s \lesssim N^{1-\alpha/2-\epsilon}.$$

With the prescribe K, we also have

$$\Big|\int_0^{N^{-K}} \frac{e^{-s}}{s}\partial_z\Big\{\prod_j \tilde{\phi}_N\big(-\mathrm{i}s[G_k]_{jj}\big)\Big\}\mathrm{d}s\Big| = \Big|\mathbb{E}_{\tilde{x}_k}\Big(\int_0^{N^{-K}} \sum_j \tilde{x}_{jk}^2[G_k^2]_{jj} e^{-s\left(1+\sum_j \tilde{x}_{jk}^2[G_k]_{jj}\right)}\mathrm{d}s\Big)\Big| \lesssim N^{-K/2},$$

and similar estimate holds if we replace $\tilde{\phi}_N$ by ϕ_N. Combining the above two displays, we can obtain that

$$\mathbb{E}_{\tilde{x}_k}f_k = -\int_0^\infty \frac{e^{-s}}{s}\partial_z\Big\{\prod_j \Big(1 + \frac{1-t}{N}u_j(z,s)\Big)\Big\}\mathrm{d}s + \mathcal{O}_\prec(N^{1-\alpha/2-\epsilon}),$$

20

where

$$u_j(z,s) = \frac{N}{1-t}\Big(\phi\big(-\mathrm{i}s[G_k]_{jj}\big)-1\Big) = -s[G_k]_{jj} + \mathsf{c}\frac{(s[G_k]_{jj})^{\frac{\alpha}{2}}}{N^{\frac{\alpha-2}{2}}(1-t)} + \frac{N}{1-t}\varepsilon_N(-\mathrm{i}s[G_k]_{jj}).$$

We introduce the approximation $\mathsf{K}_1(z,s)$ for the integrand as follows:

$$\mathsf{K}_1(z,s) := \frac{e^{-s-s(1-t)\mathrm{Tr}G_k/N}}{s}\Big(1 + \frac{\mathsf{c}}{N^{\alpha/2}}\sum_{j=1}^{M}(s[G_k]_{jj})^{\alpha/2}\Big).$$

Then our goal is to show on the event $\tilde{\Omega}_k$

$$\int_0^\infty \partial_z \delta(z,s)\mathrm{d}s \lesssim N^{1-\alpha/2-\epsilon}, \tag{57}$$

where $\delta(z,s) = \frac{e^{-s}}{s}\prod_j\big(1+\frac{1-t}{N}u_j(z,s)\big) - \mathsf{K}(z,s)$, and $\epsilon > 0$ is a small constant. By the Cauchy integral formula, we have $\left|\int_0^\infty \partial_z\delta(z,s)\mathrm{d}s\right| \lesssim t^{-2}\int_0^\infty |\delta(z_s,s)|\mathrm{d}s$, where z_s is the maximizer of $|\delta(z,s)|$ on the contour $\{z' : |z'-z| = C_k t^2/50\}$. To estimate the RHS of this inequality, we divide it into two parts,

$$\frac{1}{t^2}\int_0^\infty |\delta(z_s,s)|\mathrm{d}s = \frac{1}{t^2}\int_0^{N^\varsigma}|\delta(z_s,s)|\mathrm{d}s + \frac{1}{t^2}\int_{N^\varsigma}^\infty |\delta(z_s,s)|\mathrm{d}s = I_1 + I_2,$$

with ς being chosen later . Using the fact that $[G_k]_{jj} \lesssim t^{-2}$ on the event $\tilde{\Omega}_k$, we can obtain that

$$I_2 \lesssim \frac{1}{t^{2+\alpha}}\int_{N^\varsigma}^\infty s^{\alpha/2-1}e^{-s}\mathrm{d}s \leq e^{-N^\varsigma/3}.$$

For I_1, we further decompose it into three parts,

$$I_1 = \frac{1}{t^2}\int_0^{N^\varsigma}\Big|\frac{e^{-s}}{s}\Big(\prod_j\big(1+\frac{1-t}{N}u_j(z_s,s)\big) - e^{\sigma_N^2/N\sum_j u_j(z_s,s)}\Big)\Big|\mathrm{d}s$$

$$+ \frac{1}{t^2}\int_0^{N^\varsigma}\Big|\frac{e^{-s}}{s}\Big(e^{(1-t)/N\sum_j u_j(z_s,s)} - e^{-s(1-t)\mathrm{Tr}G_k/N}\big(1+\sum_j\mathsf{c}\frac{(s[G_k]_{jj})^{\frac{\alpha}{2}}}{N^{\frac{\alpha}{2}}} + \sum_j\varepsilon_N(-\mathrm{i}s[G_k]_{jj})\big)\Big)\Big|\mathrm{d}s$$

$$+ \frac{1}{t^2}\int_0^{N^\varsigma}\Big|\frac{e^{-s}}{s}\Big(e^{-s(1-t)\mathrm{Tr}G_k/N}\sum_j\varepsilon_N(-\mathrm{i}s[G_k]_{jj})\big)\Big|\mathrm{d}s = I_{11} + I_{12} + I_{13}.$$

Notice that on the event $\tilde{\Omega}_k$, $M_s = \max_j |u_j(z_s,s)|\sigma_N^2 \lesssim st^{-2}$, and $\mathrm{Re}\,u_j(z_s,s) = N(\mathrm{Re}\,\phi_N(-\mathrm{i}s[G_k]_{jj}) - 1) \leq 0$. Then using [15, Lemma 4.5], we have on the event $\tilde{\Omega}_k$,

$$I_{11} \leq \frac{1}{t^2}\int_0^{N^\varsigma}\frac{e^{-s}}{s}\cdot\frac{s^2}{Nt^4}e^{s^2/(Nt^4)+\sum_j\mathrm{Re}\,((1-t)u_j(z_s,s))/N}\mathrm{d}s \lesssim \frac{1}{Nt^6}\int_0^{N^\varsigma}e^{-s}se^{s^2/(Nt^4)}\mathrm{d}s \lesssim \frac{1}{Nt^6}.$$

By choosing $\varsigma < 1/3$ (say), we can obtain that $I_{11} \lesssim N^{-1}t^{-6}$. Applying the simple inequality that $|e^x - (1+x)| \leq 2|x|^2$ for $|x| \leq 1/2$, we have

$$I_{12} \lesssim \frac{1}{t^2}\int_0^{N^\varsigma}\frac{e^{-s}}{s}\Big|\sum_j\mathsf{c}\frac{(s[G_k]_{jj})^{\frac{\alpha}{2}}}{N^{\frac{\alpha}{2}}} + \sum_j\varepsilon_N(-\mathrm{i}s[G_k]_{jj})\Big|^2\mathrm{d}s$$

$$\lesssim \frac{1}{N^{\alpha-2}t^{2+2\alpha}}\int_0^{N^\varsigma}s^{\alpha-1}\mathrm{d}s \lesssim N^{\varsigma\alpha-\alpha+2}t^{-2-2\alpha} \lesssim N^{-3(\alpha-2)/5},$$

where in the last step, we chose $\varsigma < (\alpha - 2)/(4\alpha)$. Finally, for I_{13}, we can use Lemma 3.12 (ii) to obtain that, $I_{13} \lesssim N^{-(\alpha-2)\vartheta} t^{-2-2\alpha} \lesssim N^{-3(\alpha-2)/4}$. Now we may conclude the proof of (57) by combining the above estimates and possibly adjusting the constants. This gives

$$\mathbb{E}_{\tilde{x}_k}(f_k) = -\int_0^\infty \partial_z \mathsf{K}_1(z,s) \mathrm{d}s + O_\prec(N^{1-\alpha/2-\epsilon}).$$

Similarly, we can obtain that

$$\mathbb{E}_{\tilde{x}_k}(f_k \hat{f}'_k) = \int_0^\infty \int_0^\infty \partial_z \partial_{z'} \mathsf{K}_2(z,z',s,s') \mathrm{d}s \mathrm{d}s' + O_\prec(N^{1-\alpha/2-\epsilon}),$$

where

$$\mathsf{K}_2(z,z',s,s') := \frac{e^{-s-s'-s(1-t)\mathrm{Tr}G_k/N-s(1-t)\mathrm{Tr}G'_k/N}}{ss'} \left(1 + \frac{\mathsf{c}}{N^{\alpha/2}} \sum_{j=1}^M \left(s[G_k]_{jj} + s'[G'_k]_{jj}\right)^{\alpha/2}\right).$$

Notice that the estimate in Proposition 2.10 can be obtained for our $G(\tilde{X}^{(k)},z)$ as well in the same manner. Suppose that $\max\{|\mathcal{D}_r|,|\mathcal{D}_c|\} \leq N^{1-\epsilon_d}$ for some ϵ_d. Notice that $\epsilon_d \geq \epsilon_\alpha$ by definition. Hence, the claim now follows by (i) employing (56), then substituting $\sigma_N^2[G_k]_{jj}(z)$ with $\mathsf{m}_{\mathsf{mp}}(z/\sigma_N^2)$ for $j \in \mathcal{T}_r$, and utilizing the bound $[G_k]_{jj} \prec 1/t$ for $j \in \mathcal{D}_r$ with the fact $t \gg N^{-\epsilon_d/4}$; (ii) considering the estimates $\sigma_N^2 - (1-t) = \mathcal{O}(N^{\vartheta(2-\alpha)})$ (cf. (55)), $\partial_z \mathsf{m}_{\mathsf{mp}}(z/\sigma_N^2) \sim t^{-1}$, and $\partial_z^2 \mathsf{m}_{\mathsf{mp}}(z/\sigma_N^2) \sim t^{-3}$ for z within the specified domain. This enables us to further replace $\mathsf{m}_{\mathsf{mp}}(z/\sigma_N^2)$ and $\mathsf{m}_{\mathsf{mp}}(z/\sigma_N^2)$ with $\mathsf{m}_{\mathsf{mp}}^{(t)}(z)$ and $\mathsf{m}_{\mathsf{mp}}^{(t)}(z')$, respectively. $\qquad\square$

3.4 Proof of Proposition 3.6

Let us first define

$$\mathfrak{p}(z) := \mathsf{c}N^{1-\alpha/2}c_N \int_0^\infty e^{-s-sc_N \mathsf{m}_{\mathsf{mp}}^{(t)}(z)} \left(s\mathsf{m}_{\mathsf{mp}}^{(t)}(z)\right)^{\alpha/2} \mathrm{d}s,$$

$$m_{\mathsf{shift}}(z) := \frac{\mathrm{i}(\frac{z}{1-t} - c_N + 1)\mathfrak{p}(z)}{2c_N z \sqrt{(\frac{z}{1-t} - \lambda_-^{\mathsf{mp}})(\lambda_+^{\mathsf{mp}} - \frac{z}{1-t})}}.$$

Then we have the following proposition concerning the expansion of $\mathbb{E}m_X(z)$.

Proposition 3.13. *There exists some sufficiently small constant $\tau > 0$, such that for any $z \in \{\zeta : |\zeta - \bar{\zeta}_{-,t}| \leq \tau t^2, |\mathrm{Im}\,\zeta| \geq N^{-100}\}$, we have $m_{\mathsf{shift}}(z) = \mathcal{O}(t^{-1}N^{1-\alpha/2})$ and*

$$\mathbb{E}m_X(z) = \mathsf{m}_{\mathsf{mp}}^{(t)}(z) + m_{\mathsf{shift}}(z) - \frac{\mathfrak{p}(z)}{2c_N z} + \mathcal{O}(N^{1-\alpha/2-\epsilon(4-\alpha)(\alpha-2)/50}). \tag{58}$$

Furthermore, for any $z \in \{\zeta : |\zeta - \bar{\zeta}_{-,t}| \ll t^2, |\mathrm{Im}\,\zeta| \geq N^{-100}\}$,

$$tm_{\mathsf{shift}}(z) = \frac{\mathsf{c}N^{1-\alpha/2}\int_0^\infty e^{-s-sc_N \mathsf{m}_{\mathsf{mp}}(\lambda_-^{\mathsf{mp}})}\left(s\mathsf{m}_{\mathsf{mp}}(\lambda_-^{\mathsf{mp}})\right)^{\alpha/2}\mathrm{d}s}{2\sqrt{c_N}(1-\sqrt{c_N})} + \mathfrak{o}(N^{1-\alpha/2}). \tag{59}$$

Proof. By the resolvent expansion, we have for any $z \in \{\zeta : |\zeta - \bar{\zeta}_{-,t}| \leq \tau t^2, |\mathrm{Im}\,\zeta| \geq N^{-100}\}$,

$$\left[G(X^\top, z)\right]_{ii} = -\left(z + zx_i^\top G(X^{(i)}, z)x_i\right)^{-1}. \tag{60}$$

22

Let $Q = Q_{\mathsf{diag}} + Q_{\mathsf{off}}$ with $Q_{\mathsf{diag}} := \sum_{j=1}^{M} x_{ji}^2 \big[G(X^{(i)}, z) \big]_{jj}$ and $Q_{\mathsf{off}} := \sum_{\ell \neq k} x_{ki} x_{\ell i} \big[G(X^{(i)}, z) \big]_{k\ell}$. Then, we can rewrite (60) as:

$$\big[G(X^\top, z) \big]_{ii} = -\frac{1}{z(1 + Q_{\mathsf{diag}})} + \frac{Q_{\mathsf{off}}}{z(1 + Q_{\mathsf{diag}})^2} - \frac{Q_{\mathsf{off}}^2}{z(1 + Q_{\mathsf{diag}})^2 (1 + Q)}. \tag{61}$$

Taking expectation at both sides gives

$$\mathbb{E}\big[G(X^\top, z) \big]_{ii} = -\frac{1}{z} \mathbb{E}\Big[\frac{1}{1 + Q_{\mathsf{diag}}} \Big] - \frac{1}{z} \mathbb{E}\Big[\frac{Q_{\mathsf{off}}^2}{(1 + Q_{\mathsf{diag}})^2 (1 + Q)} \Big] = I_1 + I_2,$$

where the second term at the right hand side of (61) vanished due to symmetry. Notice that when $\Psi^{(i)}$ is good, we have w.h.p. that

$$|\lambda_M(\mathcal{S}(X^{(i)})) - z| = |(1 - t)\lambda_-^{\mathsf{mp}} - \bar{\zeta}_{-,t}| - |\lambda_M(\mathcal{S}(X^{(i)})) - (1 - t)\lambda_-^{\mathsf{mp}}| - |\bar{\zeta}_{-,t} - z|$$
$$\geq \sqrt{c_N} t^2 - \sqrt{c_N} t^2 / 4 - \tau t^2 \geq \sqrt{c_N} t^2 / 2,$$

where in the last step we used the fact that $|\lambda_M(\mathcal{S}(X^{(i)})) - (1 - t)\lambda_-^{\mathsf{mp}}| \leq N^{-\epsilon_b}$ w.h.p., and we also chose $\tau < \sqrt{c_N} t^2 / 4$. This together with the fact that $\Psi^{(i)}$ is good w.h.p. gives $\mathbb{P}\big(\Omega_i = \big\{ |\lambda_M(\mathcal{S}(X^{(i)})) - z| \geq \sqrt{c_N} t^2 / 2 \big\} \big) \geq 1 - N^{-D}$. Notice that $\mathrm{Re}\, Q_{\mathsf{diag}} \geq 0$ and $\mathrm{Re}\, Q \geq 0$ hold on Ω_i. Then for I_2, with the smallness of $\mathbb{P}(\Omega_i^c)$, we have $I_2 = \mathbb{E}[Q_{\mathsf{off}}^2 \mathbf{1}_{\Omega_i} / [(1 + Q_{\mathsf{diag}})^2 (1 + Q)]] + \mathcal{O}(N^{-D})$. We then bound I_2 as

$$|I_2| \leq \mathbb{E}|Q_{\mathsf{off}}|^2 \mathbf{1}_{\Omega_i} + \mathcal{O}(N^{-D}) = 2 N^{-2} \mathbb{E}\Big[\mathrm{Tr}\, G(X^{(i)}, z) \overline{G(X^{(i)}, z)} \mathbf{1}_{\Omega_i} \Big] + \mathcal{O}(N^{-1}) = \mathcal{O}_\prec(t^{-4} N^{-1}).$$

Next, we estimate I_1. Due to the smallness of $\mathbb{P}(\Omega_i^c)$, we only have to do the estimation on the event Ω_i. Specially, we have $I_1 = -\mathbb{E}[\mathbf{1}_{\Omega_i} / (z + z Q_{\mathsf{diag}})] + \mathcal{O}(N^{-D})$. Notice that $\mathrm{Re}\, Q_{\mathsf{diag}} \geq 0$ on the event Ω_i. Using the identity that for w with $\mathrm{Re}\, w > 0$, $w^{-1} = \int_0^\infty e^{-sw} \mathrm{d}s$ and setting $w = 1 + Q_{\mathsf{diag}}$, we have

$$I_1 = -\frac{1}{z} \mathbb{E}\Big[\int_0^\infty e^{-s(1 + Q_{\mathsf{diag}})} \mathrm{d}s \cdot \mathbf{1}_{\Omega_i} \Big] + \mathcal{O}(N^{-D}) = -\frac{1}{z} \mathbb{E}\Big(\mathbb{E}_{x_i} \Big[\int_0^\infty e^{-s(1 + Q_{\mathsf{diag}})} \mathrm{d}s \Big] \cdot \mathbf{1}_{\Omega_i} \Big) + \mathcal{O}(N^{-D})$$
$$= -\frac{1}{z} \mathbb{E}\Big(\int_0^\infty e^{-s} \prod_j \phi_N\big(-\mathrm{i}s \big[G(X^{(i)}, z) \big]_{jj} \big) \mathrm{d}s \cdot \mathbf{1}_{\Omega_i} \Big) + \mathcal{O}(N^{-D}).$$

Then we may proceed as the estimation in the proof of Proposition 3.11 to obtain that

$$I_1 = -\frac{1}{z} \mathbb{E}\Big[\frac{1}{1 + (1 - t) \mathrm{Tr}\, G(X^{(i)}, z) / N} \cdot \mathbf{1}_{\Omega_i} \Big] - \frac{\mathfrak{p}(z)}{z} + \mathcal{O}(N^{-3(\alpha-2)/5}).$$

Further using the $\mathcal{O}_\prec(t^{-4} N^{-1})$ bound for $\mathrm{Var}(M^{-1} \mathrm{Tr}\, G(X^{(i)}, z))$ and the fact $\mathrm{Tr}\, G(X^{(i)}, z) - \mathrm{Tr}\, G(X, z) \prec t^{-4}$, we arrive at

$$I_1 = -\frac{1}{z} \Big(\frac{1}{1 + (1 - t) \mathbb{E}\mathrm{Tr}\, G(X, z) / N} \Big) - \frac{\mathfrak{p}(z)}{z} + \mathcal{O}(N^{-3(\alpha-2)/5}) + \mathcal{O}_\prec(t^{-4} N^{-1}).$$

Collecting the estimates for I_1 and I_2, and then summing over i, we have

$$N^{-1} \mathbb{E}\mathrm{Tr}\, G(X^\top, z) = -\frac{1}{z} \Big(\frac{1}{1 + (1 - t) \mathbb{E}\mathrm{Tr}\, G(X, z) / N} \Big) - \frac{\mathfrak{p}(z)}{z} + \mathcal{O}(N^{-3(\alpha-2)/5}).$$

Using the simple equation $\mathrm{Tr}\, G(X, z) - \mathrm{Tr}\, G(X^\top, z) = (N - M)/z$, the above equation can be rewritten as:

$$c_N \mathbb{E} m_X(z) = -\frac{1}{z} \Big(\frac{1}{1 + (1 - t) c_N \mathbb{E} m_X(z)} \Big) - \frac{\mathfrak{p}(z) + 1 - c_N}{z} + \mathcal{O}(N^{-3(\alpha-2)/5}). \tag{62}$$

Notice that for $z = \bar{\zeta}_{-,t} + \mathrm{i}N^{-100K_\zeta}$, we have $(\frac{z}{1-t} + c_N - 1)^2 - \frac{4c_N z}{1-t} = \frac{c_N t^2(2-t)^2}{(1-t)^2} + \mathcal{O}(N^{-90K_\zeta})$. Then by continuity, we may choose τ sufficiently small such that for any $z \in \{\zeta : |\zeta - \bar{\zeta}_{-,t}| \le \tau t^2, |\mathrm{Im}\,\zeta| \ge N^{-100K_\zeta}\}$, we have $(\frac{z}{1-t} + c_N - 1)^2 - \frac{4c_N z}{1-t} \sim t^2$. Having this bound, we may solve the quadric equation (62) and then compare it with (2) to obtain that

$$\mathbb{E}m_X(z) = \mathsf{m}_{\mathsf{mp}}^{(t)}(z) + m_{\mathsf{shift}}(z) - \frac{\mathfrak{p}(z)}{2c_N z} + \mathcal{O}(N^{-11(\alpha-2)/20}),$$

which proves (58). Using the fact that $\mathsf{m}_{\mathsf{mp}}^{(t)}(\bar{\zeta}_{-,t} + \mathrm{i}N^{-100K_\zeta}) - \mathsf{m}_{\mathsf{mp}}(\lambda_-^{\mathsf{mp}}) \le t$, we may further derive that

$$tm_{\mathsf{shift}}(\bar{\zeta}_{-,t} + \mathrm{i}N^{-100}) = \frac{\mathsf{c}N^{1-\alpha/2} \int_0^\infty e^{-s - sc_N \mathsf{m}_{\mathsf{mp}}(\lambda_-^{\mathsf{mp}})} (s\mathsf{m}_{\mathsf{mp}}(\lambda_-^{\mathsf{mp}}))^{\alpha/2} \mathrm{d}s}{2\sqrt{c_N}(1 - \sqrt{c_N})} + \mathcal{O}(tN^{1-\alpha/2}).$$

This together with the crude bound $\mathsf{m}_{\mathsf{mp}}^{(t)}(\bar{\zeta}_{-,t} + \mathrm{i}N^{-100K_\zeta}) - \mathsf{m}_{\mathsf{mp}}^{(t)}(\bar{\zeta}_{-,t}) = \mathcal{O}(N^{-90K_\zeta})$ proves (59), which completes the proof of Proposition 3.13. $\qquad\square$

The following corollary is a direct consequence of Proposition 3.13.

Corollary 3.14. *Let τ be chosen as in Proposition 3.13. Then for any $z \in \{\zeta : |\zeta - \bar{\zeta}_{-,t}| \le \tau t^2/2, |\mathrm{Im}\,\zeta| \ge N^{-100}\}$, we have $\mathbb{E}m_X^{(k)}(z) - (\mathsf{m}_{\mathsf{mp}}^{(t)}(z))^{(k)} = \mathcal{O}(t^{-(2k+1)}N^{1-\alpha/2})$,*

Proof. The claim follows from Proposition 3.13 with Cauchy integral. We omit further details. $\qquad\square$

Proof of Proposition 3.6. Replacing $\mathbb{E}[m_X(\hat{\zeta}_{\mathsf{e}})]$ by $\mathsf{m}_{\mathsf{mp}}^{(t)}(\hat{\zeta}_{\mathsf{e}})$ in the expression of λ_{shift}, we can obtain

$$\lambda_{\mathsf{shift}} = \bar{\Phi}_t(\zeta_{\mathsf{e}}) + \left(2c_N t\lambda_-^{\mathsf{mp}} + O(t^2)\right) \cdot \left(\mathsf{m}_{\mathsf{mp}}^{(t)}(\hat{\zeta}_{\mathsf{e}}) - \mathbb{E}[m_X(\hat{\zeta}_{\mathsf{e}})]\right) + \mathcal{O}(|\zeta_{\mathsf{e}} - \hat{\zeta}_{\mathsf{e}}|). \tag{63}$$

Expanding $\bar{\Phi}_t(\zeta_{\mathsf{e}})$ around $\bar{\zeta}_{-,t}$ and using the fact that $\bar{\Phi}_t'(\bar{\zeta}_{-,t}) = 0$, we have that there exists $\tilde{\zeta} \in [\bar{\zeta}_{-,t}, \zeta_{\mathsf{e}}]$ such that $\bar{\Phi}_t(\zeta_{\mathsf{e}}) = \bar{\Phi}_t(\bar{\zeta}_{-,t}) + \bar{\Phi}_t''(\tilde{\zeta})(\zeta_{\mathsf{e}} - \bar{\zeta}_{-,t})^2 = \lambda_-^{\mathsf{mp}} + \bar{\Phi}_t''(\tilde{\zeta})(\zeta_{\mathsf{e}} - \bar{\zeta}_{-,t})^2$. Substituting this expansion back into (63), and using the bound in Corollary 3.14, (63) becomes

$$\lambda_{\mathsf{shift}} = \lambda_-^{\mathsf{mp}} + 2c_N t\lambda_-^{\mathsf{mp}}\left(\mathsf{m}_{\mathsf{mp}}^{(t)}(\hat{\zeta}_{\mathsf{e}}) - \mathbb{E}[m_X(\hat{\zeta}_{\mathsf{e}})]\right) + \bar{\Phi}_t''(\tilde{\zeta})(\zeta_{\mathsf{e}} - \bar{\zeta}_{-,t})^2 + \mathfrak{o}(N^{1-\alpha/2}).$$

Note by considering that $\tilde{\zeta} - (1-t)\lambda_-^{\mathsf{mp}} \sim t^2$, it can be easily verified that $\bar{\Phi}_t''(\tilde{\zeta}) \sim t^{-2}$.

By employing Corollary 3.14 along with the variance bounds for $m_X^{(k)}(\hat{\zeta}_{\mathsf{e}})$ in Lemma 3.4, we can conclude that

$$\bar{\Delta}_m^{(k)}(\bar{\zeta}_{-,t}) := m_X^{(k)}(\bar{\zeta}_{-,t}) - (\mathsf{m}_{\mathsf{mp}}^{(t)}(\bar{\zeta}_{-,t}))^{(k)} = \mathcal{O}_p(N^{-1/2+\epsilon/2}t^{-2-k} + N^{1-\alpha/2}t^{-2k-1}).$$

With the above probabilistic bounds in place, we may now proceed to follow the expansion detailed in the proof of Lemma 3.4, but this time substitute ζ_{e} with $\bar{\zeta}_{-,t}$ and $\mathbb{E}(m_X^{(k)}(\hat{\zeta}_{\mathsf{e}}))$ with $\bar{m}_X^{(k)}(\bar{\zeta}_{-,t})$ (cf. (49)-(54)). It becomes evident that the ZOT_ζ therein vanishes due to the fact that $\bar{\Phi}_t'(\bar{\zeta}_{-,t}) = 0$. This eventually leads to $\bar{\Delta}_\zeta := \zeta_{-,t} - \bar{\zeta}_{-,t} = \mathcal{O}_p(N^{-1/2+\epsilon/2} + N^{1-\alpha/2})$. Therefore, with $\Delta_\zeta = \mathcal{O}_p(N^{-1/2+\epsilon/2}t^6)$, we have $\bar{\Phi}_t''(\tilde{\zeta})(\zeta_{\mathsf{e}} - \bar{\zeta}_{-,t})^2 \sim t^{-2}(\bar{\Delta}_\zeta - \Delta_\zeta)^2 = \mathfrak{o}(N^{1-\alpha/2})$. Consequently, we arrive at

$$\lambda_{\mathsf{shift}} = \lambda_-^{\mathsf{mp}} + 2c_N t\lambda_-^{\mathsf{mp}}\left(\mathsf{m}_{\mathsf{mp}}^{(t)}(\hat{\zeta}_{\mathsf{e}}) - \mathbb{E}m_X(\hat{\zeta}_{\mathsf{e}})\right) + \mathfrak{o}(N^{1-\alpha/2}).$$

Recalling from (45) that $\bar{\zeta}_{-,t} - \zeta_{\mathsf{e}} \prec N^{-\beta/2}t^2$, we can deduce that $\bar{\zeta}_{-,t} - \hat{\zeta}_{\mathsf{e}} \prec N^{-\beta/2}t^2$. The claim now follows by (59) in Proposition 3.13 and the fact $\mathsf{m}_{\mathsf{mp}}(\lambda_-^{\mathsf{mp}}) = (\sqrt{c_N} - c_N)^{-1}$. $\qquad\square$

4 Beyond Gaussian divisible model

In this section, we present three Green function function comparison results, as we mentioned in the Section 1. Their proofs will be postponed to the next section. Recall the notations in (14).

4.1 Entry-wise bound

We first introduce the following shorthand notation: for any $a, b \in [M]$ and $u, v \in [N]$,

$$\mathfrak{X}_{ab} = \mathfrak{X}_{ab}(\Psi) \coloneqq \begin{cases} 1 & \text{if } a \text{ or } b \in \mathcal{T}_r, \\ t^2 & \text{if } a \in \mathcal{D}_r, b \in \mathcal{D}_r \end{cases}, \quad \mathfrak{Y}_{uv} = \mathfrak{Y}_{uv}(\Psi) \coloneqq \begin{cases} 1 & \text{if } u \text{ or } v \in \mathcal{T}_c, \\ t^2 & \text{if } u \in \mathcal{D}_c, v \in \mathcal{D}_c \end{cases},$$

$$\mathfrak{Z}_{au} = \mathfrak{Z}_{au}(\Psi) \coloneqq \begin{cases} 1 & \text{if } a \in \mathcal{T}_r \text{ or } u \in \mathcal{T}_c \\ t^2 & \text{if } a \in \mathcal{D}_r, u \in \mathcal{D}_c \end{cases}.$$

Proposition 4.1 (Entry-wise bound). *Recall* $\mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ *defined in (22). Let* $\mathsf{D}_{\leq} = \{z = E + \mathrm{i}\eta \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3) : \eta \leq N^{-\varepsilon}\}$. *Set* $10\epsilon_a \leq \varepsilon_1 \leq \epsilon_b/500$, *and set* $\varepsilon_2, \varepsilon_3$ *sufficiently small, and* $3\epsilon_1 < \varepsilon \leq \epsilon_b/100$. *Suppose that* Ψ *is good. Let* $\mathbb{P}_\Psi$ *be the probability conditioned on the event that the* (ψ_{ij}) *matrix is a given* Ψ. *Suppose that* Ψ *is good (cf. (29)). Then for each* $\delta > 0$ *and* $D > 0$, *there exists a large constant* $C > 0$ *such that*

$$\mathbb{P}_\Psi\Big(\sup_{0 \leq \gamma \leq 1} \sup_{z \in \mathsf{D}_\leq} \sup_{a,b \in [M]} |\mathfrak{X}_{ab}[G^\gamma(z)]_{ab}| \vee \sup_{0 \leq \gamma \leq 1} \sup_{z \in \mathsf{D}_\leq} \sup_{u,v \in [N]} |\mathfrak{Y}_{uv}[\mathcal{G}^\gamma(z)]_{uv}|$$

$$\vee \sup_{0 \leq \gamma \leq 1} \sup_{z \in \mathsf{D}_\leq} \sup_{a \in [M], u \in [N]} |\mathfrak{Z}_{au}[G^\gamma(z)Y^\gamma]_{au}| \geq N^\delta \Big) \leq CN^{-D},$$

The proof of Proposition 4.1 follows a similar approach to the one demonstrated in [4, Proposition 3.17]. It relies on the entry-wise bounds for the Green functions of Y^0 as provided in Theorem 2.8, which serve as an input for the subsequent comparison theorem. We defer the proof to Section 5.2.

Theorem 4.2. *Let* $F : \mathbb{R} \to \mathbb{R}$ *be a function such that*

$$\sup_{0 \leq \mu \leq d} F^{(\mu)}(x) \leq (|x| + 1)^{C_0}, \qquad \sup_{\substack{0 \leq \mu \leq d \\ |x| \leq 2N^2}} F^{(\mu)}(x) \leq N^{C_0},$$

for some real number $C_0, d > 0$. *For any* $0 - 1$ *matrix* Ψ *and complex number* z, *we define for any* $a, b \in [M]$ *and* $u, v \in [N]$,

$$\mathfrak{I}_{0,ab} = \mathfrak{I}_{0,ab}(\Psi, z) \coloneqq \max_{0 \leq \mu \leq d} \sup_{0 \leq \gamma \leq 1} \mathbb{E}_\Psi\big(\big|F^{(\mu)}(\mathfrak{X}_{ab}\mathrm{Im}\,[G^\gamma(z)]_{ab})\big|\big),$$

$$\mathfrak{I}_{1,uv} = \mathfrak{I}_{1,uv}(\Psi, z) \coloneqq \max_{0 \leq \mu \leq d} \sup_{0 \leq \gamma \leq 1} \mathbb{E}_\Psi\big(\big|F^{(\mu)}(\mathfrak{Y}_{uv}\mathrm{Im}\,[\mathcal{G}^\gamma(z)]_{uv})\big|\big),$$

$$\mathfrak{I}_{2,au} = \mathfrak{I}_{2,au}(\Psi, z) \coloneqq \max_{0 \leq \mu \leq d} \sup_{0 \leq \gamma \leq 1} \mathbb{E}_\Psi\big(\big|F^{(\mu)}(\mathfrak{Z}_{au}\mathrm{Im}\,[G^\gamma(z)Y^\gamma]_{au})\big|\big),$$

and $\Omega = \Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_w$, $Q_0 = Q_0(\varepsilon, z) \coloneqq 1 - \mathbb{P}_\Psi\big(\Omega\big)$ *with*

$$\Omega_0 = \Omega_0(\varepsilon, z) \coloneqq \Big\{ \sup_{\substack{a,b \in [M] \\ 0 \leq \gamma \leq 1}} |\mathfrak{X}_{ab}[G^\gamma(z)]_{ab}| \leq N^\varepsilon \Big\}, \Omega_1 = \Omega_1(\varepsilon, z) \coloneqq \Big\{ \sup_{\substack{u,v \in [N] \\ 0 \leq \gamma \leq 1}} |\mathfrak{Y}_{uv}[\mathcal{G}^\gamma(z)]_{uv}| \leq N^\varepsilon \Big\},$$

$$\Omega_2 = \Omega_2(\varepsilon, z) \coloneqq \Big\{ \sup_{\substack{a \in [M], u \in [N] \\ 0 \leq \gamma \leq 1}} |\mathfrak{Z}_{au}[G^\gamma(z)Y^\gamma]_{au}| \leq N^\varepsilon \Big\}, \Omega_w = \Omega_w(\varepsilon) \coloneqq \Big\{ \sup_{i \in [M], j \in [N]} |w_{ij}| \leq N^{-1/2 + \varepsilon} t \Big\}.$$

Suppose that Ψ *is good. There exist sufficiently small positive constants* $\varepsilon \leq \epsilon_b/100$ *and* ω, *and a large constant* $C > 0$ *such that for*

$$(\#_1, \#_2, \#_3) \in \{(\mathfrak{X}_{ab}\mathrm{Im}\,[G^\gamma(z)]_{ab}, \ \mathfrak{X}_{ab}\mathrm{Im}\,[G^0(z)]_{ab}, \ \mathfrak{I}_{0,ab}),$$
$$(\mathfrak{Y}_{uv}\mathrm{Im}\,[\mathcal{G}^\gamma(z)]_{uv}, \ \mathfrak{Y}_{uv}\mathrm{Im}\,[\mathcal{G}^0(z)]_{uv}, \ \mathfrak{I}_{1,uv}),$$
$$(\mathfrak{Z}_{au}\mathrm{Im}\,[G^\gamma(z)Y^\gamma]_{au}, \ \mathfrak{Z}_{au}\mathrm{Im}\,[G^0(z)Y^0]_{au}, \ \mathfrak{I}_{2,au})\},$$

we have

$$\sup_{0 \le \gamma \le 1} \left| \mathbb{E}_\Psi \big(F(\#_1) \big) - \mathbb{E}_\Psi \big(F(\#_2) \big) \right| < C N^{-\omega} (\#_3 + 1) + C Q_0 N^{C+C_0}, \tag{64}$$

for any $a, b \in [M]$ and $u, v \in [N]$. The same estimates hold if Im *'s are replaced by* Re *'s.*

4.2 Average local law

In this section, we write $m^\gamma(z) = m_{Y^\gamma}(z)$, $G^\gamma(z) = G(Y^\gamma, z)$, and $\bar{G}^\gamma(z) = G(Y^\gamma, \bar{z})$ for simplicity. Let $z_t := \lambda_{-,t} + E + \mathrm{i}\eta$. Then we have the following theorem.

Theorem 4.3. *Suppose that Ψ is good. Let us define $z_t := \lambda_{-,t} + E + \mathrm{i}\eta$. We assume that $\eta \in [N^{-\frac{2}{3}-\epsilon}, N^{-\frac{2}{3}}]$, $E \in [-N^{-\varepsilon_1}, N^{-\frac{2}{3}+\epsilon}]$ for a sufficiently small $\epsilon > 0$. Then there exists a constant $\delta_0 > 0$ such that for all integer $p \ge 3$,*

$$\sup_{0 \le \gamma \le 1} \mathbb{E}_\Psi \big(\big| N\eta \big(\operatorname{Im} m^\gamma(z_t) - \operatorname{Im} \tilde{m}^0(z_t) \big) \big|^{2p} \big) \le (1 + \mathfrak{o}(1)) \mathbb{E}_\Psi \big(\big| N\eta \big(\operatorname{Im} m^0(z_t) - \operatorname{Im} \tilde{m}^0(z_t) \big) \big|^{2p} \big) + N^{-\delta_0 p},$$

where $\tilde{m}^0(z) = m_{X+t^{1/2}\tilde{W}}(z)$. Here $\tilde{W}$ is an i.i.d. copy of W and it is also independent of X. Further, the same estimate holds if Im *'s are replaced by* Re *'s.*

The above comparison inequality directly leads to the following theorem, which is crucial for the rigidity estimate for the $\lambda_M(\mathcal{S}(Y))$, serving as a key component in proving the universality result.

Theorem 4.4 (Rigidity estimate)**.** *Suppose Ω_Ψ holds. Then, with high probability,*

$$|\lambda_M(\mathcal{S}(Y)) - \lambda_{-,t}| \le N^{-2/3+\epsilon}.$$

Proof. By Markov's inequality, Theorem 4.3 and the following local law for m^0

$$|m^0(\lambda_{-,t} + E + \mathrm{i}\eta) - m_t(\lambda_{-,t} + E + \mathrm{i}\eta)| \prec \begin{cases} \frac{1}{N\eta}, & E \ge 0, \\ \frac{1}{N(|E|+\eta)} + \frac{1}{(N\eta)^2\sqrt{|E|+\eta}}, & E \le 0, \end{cases} \tag{65}$$

we can obtain (65) with m^0 replaced by m^1 and further for the case $E \le 0$ the following

$$\operatorname{Im} m^1(\lambda_{-,t} + E + \mathrm{i}\eta) - \operatorname{Im} m_t(\lambda_{-,t} + E + \mathrm{i}\eta) \prec \frac{1}{N(|E|+\eta)} + \frac{1}{(N\eta)^2\sqrt{|E|+\eta}} + \frac{1}{N^{1+\delta_0/2}\eta}. \tag{66}$$

We remark here that the local law in (65) has been proved in [22] around the right edge for the deformed rectangular matrices, under the assumption that the original rectangular matrices satisfy the η_*-regularity. The argument can be adapted to our model, but around the left edge, again with the η_*-regularity as the input. The derivation is almost the same, and thus we do not reproduce it here.

Further, similarly to Lemma 2.6, we can prove $|\lambda_M(\mathcal{S}(V_t)) - \lambda_-^{\mathsf{mp}}| \prec N^{-2\epsilon_b}$ and $|\lambda_M(\mathcal{S}(Y)) - \lambda_-^{\mathsf{mp}}| \prec N^{-2\epsilon_b}$. By (65), and the crude lower bound on $\lambda_M(\mathcal{S}(V_t))$ implied by [61], we also have $|\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t}| \prec N^{-\frac{2}{3}+\epsilon}$. Hence, we have

$$|\lambda_M(\mathcal{S}(Y)) - \lambda_{-,t}| \prec N^{-2\epsilon_b}. \tag{67}$$

With the aid of the m^1 analogue of (65), (66) and (67), the remaining reasoning is routine and thus we omit it; see the proof of Theorem 1.4 in [36], for instance. $\square$

4.3 Green function comparison for edge universality

Theorem 4.5 (Green function comparison)**.** *Let $F : \mathbb{R} \to \mathbb{R}$ be a function whose derivatives satisfy*

$$\max_x |F^\alpha(x)|(|x| + 1)^{-C_1} \le C_1, \quad \alpha = 1, \cdots, d$$

for some constant $C_1 > 0$ and sufficiently large integer $d > 0$. Let Ψ be good. Then there exist $\epsilon_0 > 0$, $N_0 \in \mathbb{N}$ and $\delta_1 > 0$ depending on ϵ_a such that for any $\epsilon < \epsilon_0$, $N \geq N_0$ and real numbers E, E_1 and E_2 satisfying $|E|$, $|E_1|$, $|E_2| \leq N^{-2/3+\epsilon}$, $\eta_0 = N^{-2/3-\epsilon}$, we have

$$\left| \mathbb{E}_\Psi \left[F\left(N \int_{E_1}^{E_2} \operatorname{Im} m^1(\lambda_{-,t} + y + i\eta_0) \, dy \right) \right] - \mathbb{E}_\Psi \left[F\left(N \int_{E_1}^{E_2} \operatorname{Im} m^0(\lambda_{-,t} + y + i\eta_0) \, dy \right) \right] \right| \leq C N^{-\delta_1},$$
(68)

for some constant $C > 0$, and in the case $\alpha = 8/3$, (68) holds with $\lambda_{-,t}$ replaced by λ_{shift}.

Employing the above comparison inequality along with the rigidity estimate int Theorem 4.4, we can deduce the following universality result around the random edge $\lambda_{-,t}$ (and deterministic edge λ_{shift} if $\alpha = 8/3$), whose proof will be stated in the supplementary material [11].

Corollary 4.6. *For all $s \in \mathbb{R}$, we have*

$$\lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t}) \leq s \right) = \lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_M(\mathcal{S}(Y)) - \lambda_{-,t}) \leq s \right).$$
(69)

Moreover, if $\alpha = 8/3$, we have

$$\lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{\text{shift}}) \leq s \right) = \lim_{N \to \infty} \mathbb{P}\left(N^{2/3}(\lambda_M(\mathcal{S}(Y)) - \lambda_{\text{shift}}) \leq s \right).$$
(70)

Now we can prove our main theorem: Theorem 1.1.

Proof of Theorem 1.1. The conclusions (i)-(iii) in Theorem 1.1 follows from (69) in Corollary 4.6 and Theorem 2.12. To prove the critical case when $\alpha = 8/3$, i.e., (iv), from (37) in Theorem 2.11, it is easy to show that the distribution of $\lambda_{-,t}$ is asymptotically independent of the fluctuation of $\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t}$ since the former is a function of X only. It can be shown by a standard characteristic function argument that for any $s \in \mathbb{R}$,

$$\lim_{N \to \infty} \mathbb{P}\left(\gamma_N M^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{\text{shift}}) \leq s \right) = \lim_{N \to \infty} \mathbb{P}\left(M^{2/3}(\mu_M^{\text{GOE}} + 2 + \gamma_N \mathcal{X}_\alpha) \leq s \right).$$

where in the RHS $\mathcal{X}_\alpha$ is independent of GOE. Then further, together with the comparison (70) we conclude (iv). Hence, we complete the proof of Theorem 1.1. $\qquad\square$

5 Proofs for the Green function comparisons

In this section, we will mainly prove the Green function comparisons stated in the last section. We will show the details for Theorems 4.2 and 4.3 only. The proof of 4.5 is similar to 4.3, and thus will only be discussed briefly here and the details are stated in the supplementary material [11].

5.1 Some further notations

Let us introduce some additional notations. We denote by $\mathsf{E}_{(ij)}$ the standard basis for $\mathbb{R}^{M \times N}$, i.e., $[\mathsf{E}_{(ij)}]_{ab} := \delta_{ia}\delta_{jb}$. Replacement matrix notation: For any $A \in \mathbb{R}^{M \times N}$, the replacement matrix $A^\lambda_{(ij)} = A_{(ij)}(\lambda) \in \mathbb{R}^{M \times N}$ is defined as,

$$\left[A_{(ij)}(\lambda) \right]_{ab} := \begin{cases} \lambda & \text{if } (i,j) = (a,b) \\ A_{ab} & \text{if } (i,j) \neq (a,b) \end{cases}, \quad a \in [M], \quad b \in [N].$$
(71)

Let $G^{\gamma,\lambda}_{(ij)}(z) := (\mathcal{S}(Y^{\gamma,\lambda}_{(ij)}) - z)^{-1}$ be the resolvent of $\mathcal{S}(Y^{\gamma,\lambda}_{(ij)})$ with $Y^{\gamma,\lambda}_{(ij)} = (Y^\gamma)_{(ij)}(\lambda)$. We define

$$d_{ij}(\gamma, w_{ij}) := \gamma(1 - \chi_{ij})a_{ij} + \chi_{ij}b_{ij} + (1 - \gamma^2)^{1/2} t^{1/2} w_{ji},$$
$$e_{ij}(\gamma, w_{ij}) := c_{ij} + (1 - \gamma^2)^{1/2} t^{1/2} w_{ij}, \qquad i \in [M], \quad j \in [N].$$
(72)

In the sequel, for brevity, we also write $\sum_{i,j} = \sum_{i=1}^{M} \sum_{j=1}^{N}$.

5.2 Proof of Proposition 4.1

Let us prove Proposition 4.1 assuming that Theorem 4.2 holds. The proof of Theorem 4.2 is deferred to the next subsection. For $\delta > 0$ and $z = E + i\eta \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ (cf. (22)), we define

$$\mathfrak{P}_0(\delta, z, \Psi) := \mathbb{P}_\Psi \Big(\sup_{0 \leq \gamma \leq 1} \sup_{a,b \in [M]} |z^{1/2} \mathfrak{X}_{ab}[G^\gamma(z)]_{ab}| > N^\delta \Big),$$

$$\mathfrak{P}_1(\delta, z, \Psi) := \mathbb{P}_\Psi \Big(\sup_{0 \leq \gamma \leq 1} \sup_{u,v \in [N]} |z^{1/2} \mathfrak{Y}_{uv}[\mathcal{G}^\gamma(z)]_{uv}| > N^\delta \Big),$$

$$\mathfrak{P}_2(\delta, z, \Psi) := \mathbb{P}_\Psi \Big(\sup_{0 \leq \gamma \leq 1} \sup_{a \in [M], u \in [N]} |\mathfrak{Z}_{au}[G^\gamma(z)Y^\gamma]_{au}| > N^\delta \Big).$$

The following monotonicity lemma will be a useful tool.

Lemma 5.1. *Suppose that Ψ is good. Fix ε and ω as in Theorem 4.2. For all $z = E + i\eta \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, we set $z' = E' + i\eta'$ by*

$$E' = E + \frac{(1 - N^{\varepsilon/3})(\sqrt{E^2 + \eta^2} - E)}{2}, \quad \eta' = N^{\varepsilon/6}\eta. \tag{73}$$

Then for any $\delta > 0$ and $D > 0$, there exists a large constant $C > 0$ such that

$$\max_{k \in \{0,1,2\}} \mathfrak{P}_k(\delta, z, \Psi) \leq C N^C \max_{k \in \{0,1,2\}} \mathfrak{P}_k(\varepsilon/2, z', \Psi) + C N^{-D}. \tag{74}$$

Proof. This is a minor modification of [4, Lemma 4.3]. The proof requires Theorem 4.2. For brevity, the detail is provided in the supplementaries [11]. □

With the above lemma, we can prove Proposition 4.1.

Proof of Proposition 4.1. The proof is similar to the proof of Proposition 3.17 in [4]. Let ε be as in Theorem 4.2. It follows from Lemma 5.1 that for any $z_0 = \lambda_-^{\mathsf{mp}} + E_0 + i\eta_0 \in \mathsf{D}(2\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and $\eta_0 \leq N^{-\varepsilon}$, we may find $z_1 = \lambda_-^{\mathsf{mp}} + E_1 + i\eta_1$ defined through (73) such that for any $\delta > 0$,

$$\max_{k \in [0:2]} \mathfrak{P}_k(\delta, z_0, \Psi) \leq C_1 N^{C_1} \max_{k \in [0:2]} \mathfrak{P}_k(\varepsilon/2, z_1, \Psi) + C_1 N^{-D}. \tag{75}$$

Now it suffices to bound $\max_{k \in [0:2]} \mathfrak{P}_k(\varepsilon/2, z_1, \Psi)$. Notice that for $\varepsilon > 3\varepsilon_1$

$$|E_1| \lesssim |E_0| + N^{\varepsilon/3}|\eta_0| \lesssim N^{-2\epsilon_1} + N^{-2/3\varepsilon} \ll N^{-\varepsilon_1}.$$

This means that $z_1 \in (\varepsilon_1, \varepsilon_2, \varepsilon_3)$. Applying Lemma 5.1 again with $\delta = \varepsilon/2$, we can find $z_2 = \lambda_-^{\mathsf{mp}} + E_2 + i\eta_2$

$$\max_{k \in [0:2]} \mathfrak{P}_k(\varepsilon/2, z_1, \Psi) \leq C_2 N^{C_2} \max_{k \in [0:2]} \mathfrak{P}_k(\varepsilon/2, z_2, \Psi) + C_2 N^{-D},$$

where $\eta_2 = N^{\varepsilon/6}\eta_1$ and $|E_2| \ll N^{-\varepsilon_1}$. We may now repeat the above procedure until $z_m = \lambda_-^{\mathsf{mp}} + E_m + i\eta_m$ with $\eta_m \geq K N^{-\varepsilon/2}$ for some sufficiently large K. It can be computed that

$$\eta_m \lesssim N^{-\varepsilon/2} N^{\varepsilon/6} = N^{-\varepsilon/3}, \quad \text{and} \quad |E_m| \lesssim |E_0| + \sum_{i=1}^{m-1} N^{\varepsilon/3}\eta_i, \quad \eta_i = N^{\varepsilon i/6}\eta_0.$$

This implies that $|E_m| \lesssim |E_0| + N^{-\varepsilon/2} \ll N^{-\varepsilon_1}$. Then using the fact that $\max_{k \in [0:2]} \mathfrak{P}_k(\varepsilon/2, z_m, \Psi) = 0$, we can obtain that

$$\max_{k \in [0:2]} \mathfrak{P}_k(\delta, z_0, \Psi) \leq C_1 N^{C_1} \max_{k \in [0:2]} \mathfrak{P}_k(\varepsilon/2, z_1, \Psi) + C_1 N^{-D} \leq C_m N^{-D}.$$

The claim now follows by adjusting constants. □

5.3 Proof of Theorem 4.2

We need the following elementary resolvent expansion formula.

Lemma 5.2. *For any deterministic matrix $A \in \mathbb{R}^{M \times N}$, let its linearisation $\mathcal{L}(A)$ be defined as*

$$\mathcal{L}(A) = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix}. \tag{76}$$

Let $\mathcal{R}(A, z) = (z^{1/2}\mathcal{L}(A) - z)^{-1}$ be the resolvent of $\mathcal{L}(A)$. The Schur complement formula also gives

$$\mathcal{R}(A, z) = \begin{pmatrix} G(A, z) & z^{-1/2}G(A, z)A \\ z^{-1/2}A^\top G(A, z) & G(A^\top, z) \end{pmatrix}.$$

Then for any $B = A + \Delta \in \mathbb{R}^{M \times N}$, we have for any integer $s \geq 0$

$$\mathcal{R}(A, z) = \sum_{j=0}^{s} \left(\mathcal{R}(B, z)\mathcal{L}(z^{1/2}\Delta)\right)^j \mathcal{R}(B, z) + \left(\mathcal{R}(B, z)\mathcal{L}(z^{1/2}\Delta)\right)^{s+1}\mathcal{R}(A, z).$$

Proof of Theorem 4.2. During the proof, we omit the z dependence and write $d_{ij} = d_{ij}(\gamma, w_{ij})$ and $e_{ij} = e_{ij}(\gamma, w_{ij})$ for simplicity. We only show the proof for $(\#_1, \#_2, \#_3) = (\mathfrak{X}_{ab}\mathrm{Im}\,[G^\gamma(z)]_{ab}, \mathfrak{X}_{ab}\mathrm{Im}\,[G^0(z)]_{ab}, \mathfrak{I}_{0,ab})$ with $a \in \mathcal{T}_r$ or $b \in \mathcal{T}_r$, and the others can be proved similarly. Observing that

$$\frac{\partial \mathbb{E}_\Psi\left(F([\mathrm{Im}\,G^\gamma]_{ab})\right)}{\partial \gamma} = -\sum_{i,j} \mathbb{E}_\Psi\left[F^{(1)}(\mathrm{Im}\,[G^\gamma]_{ab})\mathrm{Im}\left([G^\gamma]_{ib}[G^\gamma Y^\gamma]_{aj}\right)\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right)\right]$$

$$- \sum_{i,j} \mathbb{E}_\Psi\left[F^{(1)}(\mathrm{Im}\,[G^\gamma]_{ab})\mathrm{Im}\left([G^\gamma]_{ai}[(Y^\gamma)^\top G^\gamma]_{jb}\right)\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right)\right] = -\sum_{i,j}\left[(I)_{ij} + (II)_{ij}\right],$$

and therefore it suffices to show that there exists some constant C such that

$$\sum_{i,j}\left[|(I)_{ij}| + |(II)_{ij}|\right] \leq \frac{C}{(1-\gamma^2)^{1/2}}\left(N^{-\omega}(\mathfrak{I}_{0,ab} + 1) + Q_0 N^{C_0+C}\right). \tag{77}$$

We will focus on the estimation for $(I)_{ij}$'s, while the estimates for the $(II)_{ij}$'s can be handled in an identical fanshion. To ease the presentation, we further define the shorthand notation

$$f_{(ij)}(\lambda) = U_{(ij)}(\lambda)V_{(ij)}(\lambda),$$

$$U_{(ij)}(\lambda) = F^{(1)}\left(\mathrm{Im}\,[G^{\gamma,\lambda}_{(ij)}]_{ab}\right), \quad V_{(ij)}(\lambda) = \mathrm{Im}\left([G^{\gamma,\lambda}_{(ij)}]_{ib}[G^{\gamma,\lambda}_{(ij)}Y^{\gamma,\lambda}_{(ij)}]_{aj}\right).$$

We also define $\tilde{V}_{(ij)}(\lambda) = \mathrm{Im}\left([G^{\gamma,\lambda}_{(ij)}]_{ai}[(Y^{\gamma,\lambda}_{(ij)})^\top G^{\gamma,\lambda}_{(ij)}]_{jb}\right)$. Then for any $i \in [M], j \in [N]$, $(I)_{ij}$ can be rewritten as

$$(I)_{ij} = \mathbb{E}_\Psi\left[f_{(ij)}([Y^\gamma]_{ij})\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right) \cdot (\mathbf{1}_{\psi_{ij}=0} + \mathbf{1}_{\psi_{ij}=1})\right]$$

$$= \mathbb{E}_\Psi\left[f_{(ij)}(d_{ij})\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right)\right] \cdot \mathbf{1}_{\psi_{ij}=0} + \mathbb{E}_\Psi\left[f_{(ij)}(e_{ij})\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right)\right] \cdot \mathbf{1}_{\psi_{ij}=1}$$

$$\overset{(*)}{=} \mathbb{E}_\Psi\left[f_{(ij)}(d_{ij})\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right)\right] \cdot \mathbf{1}_{\psi_{ij}=0} - \frac{\gamma}{(1-\gamma^2)^{1/2}}t^{1/2}\mathbb{E}_\Psi\left[w_{ij}f_{(ij)}(e_{ij})\right] \cdot \mathbf{1}_{\psi_{ij}=1}$$

$$= (J_1)_{ij} - (J_2)_{ij},$$

where in $(*)$, we used the fact that $\mathsf{A}_{ij} = 0$ if $\psi_{ij} = 1$.

Let us consider $(J_2)_{ij}$ first. Applying Gaussian integration by parts on w_{ij}, we have

$$|(J_2)_{ij}| = \left| \frac{\gamma t^{1/2}}{(1-\gamma^2)^{1/2}N} \mathbb{E}_{\Psi}\left[\partial_{w_{ij}} f_{(ij)}(e_{ij}) \right] \mathbf{1}_{\psi_{ij}=1} \right|$$

$$\leq \frac{\gamma t^{1/2}}{(1-\gamma^2)^{1/2}N} \mathbb{E}_{\Psi}\left[\left| \partial_{w_{ij}} f_{(ij)}(e_{ij}) \mathbf{1}_{\Omega} \right| \right] \mathbf{1}_{\psi_{ij}=1} + \frac{\gamma t^{1/2}}{(1-\gamma^2)^{1/2}N} \mathbb{E}_{\Psi}\left[\left| \partial_{w_{ij}} f_{(ij)}(e_{ij}) \mathbf{1}_{\Omega^c} \right| \right] \mathbf{1}_{\psi_{ij}=1}.$$

Notice that

$$\partial_{w_{ij}} f_{(ij)}(e_{ij}) = U_{(ij)}(e_{ij}) \cdot \partial_{w_{ij}} V_{(ij)}(e_{ij}) + V_{(ij)}(e_{ij}) \cdot \partial_{w_{ij}} U_{(ij)}(e_{ij}), \tag{78}$$

and

$$\partial_{w_{ij}} U_{(ij)}(e_{ij}) = -(1-\gamma^2)^{1/2}t^{1/2}F^{(2)}\left(\operatorname{Im}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ab}\right)\left(V_{(ij)}(e_{ij}) + \tilde{V}_{(ij)}(e_{ij})\right),$$

$$\partial_{w_{ij}} V_{(ij)}(e_{ij}) = -(1-\gamma^2)^{1/2}t^{1/2}\operatorname{Im}\Big(\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ii}\left[(Y^{\gamma,e_{ij}}_{(ij)})^{\top}G^{\gamma,e_{ij}}_{(ij)}\right]_{jb}\left[G^{\gamma,e_{ij}}_{(ij)}Y^{\gamma,e_{ij}}_{(ij)}\right]_{aj}$$

$$+ \left[G^{\gamma,e_{ij}}_{(ij)}Y^{\gamma,e_{ij}}_{(ij)}\right]_{ij}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ib}\left[G^{\gamma,e_{ij}}_{(ij)}Y^{\gamma,e_{ij}}_{(ij)}\right]_{aj} - \left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ib}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ai}$$

$$+ \left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ib}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ai}\left[(Y^{\gamma,e_{ij}}_{(ij)})^{\top}G^{\gamma,e_{ij}}_{(ij)}Y^{\gamma,e_{ij}}_{(ij)}\right]_{jj} + \left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ib}\left[G^{\gamma,e_{ij}}_{(ij)}Y^{\gamma,e_{ij}}_{(ij)}\right]_{aj}\left[G^{\gamma,e_{ij}}_{(ij)}Y^{\gamma,e_{ij}}_{(ij)}\right]_{ij} \Big). \tag{79}$$

When $\psi_{ij} = 1$, we have $i \in \mathcal{D}_r$ and $j \in \mathcal{D}_c$. Then $\mathbf{1}_{\Omega}\mathbf{1}_{\psi_{ij}=1}|V_{(ij)}(e_{ij})| \leq N^{2\varepsilon}t^{-2}$ and $\mathbf{1}_{\Omega}\mathbf{1}_{\psi_{ij}=1}|\partial_{w_{ij}} V_{(ij)}(e_{ij})| \leq N^{3\varepsilon}t^{-7/2}$. Therefore, we may find a large constant $K_1 > 0$ such that

$$|(J_2)_{ij}| \lesssim \frac{\mathbf{1}_{\psi_{ij}=1}}{N^{1-3\varepsilon}t^3} \mathbb{E}_{\Psi}\left[\left| F^{(1)}\left(\operatorname{Im}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ab}\right) \right| \right] + \frac{\mathbf{1}_{\psi_{ij}=1}}{N^{1-4\varepsilon}t^3} \mathbb{E}_{\Psi}\left[\left| F^{(2)}\left(\operatorname{Im}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ab}\right) \right| \right]$$

$$+ \frac{t^{1/2}\mathbf{1}_{\psi_{ij}=1}}{N} \mathbb{E}_{\Psi}\left[\left| \partial_{w_{ij}} f_{(ij)}(e_{ij}) \mathbf{1}_{\Omega^c} \right| \right]$$

$$\lesssim \frac{\mathbf{1}_{\psi_{ij}=1}}{N^{1-3\varepsilon}t^3} \mathbb{E}_{\Psi}\left[\left| F^{(1)}\left(\operatorname{Im}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ab}\right) \right| \right] + \frac{\mathbf{1}_{\psi_{ij}=1}}{N^{1-4\varepsilon}t^3} \mathbb{E}_{\Psi}\left[\left| F^{(2)}\left(\operatorname{Im}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ab}\right) \right| \right] + N^{K_1}Q_0\mathbf{1}_{\psi_{ij}=1},$$

where in the second step, we used the crude bound that $|\partial_{w_{ij}} f_{(ij)}(e_{ij})| \leq N^{K_1}$ for some sufficiently large K_1, which can be obtained by the fact that $\operatorname{Im} z > N^{-1}$. By the facts that $\sum_{i,j} \mathbf{1}_{\psi_{ij}=1} \leq N^{1-\epsilon_\alpha}$ and $t \gg N^{-\epsilon_\alpha/4}$, we can choose $\varepsilon < \epsilon_\alpha/16$ to obtain that

$$|(J_2)_{ij}| \lesssim \frac{\mathbf{1}_{\psi_{ij}=1}}{N^{1-\epsilon_\alpha/2}} \mathbb{E}_{\Psi}\left[\left| F^{(1)}\left(\operatorname{Im}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ab}\right) \right| + \left| F^{(2)}\left(\operatorname{Im}\left[G^{\gamma,e_{ij}}_{(ij)}\right]_{ab}\right) \right| \right] + N^{K_1}Q_0\mathbf{1}_{\psi_{ij}=1}.$$

Next, we consider $(J_1)_{ij}$. Recall that $d_{ij} = \gamma(1-\chi_{ij})a_{ij} + \chi_{ij}b_{ij} + (1-\gamma^2)^{1/2}t^{1/2}w_{ji}$. Applying Taylor expansion on $f(d_{ij})$ around 0, for an s_1 to be chosen later, we have

$$(J_1)_{ij} = \sum_{k=0}^{s_1} \frac{\mathbf{1}_{\psi_{ij}=0}}{k!} \mathbb{E}_{\Psi}\left[(d_{ij})^k f^{(k)}_{(ij)}(0)\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right) \right]$$

$$+ \frac{\mathbf{1}_{\psi_{ij}=0}}{(s_1+1)!} \mathbb{E}_{\Psi}\left[(d_{ij})^{s_1+1} f^{(s_1+1)}_{(ij)}(\tilde{d}_{ij})\left(\mathsf{A}_{ij} - \frac{\gamma t^{1/2}w_{ij}}{(1-\gamma^2)^{1/2}}\right) \right] = \sum_{k=0}^{s} (J_1)_{ij,k} + \mathsf{Rem}.$$

where $\tilde{d}_{ij} \in [0, d_{ij}]$. Before proceeding to the estimation of $(J_1)_{ij,k}$ and Rem, we first establish perturbation bounds for the entries of the resolvents, which are useful for the estimation of $f^{(k)}_{(ij)}(0)$ and $f^{(k)}_{(ij)}(\tilde{d}_{ij})$. Using Lemma 5.2 and the notation therein, we have for any $\mathfrak{u}, \mathfrak{v} \in [M+N]$,

$$\mathbf{1}_{\Omega}\left[R(Y^{\gamma,d_{ij}}_{(ij)}, z) - R(Y^{\gamma,0}_{(ij)}, z) \right]_{\mathfrak{u}\mathfrak{v}} = \sum_{j=0}^{s} \mathbf{1}_{\Omega}\left[\left(R(Y^{\gamma,d_{ij}}_{(ij)}, z)\mathcal{L}(z^{1/2}d_{ij}\mathsf{E}_{ij}) \right)^j R(Y^{\gamma,d_{ij}}_{(ij)}, z) \right]_{\mathfrak{u}\mathfrak{v}}$$

30

$$+ \mathbf{1}_\Omega \big[\big(R(Y_{(ij)}^{\gamma, d_{ij}}, z) \mathcal{L}(z^{1/2} d_{ij} \mathsf{E}_{ij}) \big)^{s+1} R(Y_{(ij)}^{\gamma,0}, z) \big]_{\mathfrak{uv}}.$$

Further using the fact that $\mathbf{1}_\Omega |d_{ij}| \leq N^{-\epsilon_b}$, $\mathbf{1}_\Omega \big| \big[R(Y_{(ij)}^{\gamma, d_{ij}}, z) \big]_{\mathfrak{uv}} \big| \leq N^\varepsilon / t^2$, and the crude bound $\| R(Y_{(ij)}^{\gamma,0}, z) \| \leq N$ when $\mathrm{Im}\, z \geq N^{-1}$, we may choose s large enough to obtain that

$$\mathbf{1}_\Omega \big| \big[R(Y_{(ij)}^{\gamma, d_{ij}}, z) - R(Y_{(ij)}^{\gamma,0}, z) \big]_{\mathfrak{uv}} \big| \lesssim \sum_{j=1}^s \Big(\frac{N^{2\varepsilon}}{t^4 N^{\epsilon_b}} \Big)^j + \Big(\frac{N^\varepsilon}{t^2 N^{\epsilon_b}} \Big)^{s+1} N \lesssim 1, \tag{80}$$

which yields directly a control of $G_{(ij)}^{\gamma,0}$, $G_{(ij)}^{\gamma,0} Y_{(ij)}^{\gamma,0}$, and $(Y_{(ij)}^{\gamma,0})^\top G_{(ij)}^{\gamma,0} Y_{(ij)}^{\gamma,0}$ on the event Ω. Here we used the fact that $Y^\top G Y$ can be written in terms of $\mathcal{G}$, which can be seen easily by singular value decomposition. Similar estimates hold if $Y_{(ij)}^{\gamma,0}$ is replaced by $Y_{(ij)}^{\gamma, \tilde{d}_{ij}}$, we omit repetitive details. By taking derivatives repeatedly similar to (78) and (79), it can be easily seen that for any integer $k \geq 0$,

$$f_{(ij)}^{(k)}(d_{ij}) \cdot \mathbf{1}_{\psi_{ij}=0} \cdot \mathbf{1}_\Omega \lesssim \frac{N^{(C_0 + 2k+2)\varepsilon}}{t^{2k+2}}. \tag{81}$$

Combining the above estimate with the perturbation bounds in (80), we have for any $x \in [0, d_{ij}]$,

$$f_{(ij)}^{(k)}(x) \cdot \mathbf{1}_{\psi_{ij}=0} \cdot \mathbf{1}_\Omega \lesssim \frac{N^{(C_0 + 2k+2)\varepsilon}}{t^{2k+2}}. \tag{82}$$

Now we may start to estimate $(J_1)_{ij,k}$ and Rem. Using the above perturbation bounds on the event Ω, we have that there exists some large $K_2 > 0$, such that

$$\begin{aligned}
|\mathsf{Rem}| &\leq \frac{\mathbf{1}_{\psi_{ij}=0}}{(s_1+1)!} \mathbb{E}_\Psi \Big[\Big| (d_{ij})^{s_1+1} f_{(ij)}^{(s_1+1)}(\tilde{d}_{ij}) \Big(\mathsf{A}_{ij} - \frac{\gamma t^{1/2} w_{ij}}{(1-\gamma^2)^{1/2}} \Big) \Big| \cdot \mathbf{1}_\Omega \Big] \\
&\quad + \frac{\mathbf{1}_{\psi_{ij}=0}}{(s_1+1)!} \mathbb{E}_\Psi \Big[\Big| (d_{ij})^{s_1+1} f_{(ij)}^{(s_1+1)}(\tilde{d}_{ij}) \Big(\mathsf{A}_{ij} - \frac{\gamma t^{1/2} w_{ij}}{(1-\gamma^2)^{1/2}} \Big) \Big| \cdot \mathbf{1}_{\Omega^c} \Big] \\
&\lesssim \frac{N^{(C_0 + 2s_1+4)\varepsilon} \mathbf{1}_{\psi_{ij}=0}}{t^{2s_1+4} N^{1/2+\epsilon_a+(s_1+1)\epsilon_b}} + N^{K_2} Q_0 \mathbf{1}_{\psi_{ij}=0}.
\end{aligned} \tag{83}$$

Therefore, with the fact that $t \gg N^{-\epsilon_b/8}$, we may choose $\varepsilon < \epsilon_b/8$ and $s_1 > C_0/4 + 6/\epsilon_b$ to obtain that

$$|\mathsf{Rem}| \lesssim N^{-3} \cdot \mathbf{1}_{\psi_{ij}=0} + N^{K_2} Q_0 \cdot \mathbf{1}_{\psi_{ij}=0}.$$

We estimate $(J_1)_{ij,k}$ for different k separately. For the case when k is even, it follows from the symmetric condition that $(J_1)_{ij,k} = 0$. Thus we mainly focus on the estimation for k is odd.

Case 1: $k \geq 5$. First note by symmetry condition, we can obtain

$$|(J_1)_{ij,k}| \lesssim \sum_{\substack{u_1+u_2 \geq 1, u_3 \geq 0 \\ u_1+u_2+u_3=(k+1)/2}} \mathbb{E}_\Psi \Big[|\mathsf{A}_{ij}|^{2u_1} |t^{1/2} w_{ij}|^{2u_2} |b_{ij}|^{2u_3} |f_{(ij)}^{(k)}(0)| \Big] \mathbf{1}_{\psi_{ij}=0} \lesssim \frac{t \mathbf{1}_{\psi_{ij}=0} \mathbb{E}_\Psi[|f_{(ij)}^{(k)}(0)|]}{N^{2+(k-3)\epsilon_b}},$$

where in the last step we also used the fact that $\mathbb{E}_\Psi(b_{ij}^k) \leq N^{-\varepsilon_b(k-2)} \mathbb{E}_\Psi(x_{ij}^2) \lesssim N^{-1-\varepsilon_b(k-2)}$ for $k \geq 2$. We need to estimate $f_{(ij)}^{(k)}(0)$ again by Taylor expansion. For an s_2 to be chosen later, there exists $\hat{d}_{ij} \in [0, d_{ij}]$ such that

$$|(J_1)_{ij,k}| \lesssim \sum_{\ell=0}^{s_2} \frac{t \mathbf{1}_{\psi_{ij}=0} \mathbb{E}_\Psi[|f_{(ij)}^{(k+\ell)}(d_{ij})|]}{N^{2+(k+\ell-3)\epsilon_b}} + \frac{t \mathbf{1}_{\psi_{ij}=0} \mathbb{E}_\Psi[|f_{(ij)}^{(k+s_2+1)}(\hat{d}_{ij})|]}{N^{2+(k+s_2-2)\epsilon_b}}. \tag{84}$$

On the event Ω^c, we may estimate the RHS in the above display as in the last step in (83), which gives

$$\Big(\sum_{\ell=0}^{s_2} \frac{t \mathbb{E}_\Psi[|f_{(ij)}^{(k+\ell)}(d_{ij})|\mathbf{1}_{\Omega^c}]}{N^{2+(k+\ell-3)\epsilon_b}} + \frac{t \mathbb{E}_\Psi[|f_{(ij)}^{(k+s_2+1)}(\hat{d}_{ij})|\mathbf{1}_{\Omega^c}]}{N^{2+(k+s_2-2)\epsilon_b}} \Big) \mathbf{1}_{\psi_{ij}=0} \lesssim N^{K_3} Q_0 \mathbf{1}_{\psi_{ij}=0}, \tag{85}$$

for some large $K_3 > 0$. On the event Ω, we may choose $s_2 > C_0 + 30 + 4/\epsilon_b$ and $\varepsilon < \epsilon_b/8$ to obtain that

$$\Big(\sum_{\ell=0}^{s_2} \frac{t}{N^{2+(k+\ell-3)\epsilon_b}} \mathbb{E}_\Psi\Big[|f_{(ij)}^{(k+\ell)}(d_{ij})|\mathbf{1}_\Omega\Big] + \frac{t}{N^{2+(k+s_2-2)\epsilon_b}} \mathbb{E}_\Psi\Big[|f_{(ij)}^{(k+s_2+1)}(\hat{d}_{ij})|\mathbf{1}_\Omega\Big] \Big) \cdot \mathbf{1}_{\psi_{ij}=0}$$

$$\lesssim \sum_{\ell=0}^{s_2} \frac{t\mathbf{1}_{\psi_{ij}=0}}{N^{2+(k+\ell-3)\epsilon_b}} \mathbb{E}_\Psi\Big[|f_{(ij)}^{(k+\ell)}(d_{ij})|\mathbf{1}_\Omega\Big] + N^{-3}\mathbf{1}_{\psi_{ij}=0}$$

$$\lesssim \sum_{\ell=0}^{s_2} \frac{N^{2(k+\ell+1)\varepsilon}\mathbf{1}_{\psi_{ij}=0}}{N^{2+(k+\ell-3)\epsilon_b} t^{2(k+\ell)+1}} \sum_{m=1}^{k+\ell+1} \mathbb{E}_\Psi\Big[|F^{(m)}\big(\mathrm{Im}\,[G_{(ij)}^{\gamma,d_{ij}}]_{ab}\big)|\Big] + N^{-3}\mathbf{1}_{\psi_{ij}=0}. \tag{86}$$

Collecting the above estimates and choosing $\varepsilon < \epsilon_b/100$, we have

$$|(J_1)_{ij,k}| \lesssim \frac{\mathbf{1}_{\psi_{ij}=0}}{N^{2+\epsilon_b/2}} \sum_{m=1}^{k+s_2+1} \mathbb{E}_\Psi\Big[|F^{(m)}\big(\mathrm{Im}\,[G_{(ij)}^{\gamma,d_{ij}}]_{ab}\big)|\Big] + \frac{\mathbf{1}_{\psi_{ij}=0}}{N^{2+\epsilon_b/2}} + N^{K_3} Q_0 \mathbf{1}_{\psi_{ij}=0}.$$

Case 2: $k = 3$. By direct calculation, we have

$$(J_1)_{ij,3} \asymp \mathbb{E}_\Psi\Big[\big(\mathsf{A}_{ij}^4 + t\mathsf{A}_{ij}^2 w_{ij}^2 + t^2 w_{ij}^4 + \mathsf{A}_{ij}^2 \mathsf{B}_{ij}^2 + t w_{ij}^2 \mathsf{B}_{ij}^2\big) f_{(ij)}^{(3)}(0)\Big]\mathbf{1}_{\psi_{ij}=0}.$$

The term $\mathsf{A}_{ij}^2 \mathsf{B}_{ij}^2$ becomes null due to the definitions of A_{ij} and B_{ij}. Concerning the remaining terms, we only show how to estimate the term involving $t w_{ij}^2 \mathsf{B}_{ij}^2$ while the others can be handled similarly. Applying Taylor expansion, and then estimating terms on Ω^c and Ω separaterly as in (84)-(86), we have for $s_3 > C_0/4 + \epsilon_b/2 + 4$,

$$\Big| \mathbb{E}_\Psi\Big[t w_{ij}^2 \mathsf{B}_{ij}^2 f_{(ij)}^{(3)}(0)\Big]\mathbf{1}_{\psi_{ij}=0} \Big| \lesssim \sum_{\ell=0}^{s_3} \frac{t\mathbf{1}_{\psi_{ij}=0}\mathbb{E}_\Psi[|f_{(ij)}^{(3+\ell)}(d_{ij})|\mathbf{1}_\Omega]}{N^{2+\ell\epsilon_b}} + N^{-3}\mathbf{1}_{\psi_{ij}=0} + N^{K_5} Q_0 \mathbf{1}_{\psi_{ij}=0},$$

for some large $K_5 > 0$. Then it remains to estimate the first term of the RHS of the above inequality. For $\ell \geq 1$, the estimate is similar to (86), we omit further details. Here we focus on the non-trivial term when $\ell = 0$. It is straightforward to compute that $f_{(ij)}^{(3)}(d_{ij})$ is the products of $F^{(\ell)}\big(\mathrm{Im}\,[G_{(ij)}^{\gamma,d_{ij}}]_{ab}\big), \ell \in [4]$, and the entries of $G_{(ij)}^{\gamma,d_{ij}}$, $G_{(ij)}^{\gamma,d_{ij}} Y_{(ij)}^{\gamma,d_{ij}}$, and $(Y_{(ij)}^{\gamma,d_{ij}})^\top G_{(ij)}^{\gamma,d_{ij}} Y_{(ij)}^{\gamma,d_{ij}}$, where the entries' indices can be $(i,i), (j,j), (i,j), (a,i), (a,j), (i,b), (j,b)$. Therefore,

$$\frac{t}{N^2} \mathbb{E}_\Psi\Big[|f_{(ij)}^{(3)}(d_{ij})| \cdot \mathbf{1}_\Omega\Big] \cdot \mathbf{1}_{\psi_{ij}=0} \leq \frac{tN^{6\varepsilon}}{N^2} \sum_{\ell=1}^{4} \mathbb{E}_\Psi\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G_{(ij)}^{\gamma,d_{ij}}]_{ab}\big)|\Big] \cdot \mathbf{1}_{\psi_{ij}=0} \cdot \mathbf{1}_{i\in\mathcal{T}_r, j\in\mathcal{T}_c}$$

$$+ \frac{N^{6\varepsilon}}{N^2 t^7} \sum_{\ell=1}^{4} \mathbb{E}_\Psi\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G_{(ij)}^{\gamma,d_{ij}}]_{ab}\big)|\Big] \cdot \mathbf{1}_{\psi_{ij}=0} \cdot (1 - \mathbf{1}_{i\in\mathcal{T}_r, j\in\mathcal{T}_c}).$$

This eventually leads to

$$|(J_1)_{ij,3}| \lesssim \frac{tN^{6\varepsilon}\mathbf{1}_{\psi_{ij}=0}\mathbf{1}_{i\in\mathcal{T}_r, j\in\mathcal{T}_c}}{N^2} \sum_{\ell=1}^{4} \mathbb{E}_\Psi\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G_{(ij)}^{\gamma,d_{ij}}]_{ab}\big)|\Big]$$

$$+ \frac{N^{6\varepsilon}\mathbf{1}_{\psi_{ij}=0}(1 - \mathbf{1}_{i\in\mathcal{T}_r, j\in\mathcal{T}_c})}{N^2 t^7} \sum_{\ell=1}^{4} \mathbb{E}_\Psi\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G_{(ij)}^{\gamma,d_{ij}}]_{ab}\big)|\Big]$$

$$+ \frac{\mathbf{1}_{\psi_{ij}=0}}{N^{2+\epsilon_b/2}} \sum_{\ell=1}^{s_3+4} \mathbb{E}_{\Psi}\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G^{\gamma,d_{ij}}_{(ij)}]_{ab}\big)|\Big] + N^{-3}\mathbf{1}_{\psi_{ij}=0} + N^{K_5}Q_0\mathbf{1}_{\psi_{ij}=0}.$$

The second term in the above display can be estimated by the fact that $|\mathcal{D}_r| \vee |\mathcal{D}_c| \le N^{1-\epsilon_d}$ for some $\epsilon_d > 0$. By the fact that $t \gg N^{-\epsilon_d/20} \vee N^{-\epsilon_b/20}$, we then have

$$|(J_1)_{ij,3}| \lesssim \frac{t^{1/2}}{N^2} \sum_{\ell=1}^{4} \mathbb{E}_{\Psi}\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G^{\gamma,d_{ij}}_{(ij)}]_{ab}\big)|\Big] \cdot \mathbf{1}_{\psi_{ij}=0} \cdot \mathbf{1}_{i\in\mathcal{T}_r, j\in\mathcal{T}_c}$$

$$+ \frac{\mathbf{1}_{\psi_{ij}=0} \cdot (1 - \mathbf{1}_{i\in\mathcal{T}_r, j\in\mathcal{T}_c})}{N^{2-\epsilon_d/2}} \sum_{\ell=1}^{4} \mathbb{E}_{\Psi}\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G^{\gamma,d_{ij}}_{(ij)}]_{ab}\big)|\Big]$$

$$+ \frac{\mathbf{1}_{\psi_{ij}=0}}{N^{2+\epsilon_b/2}} \sum_{\ell=1}^{s_3+4} \mathbb{E}_{\Psi}\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G^{\gamma,d_{ij}}_{(ij)}]_{ab}\big)|\Big] + N^{-3} \cdot \mathbf{1}_{\psi_{ij}=0} + N^{K_5}Q_0 \cdot \mathbf{1}_{\psi_{ij}=0}.$$

Case 3: $k = 1$. In this case, using the fact that $\mathbb{E}_{\Psi}[b_{ij}] = 0$, we may compute

$$(J_1)_{ij,1} = \gamma\mathbb{E}_{\Psi}\Big[\big((1-\chi_{ij})^2 a_{ij}^2 - tw_{ij}^2\big)f^{(1)}_{(ij)}(0)\Big] \cdot \mathbf{1}_{\psi_{ij}=0}.$$

Recall that $t = N\mathbb{E}(\mathsf{A}_{ij}^2) = N\mathbb{E}\big((1-\psi_{ij})^2(1-\chi_{ij})^2 a_{ij}^2\big)$. This gives

$$\big|\mathbb{E}\big((1-\chi_{ij})^2 a_{ij}^2\big) - \mathbb{E}\big(tw_{ij}^2\big)\big| \lesssim tN^{-1-\alpha/2+\alpha\epsilon_b}. \tag{87}$$

Therefore, following the same procedure as in (84)-(86), we can also obtain that for sufficiently large constant s_4,

$$|(J_1)_{ij,1}| \lesssim \frac{\mathbf{1}_{\psi_{ij}=0}}{N^{2+\epsilon_b/2}} \sum_{\ell=1}^{s_4} \mathbb{E}_{\Psi}\Big[|F^{(\ell)}\big(\mathrm{Im}\,[G^{\gamma,d_{ij}}_{(ij)}]_{ab}\big)|\Big] + N^{-3} \cdot \mathbf{1}_{\psi_{ij}=0} + N^{K_5}Q_0 \cdot \mathbf{1}_{\psi_{ij}=0}.$$

By combining the estimates of $(J_1)_{ij,k}$'s with $(J_2)_{ij}$'s, we can conclude that (77) holds when we choose $\varepsilon \le \min\{\epsilon_\alpha, \epsilon_b, \epsilon_d\}/(100)$. $\qquad\square$

5.4 Proof of Theorem 4.3

Since we need to perform the comparison at a random edge, we begin with some preliminary estimates for the derivatives w.r.t. the matrix entries of the random edge.

Lemma 5.3 ([22], Lemma 5). *Denote* $a_{k,\pm}(t) = \Phi_t(\zeta_{k,\pm}(t))$, $1 \le k \le q$. *Then* $(a_{k,\pm}(t), \zeta_{k,\pm}(t))$ *are real solutions of*

$$F_t(z, \zeta) = 0, \quad and \quad \frac{\partial F_t}{\partial \zeta}(z, \zeta) = 0,$$

where

$$F_t(z, \zeta) = 1 + \frac{t(1-c_N) - \sqrt{t^2(1-c_N)^2 + 4\zeta z}}{2\zeta} - c_N tm_X(\zeta).$$

Using the lemma above, we can derive bounds for the derivatives of the random edge $\lambda_{-,t}$ w.r.t. the matrix entries b_{ij}.

Lemma 5.4. *Suppose that Ψ is good. If we view $\lambda_{-,t}$ as a function of $\mathsf{B}_{ij}, i \in [M], j \in [N]$. For any $i \in [M]$ and $j \in [N]$, write $\lambda_{-,t}(x) = \lambda_{-,t}(\mathsf{B}_{ij} = x)$. Then for any integer $k \ge 1$ and for any $b \in [0, \mathsf{B}_{ij}]$, we have*

$$\Big|\frac{\partial^k \lambda_{-,t}}{\partial \mathsf{B}_{ij}^k}(b)\Big|\mathbf{1}_{\psi_{ij}=0} \prec \frac{1}{Nt^{2k+1}}, \qquad \Big|\frac{\partial^k \zeta_{-,t}}{\partial \mathsf{B}_{ij}^k}(b)\Big|\mathbf{1}_{\psi_{ij}=0} \prec \frac{1}{Nt^{2k+1}}. \tag{88}$$

Further, there exists some constants $C_k > 0$ such that the following deterministic bounds hold,

$$\left|\frac{\partial \lambda_{-,t}}{\partial \mathsf{B}_{ij}}(b)\right| \mathbf{1}_{\psi_{ij}=0} \leq N^{C_1}, \qquad \left|\frac{\partial^k \lambda_{-,t}}{\partial \mathsf{B}_{ij}^k}(b)\right| \mathbf{1}_{\psi_{ij}=0} \cdot \Xi(\lambda_{-,t}) \leq N^{C_k}, \tag{89}$$

where $\Xi(x)$ is a smooth cut off function which equals 0 when $x < \lambda_-^{\mathsf{mp}}/100$ and 1 when $x > \lambda_-^{\mathsf{mp}}/2$ and $|\Xi^{(n)}(x)| = \mathcal{O}(1)$ for all $n \geq 1$.

Remark 7. *Here we remark that in the second estimate of (89), we added a cutoff function, in order to get a deterministic bound for the $\lambda_{-,t}$ derivatives, which is needed when we take expectation $\mathbb{E}_\Psi$. Hence, actually, we should work with $\mathbb{E}_\Psi\big(|N\eta\big(\mathrm{Im}\,m^\gamma(z_t) - \mathrm{Im}\,\tilde{m}^0(z_t)\big)\Xi(\lambda_{-,t})|^{2p}\big)$ instead of $\mathbb{E}_\Psi\big(|N\eta\big(\mathrm{Im}\,m^\gamma(z_t) - \mathrm{Im}\,\tilde{m}^0(z_t)\big)|^{2p}\big)$ to make sure that all quantities in the expansions have bounded expectations. Adding such a cutoff factor will not complicates the expansions since again by the chain rule it boils down to the $\lambda_{-,t}$ derivatives. Hence, additional technical inputs are not needed for the comparison of the modified quantity. However, in order to ease the presentation, we will state the reasoning for the original quantity and proceed as if all random factors in the expansion have deterministic upper bound.*

Proof. Let $\psi_{ij} = 0$. To emphasis the dependence with X, we first note that $F_t(z, \zeta)$ can be rewritten as,

$$F_t(z, \zeta, X) = 1 + \frac{t(1 - c_N) - \sqrt{t^2(1 - c_N)^2 + 4\zeta z}}{2\zeta} - \frac{c_N t}{M} \mathrm{Tr} G(X, \zeta).$$

Using Lemma 5.3, we have

$$F_t(\lambda_{-,t}, \zeta_{-,t}, X) = 0, \quad \text{and} \quad \frac{\partial F_t}{\partial \zeta}(\lambda_{-,t}, \zeta_{-,t}, X) = 0. \tag{90}$$

Then taking derivative of (90) gives

$$\frac{\partial \lambda_{-,t}}{\partial \mathsf{B}_{ij}} \frac{\partial F_t}{\partial z}(\lambda_{-,t}, \zeta_{-,t}, X) + \frac{\partial F_t}{\partial x_{ij}}(\lambda_{-,t}, \zeta_{-,t}, X)) = 0.$$

Therefore, we may solve the above equation to obtain that

$$\frac{\partial \lambda_{-,t}}{\partial \mathsf{B}_{ij}} = \frac{2c_N t \sqrt{t^2(1 - c_N)^2 + 4\lambda_{-,t}\zeta_{-,t}}}{M}\big[X^\top(G(X, \zeta_{-,t}))^2\big]_{ji}. \tag{91}$$

Notice that

$$\begin{aligned}
\big|\big[X^\top(G(X, \zeta_{-,t}))^2\big]_{ji}\big| &\overset{(i)}{\leq} \big|\big[X^\top(G(X, \zeta_{-,t}))^2 X\big]_{jj}\big|^{1/2} \cdot \big|\big[(G(X, \zeta_{-,t}))^2\big]_{ii}\big|^{1/2} \\
&= \big|\big[G(X^\top, \zeta_{-,t})\big]_{jj} + \zeta_{-,t}\big[(G(X^\top, \zeta_{-,t}))^2\big]_{jj}\big|^{1/2} \cdot \big|\big[(G(X, \zeta_{-,t}))^2\big]_{ii}\big|^{1/2} \\
&\lesssim \big(\|G(X^\top, \zeta_{-,t})\|^{1/2} + |\zeta_{-,t}|\|G(X^\top, \zeta_{-,t})\|\big) \cdot \|G(X, \zeta_{-,t})\| \overset{(ii)}{\prec} t^{-4}, \tag{92}
\end{aligned}$$

where in (i) we applied Cauchy-Schwarz inequality, and in (ii) we used Lemma 3.2 (i). Therefore, we can obtain that $\partial_{\mathsf{B}_{ij}} \lambda_{-,t}(b)$.

Next we view $\Phi_t(\zeta)$ as a function of X, and write $\Phi_t(\zeta, X) = \Phi_t(\zeta)$. By Lemma 3.1, we have $\frac{\partial \Phi_t}{\partial \zeta}(\zeta_{-,t}, X) = 0$. Further taking derivative w.r.t B_{ij} on this equation gives

$$\frac{\partial^2 \Phi_t}{\partial \zeta^2}(\zeta_{-,t}, X)\frac{\partial \zeta_{-,t}}{\partial \mathsf{B}_{ij}} + \frac{\partial^2 \Phi_t}{\partial \zeta \partial x_{ij}}(\zeta_{-,t}, X) = 0. \tag{93}$$

By direct calculation, we have

$$\frac{\partial^2 \Phi_t}{\partial \zeta \partial x_{ij}}(\zeta_{-,t}, X) = \frac{4c_N t}{M}[X^\top(G(X, \zeta_{-,t}))^2]_{ji} - \frac{2c_N^2 t^2 m_X(\zeta_{-,t})}{M}[X^\top(G(X, \zeta_{-,t}))^2]_{ji}$$

34

$$+ \frac{8c_N t\zeta_{-,t}(1 - c_N t m_X(\zeta_{-,t}))}{M}[X^\top(G(X,\zeta_{-,t}))^3]_{ji} - \frac{4c_N^2 t^2 \zeta_{-,t} m'_X(\zeta_{-,t})}{M}[X^\top(G(X,\zeta_{-,t}))^2]_{ji}$$

$$+ \frac{4c_N(1 - c_N)t^2}{M}[X^\top(G(X,\zeta_{-,t}))^2]_{ji}.$$

A similar argument as in (92) leads to $\left[X^\top(G(X,\zeta_{-,t}))^3\right]_{ji} \prec t^{-6}$. This together with the fact that $c_N t m_X(\zeta_{-,t}) \prec t^{1/2}$ and $m'_X(\zeta_{-,t}) \sim t^{-1}$ gives $\frac{\partial^2 \Phi_t}{\partial \zeta \partial x_{ij}}(\zeta_{-,t}, X) \prec 1/(Nt^5)$. We can also compute that

$$\frac{\partial^2 \Phi_t}{\partial \zeta^2}(\zeta_{-,t}, X) = -2c_N t m''_X(\zeta_{-,t})\zeta_{-,t}(1 - c_N t m_X(\zeta_{-,t})) - 4c_N t m'_X(\zeta_{-,t})(1 - c_N t m_X(\zeta_{-,t}))$$

$$+ 2\zeta_{-,t}(c_N t m'_X(\zeta_{-,t}))^2 - c_N(1 - c_N)t^2 m''_X(\zeta_{-,t}). \tag{94}$$

Using Lemma 2.2 with the fact $\zeta_{-,t} - \lambda_M(\mathcal{S}(X)) \sim t^2$ w.h.p., we have w.h.p. that $\frac{\partial^2 \Phi_t}{\partial \zeta^2}(\zeta_{-,t}, X) \sim t^2$. Combining the above bounds gives $\partial_{\mathsf{B}_{ij}}\zeta_{-,t} \prec 1/(Nt^3)$.

It is worth noting that for any integer $k \geq 2$, the $\partial^k_{\mathsf{B}_{ij}}\lambda_{-,t}$ can be expressed as a function of $\partial^\ell_{\mathsf{B}_{ij}}\lambda_{-,t}$ and $\partial^\ell_{\mathsf{B}_{ij}}\zeta_{-,t}$, where ℓ ranges from 0 to $k - 1$. Similarly, $\partial^k_{\mathsf{B}_{ij}}\zeta_{-,t}$ is solely dependent on $\partial^\ell_{\mathsf{B}_{ij}}\zeta_{-,t}$, where ℓ ranges from 0 to $k - 1$. By employing the product rule and adopting a similar argument as used in (92) to bound the Green function entries, we can observe that the order of $\partial^\ell_{\mathsf{B}_{ij}}\lambda_{-,t}$ is determined by the term that includes $\partial^{k-1}_{\mathsf{B}_{ij}}[X^\top(G(X,\zeta_{-,t}))^2]_{ji}$. Similarly, the order of $\partial^\ell_{\mathsf{B}_{ij}}\zeta_{-,t}$ is determined by the term that includes $\partial^{k-1}_{\mathsf{B}_{ij}}[X^\top(G(X,\zeta_{-,t}))^3]_{ji}$. This allows us to conclude that for any $k \geq 1$

$$\frac{\partial^k \lambda_{-,t}}{\partial \mathsf{B}^k_{ij}} \prec \frac{1}{Nt^{2k+1}}, \qquad \frac{\partial^k \zeta_{-,t}}{\partial \mathsf{B}^k_{ij}} \prec \frac{1}{Nt^{2k+1}}.$$

The claim now follows by noting that the above bounds still hold when we replace B_{ij} in X with some other $b \in [0, \mathsf{B}_{ij}]$. The reason behind this is that the replacement matrix still satisfies the η^*-regularity condition, ensuring that the corresponding $\zeta_{-,t}$ and λ_M still satisfy Lemma 3.2 (i).

Next, we prove a deterministic upper bound for $\partial_{\mathsf{B}_{ij}}\lambda_{-,t}$. For notational simplicity, we will only work on the original matrix X, and the argument holds for the replacement matrix $X_{(ij)}(b)$. In view of (91), it suffices to obtain deterministic upper bounds for $\lambda_{-,t}$, $\zeta_{-,t}$, and $[X^\top(G(X,\zeta_{-,t}))^2]_{ji}$. We may first apply Cauchy interlacing theorem to obtain an upper bound for $\zeta_{-,t}$ as follows:

$$\zeta_{-,t} \leq \lambda_M(\mathcal{S}(X)) \leq \lambda_{M-|\mathcal{D}_r|}(\mathcal{S}(\mathsf{B}^{(\mathcal{D}_r)})) \leq N^{2-2\epsilon_b}, \tag{95}$$

where in the last step we used the fact that the entries of $\mathcal{S}(\mathsf{B}^{(\mathcal{D}_r)})$ are bounded by $N^{-\epsilon_b}$. From (25), we have $c_N t m_X(\zeta_{-,t}) = c_N t m_t(\lambda_{-,t})/(1 + c_N t m_t(\lambda_{-,t}))$, which gives the deterministic bound $m_X(\zeta_{-,t}) \leq (c_N t)^{-1}$. Using this deterministic bound, we have that there exists some constant $C > 0$ such that

$$\frac{1}{M} \leq \frac{1}{M}\sum_{i=1}^{M}\frac{\lambda_M(\mathcal{S}(X)) - \zeta_{-,t}}{\lambda_i(\mathcal{S}(X)) - \zeta_{-,t}} \leq Ct^{-1}(\lambda_M(\mathcal{S}(X)) - \zeta_{-,t}).$$

This together with the fact that $\lambda_M(\mathcal{S}(X)) \geq \zeta_{-,t}$ (cf. Lemma 3.2 (i)) gives $\lambda_M(\mathcal{S}(X)) - \zeta_{-,t} \geq C^{-1}t/M$. Therefore, we are able to obtain deterministic bounds for the high order derivatives $m_X^{(k)}(\zeta_{-,t})$ as well as the spectral norm of $G(X, \zeta_{-,t})$. We can also obtain that

$$|\lambda_{-,t}| = \left|\left[1 - c_N t m_X(\zeta_{-,t})\right]^2\zeta_{-,t} + (1 - c_N)t\left[1 - c_N t m_X(\zeta_{-,t})\right]\right| \lesssim N^{2-2\epsilon_b}.$$

For the upper bound of $|[X^\top(G(X,\zeta_{-,t}))^2]_{ji}|$, we have

$$|[X^\top(G(X,\zeta_{-,t}))^2]_{ji}| \leq \left|[X^\top(G(X,\zeta_{-,t}))^2 X]_{jj}\right|^{1/2} \cdot \left|[(G(X,\zeta_{-,t}))^2]_{ii}\right|^{1/2}$$

$$\leq \|(G(X,\zeta_{-,t}))^2 XX^\top\|^{1/2} \cdot \|(G(X,\zeta_{-,t}))^2\|^{1/2}$$

$$\leq \left(\|G(X,\zeta_{-,t})\|^{1/2} + |\zeta_{-,t}|\|G(X,\zeta_{-,t})\|\right) \cdot \|G(X,\zeta_{-,t})\| \lesssim N^{2-2\epsilon_b}M^2t^{-2}.$$

Collecting the above bounds proves the first bound in (89).

To prove the second bound in (89), it suffices to provide a lower bound for $\frac{\partial^2 \Phi_t}{\partial \zeta^2}(\zeta_{-,t}, X)$ (cf. (93)). When $\lambda_{-,t} \geq \lambda_-^{\mathsf{mp}}/100$, we have

$$|c_N tm_X(\zeta_{-,t})| = \left|\frac{c_N tm_t(\lambda_{-,t})}{1 + c_N tm_t(\lambda_{-,t})}\right| \leq |c_N tm_t(\lambda_{-,t})| \lesssim \frac{t^{1/2}}{|\lambda_{-t}|} \lesssim t^{1/2}.$$

Therefore, using Cauchy-Schwarz inequality, we have

$$(c_N tm'_X(\zeta_{-,t}))^2 \leq \frac{c_N tm_X(\zeta_{-,t}) \cdot c_N tm''_X(\zeta_{-,t})}{2} \ll c_N tm''_X(\zeta_{-,t}).$$

This implies that $-2c_N tm''_X(\zeta_{-,t})\zeta_{-,t}(1-c_N tm_X(\zeta_{-,t})) + 2\zeta_{-,t}(c_N tm'_X(\zeta_{-,t}))^2 < 0$. Then using (94), we may lower bound $\frac{\partial^2 \Phi_t}{\partial \zeta^2}(\zeta_{-,t}, X)$ as follows:

$$\left|\frac{\partial^2 \Phi_t}{\partial \zeta^2}(\zeta_{-,t}, X)\right| > c_N(1-c_N)t^2 m''_X(\zeta_{-,t}) \geq \frac{2c_N(1-c_N)t^2}{M(\lambda_M(\mathcal{S}(X)) - \zeta_{-,t})^3} \geq \frac{2c_N(1-c_N)t^2}{MN^{6-6\epsilon_b}},$$

where in the last step we used (95). $\qquad\square$

Next, we start the proof of Theorem 4.3.

Proof of Theorem 4.3. We begin by collecting some notation to simplify the presentation of the proof. Consider $\tilde{w}_{ij}$ as the (i,j)-entry of $\tilde{W}$, and define $\tilde{Y}^\gamma$ analogously to Y^γ, with the substitution of W by $\tilde{W}$. Recall (72) and we write $d_{ij} = d_{ij}(\gamma, w_{ij})$, $e_{ij} = e_{ij}(\gamma, w_{ij})$, $\tilde{d}_{ij} = d_{ij}(0, \tilde{w}_{ij})$, and $\tilde{e}_{ij} = \tilde{e}_{ij}(0, \tilde{w}_{ij})$ in the sequel. To emphasize that $\lambda_{-,t}$ is a function of X, we introduce the notation $\lambda_{-,t}^{(ij)}(\beta) = \lambda_{-,t}(X_{(ij)}^\beta)$. Consequently, we define $z_t^{(ij)}(\beta) = \lambda_{-,t}^{(ij)}(\beta) + E + i\eta$. For simplicity, we use the shorthand notation $G_{(ij)}^{\gamma,\lambda,\beta}$ as $G_{(ij)}^{\gamma,\lambda}(z_t^{(ij)}(\beta))$, and we define $\tilde{G}_{(ij)}^{\gamma,\lambda,\beta}$ analogously, replacing W with $\tilde{W}$.

We will focus on the estimation of $\frac{\partial \mathbb{E}_\Psi(|N\eta(\operatorname{Im}m^\gamma(z_t) - \operatorname{Im}\tilde{m}^0(z_t))|^{2p})}{\partial \gamma}$. To this end, let us define $f_{\gamma,(ab),(ij)}(\lambda, \beta) = \operatorname{Im}[G_{(ij)}^{\gamma,\lambda,\beta}]_{ab}$, $\tilde{f}_{\gamma,(ab),(ij)}(\lambda,\beta) = \operatorname{Im}[\tilde{G}_{(ij)}^{\gamma,\lambda,\beta}]_{ab}$, $g_{(ij)}(\lambda,\beta) = \eta\operatorname{Im}[(G_{(ij)}^{\gamma,\lambda,\beta})^2 Y_{(ij)}^{\gamma,\lambda,\beta}]_{ij}$, and $F_p(\lambda, \tilde{\lambda}, \beta) = (\eta\sum_a f_{\gamma,(aa),(ij)}(\lambda,\beta) - \eta\sum_a \tilde{f}_{0,(aa),(ij)}(\tilde{\lambda},\beta))^p$. Some elementary calculation gives

$$\frac{\partial \mathbb{E}_\Psi\left(|N\eta(\operatorname{Im}m^\gamma(z_t) - \operatorname{Im}\tilde{m}^0(z_t))|^{2p}\right)}{\partial \gamma} = -2p\sum_{i,j}\left((J_1)_{ij} + (J_2)_{ij}\right),$$

where

$$(J_1)_{ij} = \mathbb{E}_\Psi\left[g_{(ij)}(d_{ij}, \chi_{ij}b_{ij})\mathcal{E}_{ij}F_{2p-1}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})\right] \cdot \mathbf{1}_{\psi_{ij}=0},$$

$$(J_2)_{ij} = -\frac{\gamma t^{1/2}}{(1-\gamma^2)^{1/2}}\mathbb{E}_\Psi\left[w_{ij}g_{(ij)}(e_{ij}, c_{ij})F_{2p-1}(e_{ij}, \tilde{e}_{ij}, c_{ij})\right] \cdot \mathbf{1}_{\psi_{ij}=1},$$

$$\mathcal{E}_{ij} = (1-\chi_{ij})a_{ij} - \gamma t^{1/2}(1-\gamma^2)^{-1/2}w_{ij}.$$

For $(J_2)_{ij}$, we may apply Gaussian integration by parts to obtain that

$$(J_2)_{ij} = -\frac{\gamma t^{1/2}}{(1-\gamma^2)^{1/2}N}\Big(\mathbb{E}_\Psi\left[\partial_{w_{ij}}\{g_{(ij)}(e_{ij}, c_{ij})\}F_{2p-1}(e_{ij}, \tilde{e}_{ij}, c_{ij})\right]$$
$$+ (2p-1)\mathbb{E}_\Psi\left[g_{(ij)}(e_{ij}, c_{ij})\partial_{w_{ij}}\{\eta\sum_a f_{\gamma,(aa),(ij)}(e_{ij}, c_{ij})\}F_{2p-2}(e_{ij}, \tilde{e}_{ij}, c_{ij})\right]\Big) \cdot \mathbf{1}_{\psi_{ij}=1}.$$

36

Note by directly calculation, we have

$$\partial_{w_{ij}}\{g_{(ij)}(e_{ij},c_{ij})\} = \Big(\mathrm{Im}\,\big[(G_{(ij)}^{\gamma,e_{ij},c_{ij}})^2\big]_{ii} - 2\mathrm{Im}\,\big(\big[(G_{(ij)}^{\gamma,e_{ij},c_{ij}})^2\big]_{ii}\big[(Y_{(ij)}^{\gamma,e_{ij}})^\top G_{(ij)}^{\gamma,e_{ij},c_{ij}}Y_{(ij)}^{\gamma,e_{ij}}\big]_{jj}\big)$$
$$- 2\mathrm{Im}\,\big(\big[(G_{(ij)}^{\gamma,e_{ij},c_{ij}})^2 Y_{(ij)}^{\gamma,e_{ij}}\big]_{ij}\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}Y_{(ij)}^{\gamma,e_{ij}}\big]_{ij}\big)\Big)(1-\gamma^2)^{1/2}t^{1/2}\eta, \qquad (96)$$

and $\partial_{w_{ij}}\{\eta\sum_a f_{\gamma,(aa),(ij)}(e_{ij},c_{ij})\} = -2t^{1/2}(1-\gamma^2)^{1/2}g_{(ij)}(e_{ij},c_{ij})$. Using Wald's identity with the fact that $i\in\mathcal{D}_r$ and $j\in\mathcal{D}_c$ when $\psi_{ij}=1$, we can obtain that

$$\big|\big[(G_{(ij)}^{\gamma,e_{ij},c_{ij}})^2\big]_{ii}\big| \le \sum_a \big|\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}\big]_{ai}\big|^2 = \frac{\mathrm{Im}\,\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}\big]_{ii}}{\eta} \prec t^{-2}\eta^{-1}, \qquad (97)$$

and

$$\big|\big[(G_{(ij)}^{\gamma,e_{ij},c_{ij}})^2 Y_{(ij)}^{\gamma,e_{ij}}\big]_{ij}\big| \le \sum_a \big|\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}Y_{(ij)}^{\gamma,e_{ij}}\big]_{aj}\big|^2 + \sum_a \big|\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}\big]_{ia}\big|^2$$
$$\overset{(i)}{=} \big[(Y_{(ij)}^{\gamma,e_{ij}})^\top|G_{(ij)}^{\gamma,e_{ij},c_{ij}}|^2 Y_{(ij)}^{\gamma,e_{ij}}\big]_{jj} + \eta^{-1}\mathrm{Im}\,\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}\big]_{ii}$$
$$\overset{(ii)}{=} \big[(Y_{(ij)}^{\gamma,e_{ij}})^\top Y_{(ij)}^{\gamma,e_{ij}}|\mathcal{G}_{(ij)}^{\gamma,e_{ij},c_{ij}}|^2\big]_{jj} + \eta^{-1}\mathrm{Im}\,\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}\big]_{ii}$$
$$= \big[\bar{\mathcal{G}}_{(ij)}^{\gamma,e_{ij},c_{ij}}\big]_{jj} + z\big[|\mathcal{G}_{(ij)}^{\gamma,e_{ij},c_{ij}}|^2\big]_{jj} + \eta^{-1}\mathrm{Im}\,\big[G_{(ij)}^{\gamma,e_{ij},c_{ij}}\big]_{ii} \overset{(iii)}{\prec} t^{-2}\eta^{-1}, \qquad (98)$$

where in (i) we applied Wald's identity, in (ii) we used the fact that for any $A\in\mathbb{R}^{M\times N}$, $(AA^\top - z)^{-1}A = A^\top(AA^\top - z)^{-1}$ when z does not lie inside the spectrum of A, and in (iii) we estimate $\big[|\mathcal{G}_{(ij)}^{\gamma,e_{ij},c_{ij}}|^2\big]_{jj}$ in a similar way as done in (97).

Combining the above estimates with the fact that $\sum_{i,j}\mathbf{1}_{\psi_{ij}=1} \le N^{1-\epsilon_\alpha}$, we arrive at

$$|(J_2)_{ij}| \lesssim \frac{\gamma}{(1-\gamma^2)^{1/2}N^{1-\epsilon_\alpha}}\sum_{k=1}^2 \mathbb{E}_\Psi\Big[|\mathcal{O}_\prec(N^{-\epsilon_\alpha}t^{-3})|\cdot|F_{2p-k}(e_{ij},\tilde{e}_{ij},c_{ij})|\Big]\cdot\mathbf{1}_{\psi_{ij}=1}.$$

We may then apply Young's inequality as the following:

$$\mathbb{E}_\Psi\Big[|\mathcal{O}_\prec(N^{-\epsilon_\alpha}t^{-3})|\cdot|F_{2p-1}(e_{ij},\tilde{e}_{ij},c_{ij})|\Big] \overset{(*)}{=} \mathbb{E}_\Psi\Big[|\mathcal{O}_\prec(N^{-\epsilon_\alpha}t^{-3})|\cdot\frac{|F_{2p-1}(e_{ij},\tilde{e}_{ij},c_{ij})|}{\log N}\Big]$$
$$\overset{(**)}{\lesssim} (\log N)^{\frac{2p}{1-2p}}\mathbb{E}_\Psi\Big[F_{2p}(e_{ij},\tilde{e}_{ij},c_{ij})\Big] + N^{-\epsilon_\alpha p}, \qquad (99)$$

where in $(*)$ we used the definition of stochastic domination, and in $(**)$ we used the fact that $t \gg N^{-\epsilon_\alpha/6}$. Similar argument can be applied to the second term involving F_{2p-2}, and therefore,

$$|(J_2)_{ij}| \lesssim \frac{\gamma}{(1-\gamma^2)^{1/2}N^{1-\epsilon_\alpha}}\Big((\log N)^{\frac{2p}{1-2p}}\mathbb{E}_\Psi\Big[F_{2p}(e_{ij},\tilde{e}_{ij},c_{ij})\Big] + N^{-\epsilon_\alpha p/2}\Big)\cdot\mathbf{1}_{\psi_{ij}=1}.$$

Next, we consider $(J_1)_{ij}$. Observe that by repeatedly taking derivatives w.r.t. d_{ij}, it can be easily seen that $\partial_{d_{ij}}^k\{g_{(ij)}(d_{ij},\chi_{ij}b_{ij})\}/\eta$ can be expressed as a linear combination of the imaginary parts of $\mathfrak{A}(a_1,a_2,a_3)\cdot\mathfrak{B}$, where $\mathfrak{A}(a_1,a_2,a_3) = \big(\big[G_{(ij)}^{\gamma,d_{ij},\chi_{ij}b_{ij}}Y_{(ij)}^{\gamma,d_{ij}}\big]_{ij}\big)^{a_1}\cdot\big(\big[G_{(ij)}^{\gamma,d_{ij},\chi_{ij}b_{ij}}\big]_{ii}\big)^{a_2}\cdot\big(\big[(Y_{(ij)}^{\gamma,d_{ij}})^\top G_{(ij)}^{\gamma,d_{ij},\chi_{ij}b_{ij}}Y_{(ij)}^{\gamma,d_{ij}}\big]_{jj}\big)^{a_3}$ for any integer $a_1,a_2,a_3 \ge 0$, and $\mathfrak{B} \in \{\big[(Y_{(ij)}^{\gamma,d_{ij}})^\top(G_{(ij)}^{\gamma,d_{ij},\chi_{ij}b_{ij}})^2 Y_{(ij)}^{\gamma,d_{ij}}\big]_{jj},\big[(G_{(ij)}^{\gamma,d_{ij},\chi_{ij}b_{ij}})^2 Y_{(ij)}^{\gamma,d_{ij}}\big]_{ij},\big[(G_{(ij)}^{\gamma,d_{ij},\chi_{ij}b_{ij}})^2\big]_{ii}\}$. The same holds for $\partial_{d_{ij}}^k\{\sum_a f_{\gamma,(aa),(ij)}(d_{ij},\chi_{ij}b_{ij})\}$ since $\partial_{d_{ij}}\{\eta\sum_a f_{\gamma,(aa),(ij)}(d_{ij},\chi_{ij}b_{ij})\} = -2g_{(ij)}(d_{ij},\chi_{ij}b_{ij})$. This together with a similar argument as (97) and (98) implies that

$$|\mathfrak{A}(a_1,a_2,a_3)|\cdot\mathbf{1}_{\psi_{ij}=0} \prec t^{-2(a_1+a_2+a_3)}, \quad\text{and}\quad |\mathfrak{B}|\cdot\mathbf{1}_{\psi_{ij}=0} \prec t^{-2}\eta^{-1}. \qquad (100)$$

The estimation of $(J_1)_{ij}$ relies on a careful analysis of expansion. Here, we introduce $g_{(ij)}^{(k_1,k_2)}$ as the mixed (k_1,k_2)-th order derivative of $g_{(ij)}(\lambda, \beta)$ w.r.t. λ and β, and $F_{2p-1}^{(k_1,k_2,k_3)}$ represents the mixed (k_1, k_2, k_3)-th order derivative of $F_{2p-1}(\lambda, \tilde{\lambda}, \beta)$ w.r.t. λ, $\tilde{\lambda}$, and β. Applying Taylor expansion on $g_{(ij)}(d_{ij}, \chi_{ij}b_{ij})$ on the first variable around 0, we have for an s_1 to be chosen later, there exists $\bar{d}_{ij} \in [0, d_{ij}]$ such that,

$$
(J_1)_{ij} = \sum_{k=0}^{s_1} \mathbb{E}_\Psi \left[\frac{d_{ij}^k g_{(ij)}^{(k,0)}(0, \chi_{ij}b_{ij})}{k!} \mathcal{E}_{ij} F_{2p-1}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij}) \right] \cdot \mathbf{1}_{\psi_{ij}=0}
$$
$$
+ \mathbb{E}_\Psi \left[\frac{d_{ij}^{s_1+1} g_{(ij)}^{(s_1+1,0)}(\bar{d}_{ij}, \chi_{ij}b_{ij})}{(s_1+1)!} \mathcal{E}_{ij} F_{2p-1}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij}) \right] \cdot \mathbf{1}_{\psi_{ij}=0} = \sum_{k=0}^{s_1} \frac{1}{k!}(J_1)_{ij,k} + \mathsf{Rem}_1.
$$

By the entries bound in Proposition 4.1, (97), (98), and the perturbation argument in (80), we may crudely bound the above remainder term as follows:

$$
|\mathsf{Rem}_1| \lesssim \frac{\mathbf{1}_{\psi_{ij}=0}}{N^{(s_1+2)\epsilon_b}} \mathbb{E}_\Psi \left[|g_{(ij)}^{(s_1+1,0)}(\bar{d}_{ij}, \chi_{ij}b_{ij})| \cdot |F_{2p-1}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})| \right] \lesssim \frac{N^\epsilon N^{2p}}{N^{(s_1+2)\epsilon_b} t^{2s_1+4}} \lesssim N^{-p},
$$

where in the second inequality, we used the deterministic bound $|F_{2p-1}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})| \leq N^{2p-1}$, and in the last step, we chose $s_1 > 6p/\epsilon_b$ and used the fact that $t \gg N^{-\epsilon_b/8}$. We may apply similar argument to expand the first two variables of $F_{2p-1}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})$ in $(J_1)_{ij,k}$ to obtain that

$$
(J_1)_{ij,k} = \sum_{\ell=0}^{s_2} \sum_{m=0}^{\ell} \mathbb{E}_\Psi \left[\frac{d_{ij}^{k+m} \tilde{d}_{ij}^{\ell-m} g_{(ij)}^{(k,0)}(0, \chi_{ij}b_{ij})}{m!(\ell-m)!} \mathcal{E}_{ij} F_{2p-1}^{(m,\ell-m,0)}(0, 0, \chi_{ij}b_{ij}) \right] \cdot \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-p})
$$
$$
= \sum_{\ell=0}^{s_2} (J_1)_{ij,k\ell} + \mathcal{O}(N^{-p}).
$$

where s_2 is a large integer satisfying $s_2 > 6p/\epsilon_b$. To estimate $(J_1)_{ij,k\ell}$, we start by introducing the notation $t_{ij} = t^2 \cdot (1 - \mathbf{1}_{i \in \mathcal{T}_r, j \in \mathcal{T}_c}) + \mathbf{1}_{i \in \mathcal{T}_r, j \in \mathcal{T}_c}$ for presentation simplicity. Note by the chain rule, we have for any integer $\ell \geq 0$ and $m \leq \ell$,

$$
F_{2p-1}^{m,\ell-m,0}(0, 0, \chi_{ij}b_{ij}) = \sum_{k=1}^{\ell \wedge (2p-1)} \mathcal{C}_{k,m}^{\chi_{ij}b_{ij}} F_{2p-1-k}(0, 0, \chi_{ij}b_{ij}) + \mathcal{C}_{\ell+1,m}^{\chi_{ij}b_{ij}} \mathbf{1}_{\ell \geq (2p-1)}, \tag{101}
$$

where for all $k \in [\ell+1]$, $m \in [\ell]$, $\mathcal{C}_{k,m}^{\chi_{ij}b_{ij}}$ are polynomials of the following terms

$$
[G_{(ij)}^{\gamma,0,\chi_{ij}b_{ij}} Y_{(ij)}^{\gamma,0}]_{ij}, [G_{(ij)}^{\gamma,0,\chi_{ij}b_{ij}}]_{ii}, [(Y_{(ij)}^{\gamma,0})^\top G_{(ij)}^{\gamma,0,\chi_{ij}b_{ij}} Y_{(ij)}^{\gamma,0}]_{jj}, [(G_{(ij)}^{\gamma,0,\chi_{ij}b_{ij}})^2 Y_{(ij)}^{\gamma,0}]_{ij}, [(G_{(ij)}^{\gamma,0,\chi_{ij}b_{ij}})^2]_{ii},
$$
$$
[\tilde{G}_{(ij)}^{0,0,\chi_{ij}b_{ij}} \tilde{Y}_{(ij)}^{0,0}]_{ij}, [\tilde{G}_{(ij)}^{0,0,\chi_{ij}b_{ij}}]_{ii}, [(\tilde{Y}_{(ij)}^{0,0})^\top \tilde{G}_{(ij)}^{0,0,\chi_{ij}b_{ij}} \tilde{Y}_{(ij)}^{0,0}]_{jj}, [(\tilde{G}_{(ij)}^{0,0,\chi_{ij}b_{ij}})^2 \tilde{Y}_{(ij)}^{0,0}]_{ij}, [(\tilde{G}_{(ij)}^{0,0,\chi_{ij}b_{ij}})^2]_{ii}.
$$

After carrying out a similar derivation as shown in (96)-(98) and employing the perturbation argument described in (80), it can be easily verified that $\mathcal{C}_{k,m}^{\chi_{ij}b_{ij}} \cdot \mathbf{1}_{\psi_{ij}=0} \prec t_{ij}^{-(\ell+1)}$.

Plugging (101) into $(J_1)_{ij,k\ell}$, we have

$$
(J_1)_{ij,k\ell} = \sum_{n=1}^{\ell \wedge (2p-1)} \sum_{m=0}^{\ell} \mathbb{E}_\Psi \left[\frac{d_{ij}^{k+m} \tilde{d}_{ij}^{\ell-m} \mathcal{E}_{ij}}{m!(\ell-m)!} g_{(ij)}^{(k,0)}(0, \chi_{ij}b_{ij}) \mathcal{C}_{n,m}^{\chi_{ij}b_{ij}} F_{2p-1-n}(0, 0, \chi_{ij}b_{ij}) \right] \cdot \mathbf{1}_{\psi_{ij}=0}
$$
$$
+ \sum_{m=0}^{\ell} \mathbb{E}_\Psi \left[\frac{d_{ij}^{k+m} \tilde{d}_{ij}^{\ell-m} \mathcal{E}_{ij}}{m!(\ell-m)!} g_{(ij)}^{(k,0)}(0, \chi_{ij}b_{ij}) \mathcal{C}_{\ell+1,m}^{\chi_{ij}b_{ij}} \mathbf{1}_{\ell \geq (2p-1)} \right] \cdot \mathbf{1}_{\psi_{ij}=0} = (\mathsf{T}_1)_{ij,k\ell} + (\mathsf{T}_2)_{ij,k\ell}.
$$

For $(\mathsf{T}_2)_{ij,k\ell}$, we only need to consider the case when $\ell \geq 2p-1$. Using $g^{(k,0)}_{(ij)}(0, \chi_{ij}b_{ij}) \prec t_{ij}^{-(k+1)}$ with the fact that $t \gg N^{-\epsilon_b/8}$, we can conclude that $|(\mathsf{T}_2)_{ij,k\ell}| \lesssim N^{-\epsilon_b p}$. Next, we focus on the estimation of $(\mathsf{T}_1)_{ij,k\ell}$.

When $k + \ell$ is even, we have by the law of total expectation that,

$$
(\mathsf{T}_1)_{ij,k\ell} = \sum_{n=1}^{\ell \wedge (2p-1)} \sum_{m=0}^{\ell} \mathbb{E}_{\Psi} \Big[\frac{(\gamma a_{ij} + (1-\gamma^2)^{1/2} t^{1/2} w_{ij})^{k+m} (t^{1/2} \tilde{w}_{ij})^{\ell-m}}{m!(\ell-m)!} g^{(k,0)}_{(ij)}(0,0)
$$
$$
\times \Big(a_{ij} - \frac{\gamma t^{1/2} w_{ij}}{(1-\gamma^2)^{1/2}} \Big) \mathcal{C}^0_{n,m} F_{2p-1-n}(0,0,0) \Big] \cdot \mathbb{P}(\chi_{ij} = 0) \cdot \mathbf{1}_{\psi_{ij}=0}
$$
$$
- \sum_{n=1}^{\ell \wedge (2p-1)} \sum_{m=0}^{\ell} \mathbb{E}_{\Psi} \Big[\frac{\gamma (b_{ij} + (1-\gamma^2)^{1/2} t^{1/2} w_{ij})^{k+m} (b_{ij} + t^{1/2} \tilde{w}_{ij})^{\ell-m} t^{1/2} w_{ij}}{(1-\gamma^2)^{1/2} m!(\ell-m)!}
$$
$$
\times g^{(k,0)}_{(ij)}(0, b_{ij}) \mathcal{C}^{b_{ij}}_{n,m} F_{2p-1-n}(0,0,b_{ij}) \Big] \cdot \mathbb{P}(\chi_{ij} = 1) \cdot \mathbf{1}_{\psi_{ij}=0}. \tag{102}
$$

From the above equation, one can easily verify that $(\mathsf{T}_1)_{ij,k\ell} = 0$ when $k + \ell$ is even. Therefore, in the rest of the estimation, we consider the case of $k + \ell$ is odd. In this case, we need to further expand out $\chi_{ij}b_{ij}$ in $\mathcal{C}^{\chi_{ij}b_{ij}}_{n,m}$, $g^{(k,0)}_{(ij)}(0, \chi_{ij}b_{ij})$ and $F_{2p-1-n}(0,0,\chi_{ij}b_{ij})$.

First note by Taylor expansion, for any $s_3 \geq 0$ there exists $b^{(1)}_{ij} \in [0, \chi_{ij}b_{ij}]$ such that

$$
g^{(k,0)}_{(ij)}(0, \chi_{ij}b_{ij}) = \sum_{q=0}^{s_3} \frac{(\chi_{ij}b_{ij})^q}{q!} g^{(k,q)}_{(ij)}(0,0) + \frac{(\chi_{ij}b_{ij})^{s_3+1}}{(s_3+1)!} g^{(k,s_3+1)}_{(ij)}(0, b^{(1)}_{ij}). \tag{103}
$$

By Faà di Bruno's formula, for $q \geq 1$, $g^{(k,q)}_{(ij)}(\lambda, \beta)$ can be expressed as

$$
g^{(k,q)}_{(ij)}(\lambda, \beta) = \sum_{(u_1, \cdots, u_q)} \frac{q!}{u_1! u_2! \cdots u_q!} \partial_z^{u_1 + \cdots + u_q} g^{(k,0)}_{(ij)}(\lambda, \beta) \cdot \prod_{v=1}^{q} \Big(\frac{\partial^v \lambda^{(ij)}_{-,t}}{\partial \beta^v}(\beta) \Big)^{u_v}, \tag{104}
$$

where the sum $\sum_{(u_1, \cdots, u_q)}$ is over all q-tuples of nonnegative integers $(u_1, \cdots, u_q)$ satisfying $\sum_{i=1}^{q} i u_i = q$. We may then use (88) in Lemma 5.4 to bound the derivatives of $\lambda^{(ij)}_{-,t}$ and a Cauchy integral argument to bound the derivatives of $g^{(k,0)}_{(ij)}$ w.r.t z, which gives

$$
g^{(k,q)}_{(ij)}(0,0) \cdot \mathbf{1}_{\psi_{ij}=0} \prec \sum_{(u_1, \cdots, u_q)} \frac{1}{\eta^{u_1 + \cdots + u_q} t_{ij}^{k+1}} \prod_{v=1}^{q} \frac{1}{N^{u_v} t^{(2v+1)u_v}} \prec \frac{1}{N\eta t^{3q} t_{ij}^{k+1}}, \quad q \geq 1. \tag{105}
$$

and the same bound holds for $g^{(k,q)}_{(ij)}(0, b^{(1)}_{ij})$. Therefore, by choosing $s_3 > 6p/\epsilon_b$ together with the facts that $\mathcal{C}^{\chi_{ij}b_{ij}}_{n,m} \prec t_{ij}^{-(k+1)}$, $|F_{2p-1-n}(0,0,\chi_{ij}b_{ij})| \lesssim N^{2p-1-n}$, we can obtain that

$$
(\mathsf{T}_1)_{ij,k\ell} = \sum_{n=1}^{\ell \wedge (2p-1)} \sum_{q=0}^{s_3} \sum_{m=0}^{\ell} \mathbb{E}_{\Psi} \Big[\frac{d_{ij}^{k+m} \tilde{d}_{ij}^{\ell-m} (\chi_{ij}b_{ij})^q \mathcal{E}_{ij}}{q! m!(\ell-m)!} g^{(k,q)}_{(ij)}(0,0) \mathcal{C}^{\chi_{ij}b_{ij}}_{n,m}
$$
$$
\times F_{2p-1-n}(0,0,\chi_{ij}b_{ij}) \Big] \cdot \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\epsilon_b p}) = \sum_{n=1}^{\ell \wedge (2p-1)} \sum_{q=0}^{s_3} (\mathsf{T}_1)_{ij,k\ell,nq} + \mathcal{O}(N^{-\epsilon_b p}).
$$

For $(\mathsf{T}_1)_{ij,k\ell,nq}$, the term $\mathcal{C}_{n,m}^{\chi_{ij}b_{ij}}$ can be expanded in a similar way as done for $g_{(ij)}^{(k,0)}(0,\chi_{ij}b_{ij})$ in (103) and (104), we omit the details. This leads to

$$(\mathsf{T}_1)_{ij,k\ell,nq} = \sum_{r=0}^{s_4}\sum_{m=0}^{\ell} \mathbb{E}_\Psi\Big[\frac{d_{ij}^{k+m}\tilde{d}_{ij}^{\ell-m}(\chi_{ij}b_{ij})^{q+r}\mathcal{E}_{ij}}{r!q!m!(\ell-m)!}g_{(ij)}^{(k,q)}(0,0)$$
$$\times\,\mathcal{C}_{n,m}^{(r),0}F_{2p-1-n}(0,0,\chi_{ij}b_{ij})\Big]\cdot\mathbf{1}_{\psi_{ij}=0}+\mathcal{O}(N^{-\epsilon_b p}),$$

where $s_4 > 6p/\epsilon_b$ and

$$\mathcal{C}_{n,m}^{(r),0} = \frac{\partial^r \mathcal{C}_{n,m}^\beta}{\partial\beta^r}\Big|_{\beta=0}, \quad\text{and}\quad \mathcal{C}_{n,m}^{(r),0} \prec \frac{1}{N\eta t^{3r}t_{ij}^{\ell+1}}, \quad r\ge 1. \tag{106}$$

Next, we deal with $F_{2p-1-n}(0,0,\chi_{ij}b_{ij})$. For any $s\ge 0$, we can compute that

$$F_{2p-1-n}^{(0,0,s)}(0,0,0) = \sum_{(u_1,\cdots,u_s)}\frac{s!}{u_1!u_2!\cdots u_s!}\partial_z^{u_1+\cdots+u_s}F_{2p-1-n}(0,0,0)\cdot\prod_{\mathsf{w}=1}^{s}\Big(\frac{\partial^{\mathsf{w}}\lambda_{-,t}^{(ij)}}{\partial\beta^{\mathsf{w}}}(0)\Big)^{u_{\mathsf{w}}}, \tag{107}$$

and for any integer $\vartheta\ge 0$,

$$\partial_z^\vartheta F_{2p-1-n}(0,0,0) = \sum_{\substack{(v_1,\cdots,v_\vartheta)\\v_1+\cdots+v_\vartheta\le 2p-1-n}}\frac{\vartheta!}{u_1!v_2!\cdots v_\vartheta!}F_{2p-1-n-(v_1+\cdots+v_\vartheta)}(0,0,0)$$
$$\times\prod_{\mathsf{w}=1}^{\vartheta}\Big(\eta\mathrm{Im}\,\mathrm{Tr}\big(G_{(ij)}^{\gamma,0,0}\big)^{\mathsf{w}+1}-\eta\mathrm{Im}\,\mathrm{Tr}\big(\tilde{G}_{(ij)}^{0,0,0}\big)^{\mathsf{w}+1}\Big)^{v_{\mathsf{w}}}. \tag{108}$$

Combining the above two expression, and using Lemma 5.4, we can estimate the remainder term as done for Rem_3, which gives

$$(\mathsf{T}_1)_{ij,k\ell,nq} = \sum_{r=0}^{s_4}\sum_{s=0}^{s_5}\sum_{m=0}^{\ell}\mathbb{E}_\Psi\Big[\frac{d_{ij}^{k+m}\tilde{d}_{ij}^{\ell-m}(\chi_{ij}b_{ij})^{q+r+s}\mathcal{E}_{ij}}{s!r!q!m!(\ell-m)!}\Big]$$
$$\times\,\mathbb{E}_\Psi\Big[g_{(ij)}^{(k,q)}(0,0)\mathcal{C}_{n,m}^{0,(r)}F_{2p-1-n}^{(0,0,s)}(0,0,0)\Big]\cdot\mathbf{1}_{\psi_{ij}=0}+\mathcal{O}(N^{-\epsilon_b p}), \tag{109}$$

for some large enough integer s_5. Here we also used the independency between the random variables. Then it suffices to estimate $(\mathsf{T}_1)_{ij,k\ell,nq}$ in two different cases, $k+\ell=1$ and $k+\ell\ge 3$ (recall that we only need to consider the case when $k+\ell$ is odd, cf. (102)).

    Case 1: $k+\ell\ge 3$. From (109), using the estimates (105) and (106), and the fact that $\mathbb{E}(b_{ij}^2)\lesssim N^{-1}$, $\mathbb{E}((1-\chi_{ij})a_{ij}^2)\asymp t\mathbb{E}(w_{ij}^2)=t/N$, we have

$$(\mathsf{T}_1)_{ij,k\ell,nq} = \sum_{r=0}^{s_4}\sum_{s=0}^{s_5}\mathcal{O}\Big(\frac{t}{N^{2+(k+\ell+q+r+s-3)\epsilon_b}}\Big)$$
$$\times\,\mathbb{E}_\Psi\Big[\mathcal{O}_\prec\Big(\frac{1}{t^{3(q+r)}t_{ij}^{k+\ell+2}}\Big)F_{2p-1-n}^{(0,0,s)}(0,0,0)\Big]\cdot\mathbf{1}_{\psi_{ij}=0}+\mathcal{O}(N^{-\epsilon_b p}). \tag{110}$$

Note that we have already derived the expression of $F_{2p-1-n}^{(0,0,s)}(0,0,0)$ in (107) and (108). Then using the following inequality:

$$\big|\eta\mathrm{Im}\,\mathrm{Tr}\big(G_{(ij)}^{\gamma,0,0}\big)^{\mathsf{w}+1}-\eta\mathrm{Im}\,\mathrm{Tr}\big(\tilde{G}_{(ij)}^{0,0,0}\big)^{\mathsf{w}+1}\big|^{v_{\mathsf{w}}}\lesssim \big|\eta\mathrm{Im}\,\mathrm{Tr}\big(G_{(ij)}^{\gamma,0,0}\big)^{\mathsf{w}+1}\big|^{v_{\mathsf{w}}}+\big|\eta\mathrm{Im}\,\mathrm{Tr}\big(\tilde{G}_{(ij)}^{0,0,0}\big)^{\mathsf{w}+1}\big|^{v_{\mathsf{w}}}$$
$$\le \eta^{-\mathsf{w}v_{\mathsf{w}}}\big(\big|\eta\mathrm{Im}\,\mathrm{Tr}G_{(ij)}^{\gamma,0,0}\big|^{v_{\mathsf{w}}}+\big|\eta\mathrm{Im}\,\mathrm{Tr}\tilde{G}_{(ij)}^{0,0,0}\big|^{v_{\mathsf{w}}}\big)\lesssim \eta^{-\mathsf{w}v_{\mathsf{w}}}\big(|F_{v_{\mathsf{w}}}(0,0,0)|+\big|\eta\mathrm{Im}\,\mathrm{Tr}\tilde{G}_{(ij)}^{0,0,0}\big|^{v_{\mathsf{w}}}\big), \tag{111}$$

together with Lemma 5.4 and the fact that $\eta \operatorname{Im} \operatorname{Tr} \tilde{G}_{(ij)}^{0,0,0} \prec N\eta\sqrt{|E|+\eta} \leq N^{1-\varepsilon_1/2}\eta$ (this can be done by bounding $\left(\eta \operatorname{Im} \operatorname{Tr} \tilde{G}_{(ij)}^{0,0,0} - \eta \operatorname{Im} \operatorname{Tr} \tilde{G}_{(ij)}^{0,d_{ij},\chi_{ij}b_{ij}}\right) \cdot \mathbf{1}_{\psi_{ij}=0}$ through Taylor expansion and then using local law for the Gaussian divisible model (cf. (18)) that $\left(\eta \operatorname{Im} \operatorname{Tr} \tilde{G}_{(ij)}^{0,d_{ij},\chi_{ij}b_{ij}} - N\eta \operatorname{Im} m_t(z_t)\right) \cdot \mathbf{1}_{\psi_{ij}=0} \prec 1$ with $\operatorname{Im} m_t(z_t) \prec \sqrt{|E|+\eta}$ (cf. (33))), we can obtain that

$$|(\mathsf{T}_1)_{ij,k\ell,nq}| \leq \frac{1}{N^2} \sum_{\mathsf{a}=0}^{2p-1-n} \mathbb{E}_\Psi \left[\mathcal{O}_\prec\left(\frac{t}{N^{(k+\ell+\mathsf{a}-3)\epsilon_b} t^{3\mathsf{a}} t_{ij}^{(k+\ell+2)}} \right) |F_{2p-1-n-\mathsf{a}}(0,0,0)| \right] \cdot \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\epsilon_b p}).$$

Substituting this back into $(\mathsf{T}_1)_{ij,k\ell}$ and considering that $t \gg N^{-\epsilon_b/100} \vee N^{-\epsilon_d/20}$, a straightforward calculation yields that: if $k+\ell \geq 5$,

$$|(\mathsf{T}_1)_{ij,k\ell}| \leq \frac{1}{N^2} \sum_{n=1}^{2p-1} \mathbb{E}_\Psi \left[\mathcal{O}_\prec\left(\frac{1}{N^{(n+1)\epsilon_b/10}} \right) |F_{2p-1-n}(0,0,0)| \right] \cdot \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\epsilon_b p}), \quad (112)$$

and if $k+\ell \geq 3$,

$$|(\mathsf{T}_1)_{ij,k\ell}| \leq \sum_{n=1}^{\ell \wedge (2p-1)} \sum_{\mathsf{a}=0}^{2p-1-n} \left(\frac{\mathbf{1}_{\psi_{ij}=0}(1-\mathbf{1}_{i\in\mathcal{T}_r,j\in\mathcal{T}_c})}{N^{2-\epsilon_d}} \mathbb{E}_\Psi \left[\mathcal{O}_\prec\left(\frac{1}{N^{\mathsf{a}\epsilon_b/10+\epsilon_d/2}} \right) |F_{2p-1-n-\mathsf{a}}(0,0,0)| \right] \right.$$
$$\left. + \frac{\mathbf{1}_{\psi_{ij}=0}\mathbf{1}_{i\in\mathcal{T}_r,j\in\mathcal{T}_c}}{N^2} \mathbb{E}_\Psi \left[\mathcal{O}_\prec\left(\frac{t}{N^{\mathsf{a}\epsilon_b/10}} \right) |F_{2p-1-n-\mathsf{a}}(0,0,0)| \right] \right) + \mathcal{O}(N^{-\epsilon_b p}). \quad (113)$$

Next, we shall replace $F_{2p-1-n}(0,0,0) \cdot \mathbf{1}_{\psi_{ij}=0}$ back by $F_{2p-1-n}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij}) \cdot \mathbf{1}_{\psi_{ij}=0}$. Applying Taylor expansion on the third variable and then using (107)-(111), we can obtain that

$$|F_{2p-1-n}(0,0,0)| \leq \sum_{\mathsf{a}=0}^{2p-1-n} \mathcal{O}_\prec\left(N^{-\epsilon_b \mathsf{a}/10} \right) \cdot |F_{2p-1-n-\mathsf{a}}(0,0,\chi_{ij}b_{ij})| + \mathcal{O}_\prec(N^{-\epsilon_b p}).$$

Therefore, we have that (112) and (113) remain valid, with $(0,0,0)$ replaced by $(0,0,\chi_{ij}b_{ij})$. Using Taylor expansion again, for a large enough integer s_7, there exists $d_{1,ij} \in [0,d_{ij}], d_{2,ij} \in [0,\tilde{d}_{ij}]$ such that

$$F_{2p-1-n}(0,0,\chi_{ij}b_{ij}) = \sum_{u=0}^{s_7} \sum_{v=0}^{u} \frac{(-d_{ij})^v(-\tilde{d}_{ij})^{u-v}}{v!(u-v)!} \cdot F_{2p-1-n}^{(v,u-v,0)}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})$$
$$+ \sum_{\ell=0}^{s_7+1} \frac{(-d_{ij})^v(-\tilde{d}_{ij})^{s_7+1-v}}{v!(s_7+1-v)!} \cdot F_{2p-1-n}^{(v,s_7+1-v,0)}(d_{1,ij}, d_{2,ij}, \chi_{ij}b_{ij}).$$

Then we may use (101)(with minor modification that replace $(0,0,\chi_{ij}b_{ij})$ by $(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})$) to transform $F_{2p-1-n}^{(v,u-v,0)}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})$ to $F_{2p-1-r}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})$ for some $r \geq n$. It can also be easily checked that the resulting coefficients of F_{2p-1-r} can be compensated by bounding $|d_{ij}|, |\tilde{d}_{ij}|$ by $N^{-\epsilon_b}$ (w.h.p). This finally confirms that (112) and (113) still hold when $(0,0,0)$ are replaced by $(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij})$.

Therefore, using straightforward power counting and applying Young's inequality as shown in (99), we may conclude that when $k+\ell \geq 3$, there exits some constants $K = K(p) > 0$ and $\delta = \delta(\epsilon_a, \epsilon_b, \epsilon_d) > 0$, such that

$$|(\mathsf{J}_1)_{ij,k\ell}| \lesssim \frac{\mathbf{1}_{\psi_{ij}=0}}{N^2} \left((\log N)^{-K} \mathbb{E}_\Psi \left[F_{2p}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij}) \right] + N^{-\delta p} \right)$$
$$+ \frac{\mathbf{1}_{\psi_{ij}=0}(1-\mathbf{1}_{i\in\mathcal{T}_r,j\in\mathcal{T}_c})}{N^{2-\epsilon_d}} \left((\log N)^{-K} \mathbb{E}_\Psi \left[F_{2p}(d_{ij}, \tilde{d}_{ij}, \chi_{ij}b_{ij}) \right] + N^{-\delta p} \right). \quad (114)$$

Case 2: $k + \ell = 1$. Recall from (109) that

$$(\mathsf{T}_1)_{ij,k\ell,nq} = \sum_{r=0}^{s_4} \sum_{s=0}^{s_5} \sum_{m=0}^{\ell} \mathbb{E}_\Psi \left[\frac{d_{ij}^{k+m} \tilde{d}_{ij}^{\ell-m} (\chi_{ij} b_{ij})^{q+r+s} \mathcal{E}_{ij}}{s! r! q! m! (\ell-m)!} \right]$$
$$\times \mathbb{E}_\Psi \left[g_{(ij)}^{(k,q)}(0,0) \mathcal{C}_{n,m}^{0,(r)} F_{2p-1-n}^{(0,0,s)}(0,0,0) \right] \cdot \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\epsilon_b p}).$$

Case 2.1: $q + r + s$ is odd. In this case, we can directly compute that

$$\sum_{m=0}^{\ell} \mathbb{E}_\Psi \left[\frac{d_{ij}^{k+m} \tilde{d}_{ij}^{\ell-m} (\chi_{ij} b_{ij})^{q+r+s} \mathcal{E}_{ij}}{s! r! q! m! (\ell-m)!} \right] = \mathbb{E}_\Psi \left[\frac{\gamma((1-\chi_{ij})a_{ij}^2 - t w_{ij}^2)(\chi_{ij} b_{ij})^{q+r+s}}{s! r! q!} \right] = 0.$$

Thus, we have $(\mathsf{T}_1)_{ij,k\ell,nq} = \mathcal{O}(N^{-\epsilon_b p})$ in this case .

Case 2.2: $q+r+s \geq 0$ is even. Using (87) and the simple facts that $\chi_{ij}(1-\chi_{ij}) = 0$ and $\mathbb{E}(b_{ij}^2) \lesssim N^{-1}$, we have

$$\sum_{m=0}^{\ell} \mathbb{E}_\Psi \left[\frac{d_{ij}^{k+m} \tilde{d}_{ij}^{\ell-m} (\chi_{ij} b_{ij})^{q+r+s} \mathcal{E}_{ij}}{s! r! q! m! (\ell-m)!} \right] = \frac{-\gamma t^{1/2}}{(1-\gamma^2)^{1/2}} \mathbb{E}_\Psi \left[\frac{d_{ij} w_{ij} (\chi_{ij} b_{ij})^{q+r+s}}{s! r! q!} \right]$$
$$= \frac{t}{N^2} \left(\mathcal{O}\left(\frac{\mathbf{1}_{q+r+s \geq 2}}{N^{(q+r+s-2)\epsilon_b}} \right) + \mathcal{O}\left(\frac{\mathbf{1}_{q+r+s=0}}{N^{\epsilon_b}} \right) \right).$$

Further using (105) and (106), we can obtain that

$$(\mathsf{T}_1)_{ij,k\ell,nq} = \sum_{r=0}^{s_4} \sum_{s=0}^{s_5} \frac{t}{N^2} \left(\mathcal{O}\left(\frac{\mathbf{1}_{q+r+s \geq 2}}{N^{(q+r+s-2)\epsilon_b}} \right) + \mathcal{O}\left(\frac{\mathbf{1}_{q+r+s=0}}{N^{\epsilon_b}} \right) \right)$$
$$\times \mathbb{E}_\Psi \left[\mathcal{O}_\prec \left(\frac{1}{(N\eta \mathbf{1}_{q+r \geq 1} + \mathbf{1}_{q+r=0}) t^{3(q+r)} t_{ij}^3} \right) F_{2p-1-n}^{(0,0,s)}(0,0,0) \right] \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\epsilon_b p}).$$

Observing that the above equation has a similar form to (110), we may proceed in a similar manner as in Case 1 to estimate $(\mathsf{T}_1)_{ij,k\ell,nq}$. We will omit the repetitive details for brevity. Consequently, we can conclude that, by possibly adjusting the constants, (114) also holds when $k + \ell = 1$.

Combining Case 1, Case 2, and the estimates for $(J_2)_{ij}$'s, we arrive at

$$\sum_{i,j} |(I)_{ij}| \lesssim \frac{\mathbf{1}_{\psi_{ij}=1}}{N^{1-\epsilon_\alpha}} \sum_{i,j} \left((\log N)^{\frac{2p}{1-2p}} \mathbb{E}_\Psi \left[F_{2p}(e_{ij}, \tilde{e}_{ij}, c_{ij}) \right] + N^{-\epsilon_\alpha p/2} \right)$$
$$+ \sum_{i,j} \frac{\mathbf{1}_{\psi_{ij}=0}}{N^2} \left((\log N)^{-K} \mathbb{E}_\Psi \left[F_{2p}(d_{ij}, \tilde{d}_{ij}, \chi_{ij} b_{ij}) \right] + N^{-\delta p} \right)$$
$$+ \sum_{i,j} \frac{\mathbf{1}_{\psi_{ij}=0}(1 - \mathbf{1}_{i \in \mathcal{T}_r, j \in \mathcal{T}_c})}{N^{2-\epsilon_d}} \left((\log N)^{-K} \mathbb{E}_\Psi \left[F_{2p}(d_{ij}, \tilde{d}_{ij}, \chi_{ij} b_{ij}) \right] + N^{-\delta p} \right)$$
$$\lesssim (\log N)^{-(K \wedge \frac{2p}{2p-1})} \mathbb{E}_\Psi \left[\left| N\eta (\operatorname{Im} m^\gamma(z) - \operatorname{Im} \tilde{m}^0(z)) \right|^{2p} \right] + N^{-\tilde{\delta} p},$$

where $\tilde{\delta} = \tilde{\delta}(\epsilon_a, \epsilon_b, \epsilon_d) > 0$. Therefore, for any $0 \leq \gamma \leq 1$,

$$\mathbb{E}_\Psi \left(\left| N\eta (\operatorname{Im} m^\gamma(z_t) - \operatorname{Im} \tilde{m}^0(z_t)) \right|^{2p} \right) - \mathbb{E}_\Psi \left(\left| N\eta (\operatorname{Im} m^0(z_t) - \operatorname{Im} \tilde{m}^0(z_t)) \right|^{2p} \right)$$
$$= \int_0^\gamma \frac{\partial \mathbb{E} \left(\left| N\eta (\operatorname{Im} m^{\gamma'}(z_t) - \operatorname{Im} \tilde{m}^0(z_t)) \right|^{2p} \right)}{\partial \gamma'} d\gamma'. \tag{115}$$

Taking supremum over γ, and using the estimates above, we have

$$\sup_{0 \leq \gamma \leq 1} \mathbb{E}_\Psi \Big(\big| N\eta \big(\mathrm{Im}\, m^\gamma(z_t) - \mathrm{Im}\, \tilde{m}^0(z_t) \big) \big|^{2p} \Big) - \mathbb{E}_\Psi \Big(\big| N\eta \big(\mathrm{Im}\, m^0(z_t) - \mathrm{Im}\, \tilde{m}^0(z_t) \big) \big|^{2p} \Big)$$

$$\lesssim (\log N)^{-(K \wedge \frac{2p}{2p-1})} \sup_{0 \leq \gamma \leq 1} \mathbb{E}_\Psi \Big[\big| N\eta \big(\mathrm{Im}\, m^\gamma(z_t) - \mathrm{Im}\, \tilde{m}^0(z_t) \big) \big|^{2p} \Big] + N^{-\tilde{\delta}p}. \tag{116}$$

The claim now follows by rearranging the terms.

$\square$

5.5 Proof of Theorem 4.5

The proof of Theorem 4.5 is essentially the same as Theorem 4.3. We outline the proof here while the detailed proof can be found in [11].

Using the same notation as in the proof of Theorem 4.3 and further defining $h_{\gamma,(ij)}(\lambda, \beta) := \eta_0 \sum_a f_{\gamma,(aa),(ij)}(\lambda, \beta)$ and $\mathsf{H}_{(ij)}(\lambda, \beta) := F'\big(h_{\gamma,(ij)}(\lambda, \beta)\big) g_{(ij)}(\lambda, \beta)$. Observe that

$$\frac{\partial \mathbb{E}_\Psi \big(F(N\eta_0 \mathrm{Im}\, m^\gamma(z_t)) \big)}{\partial \gamma} = -2 \Big(\sum_{i,j} (I_1)_{ij} - (I_2)_{ij} \Big),$$

where $(I_1)_{ij} = \mathbb{E}_\Psi \big[\mathsf{A}_{ij} \mathsf{H}_{(ij)}([Y^\gamma]_{ij}, X_{ij}) \big]$ and $(I_2)_{ij} = \gamma(1-\gamma^2)^{-1/2} t^{1/2} \mathbb{E}_\Psi \big[w_{ij} \mathsf{H}_{(ij)}([Y^\gamma]_{ij}, X_{ij}) \big]$. We estimate them by considering the cases $\psi_{ij} = 1$ and $\psi_{ij} = 0$ separately. For $(I_2)_{ij}$, in both cases, we can estimate it by Gaussian integration by part, which leads to

$$(I_2)_{ij} = \frac{\gamma t^{1/2}}{(1-\gamma^2)^{1/2} N} \Big(\mathbb{E}_\Psi \big[\partial_{w_{ij}} \big\{ \mathsf{H}_{(ij)}(d_{ij}, \chi_{ij} b_{ij}) \big\} \big] \cdot \mathbf{1}_{\psi_{ij}=0} + \mathbb{E}_\Psi \big[\partial_{w_{ij}} \big\{ \mathsf{H}_{(ij)}(e_{ij}, c_{ij}) \big\} \big] \cdot \mathbf{1}_{\psi_{ij}=1} \Big).$$

The term involving $\mathbf{1}_{\psi_{ij}=1}$ can be estimated directly by the fact that $t^{1/2} N^{-1} \cdot \sum_{i,j} \mathbf{1}_{\psi_{ij}=1} \sim t^{1/2} N^{-1} \cdot N^{1-\epsilon_\alpha} = \mathfrak{o}(1)$. Therefore, by the definition of d_{ij}, we have

$$(I_2)_{ij} \approx \frac{\gamma t}{N} \mathbb{E}_\Psi \big[\partial_{d_{ij}} \big\{ \mathsf{H}_{(ij)}(d_{ij}, \chi_{ij} b_{ij}) \big\} \big] \cdot \mathbf{1}_{\psi_{ij}=0}. \tag{117}$$

For $(I_1)_{ij}$, we only need to consider the case $\psi_{ij} = \chi_{ij} = 0$ since $\mathsf{A}_{ij} \mathbf{1}_{\psi_{ij}=1 \text{ or } \chi_{ij}=1} = 0$. Using Taylor expansion and the law of total expectation gives

$$(I_1)_{ij} \approx \sum_k \frac{1}{k!} \mathbb{E}_\Psi[a_{ij} d_{ij}^k | \chi_{ij} = 0] \cdot \mathbb{E}_\Psi \big[\partial_{d_{ij}}^k \big\{ \mathsf{H}_{(ij)}(d_{ij}, \chi_{ij} b_{ij}) \big\} | \chi_{ij} = 0 \big] \cdot \mathbb{P}(\chi_{ij} = 0) \cdot \mathbf{1}_{\psi_{ij}=0}.$$

For even values of k, it holds that $\mathbb{E}_\Psi[a_{ij} d_{ij}^k | \chi_{ij} = 0] = 0$. In the case where $k \geq 3$, we have $\mathbb{E}_\Psi[a_{ij} d_{ij}^k | \chi_{ij} = 0] \sim N^{-2-\varepsilon}$ for some small $\varepsilon > 0$, effectively compensating for the size of the summation $\sum_{i,j}$. Consequently, we arrive at

$$(I_1)_{ij} \approx \mathbb{E}_\Psi[\gamma a_{ij}^2] \mathbb{P}(\chi_{ij} = 0) \cdot \mathbb{E}_\Psi \big[\partial_{d_{ij}} \big\{ \mathsf{H}_{(ij)}(d_{ij}, \chi_{ij} b_{ij}) \big\} | \chi_{ij} = 0 \big] \cdot \mathbf{1}_{\psi_{ij}=0}. \tag{118}$$

In view of (117) and (118), we can conclude the proof by leveraging the moment matching (87) and exploiting the smallness of $\big| \mathbb{E}_\Psi \big[\partial_{d_{ij}} \big\{ \mathsf{H}_{(ij)}(d_{ij}, \chi_{ij} b_{ij}) \big\} \big] - \mathbb{E}_\Psi \big[\partial_{d_{ij}} \big\{ \mathsf{H}_{(ij)}(d_{ij}, \chi_{ij} b_{ij}) \big\} | \chi_{ij} = 0 \big] \big|$.

Supplementary information. The remaining technical proofs are stated in the supplementary material [11].

Declarations

Not applicable.

References

[1] Adhikari A, Huang J (2020) Dyson Brownian motion for general β and potential at the edge. Probab Theory Related Fields 178(3-4):893–950. https://doi.org/10.1007/s00440-020-00992-9, URL https://doi.org/10.1007/s00440-020-00992-9

[2] Aggarwal A (2019) Bulk universality for generalized Wigner matrices with few moments. Probab Theory Related Fields 173(1-2):375–432. https://doi.org/10.1007/s00440-018-0836-y, URL https://doi.org/10.1007/s00440-018-0836-y

[3] Aggarwal A, Huang J (2023) Edge rigidity of dyson brownian motion with general initial data, URL https://arxiv.org/abs/2308.04236, arXiv:2308.04236

[4] Aggarwal A, Lopatto P, Yau HT (2021) GOE statistics for Lévy matrices. J Eur Math Soc (JEMS) 23(11):3707–3800. https://doi.org/10.4171/jems/1089, URL https://doi.org/10.4171/jems/1089

[5] Aggarwal A, Bordenave C, Lopatto P (2022) Mobility edge of lévy matrices, URL https://arxiv.org/abs/2210.09458, arXiv:2210.09458

[6] Alt J, Erdős L, Krüger T, et al (2020) Correlated random matrices: band rigidity and edge universality. Ann Probab 48(2):963–1001. https://doi.org/10.1214/19-AOP1379, URL https://doi.org/10.1214/19-AOP1379

[7] Auffinger A, Ben Arous G, Péché S (2009) Poisson convergence for the largest eigenvalues of heavy tailed random matrices. Ann Inst Henri Poincaré Probab Stat 45(3):589–610. https://doi.org/10.1214/08-AIHP188, URL https://doi.org/10.1214/08-AIHP188

[8] Bai Z, Silverstein JW (2010) Spectral analysis of large dimensional random matrices, 2nd edn. Springer Series in Statistics, Springer, New York, https://doi.org/10.1007/978-1-4419-0661-8, URL https://doi.org/10.1007/978-1-4419-0661-8

[9] Bai ZD, Yin YQ (1993) Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. Ann Probab 21(3):1275–1294. URL http://links.jstor.org/sici?sici=0091-1798(199307)21:3⟨1275:LOTSEO⟩2.0.CO;2-2&origin=MSN

[10] Bao Z, Pan G, Zhou W (2015) Universality for the largest eigenvalue of sample covariance matrices with general population. Ann Statist 43(1):382–421. https://doi.org/10.1214/14-AOS1281, URL https://doi.org/10.1214/14-AOS1281

[11] Bao Z, Lee J, Xu X (2023) Supplementary material for "phase transition for the smallest eigenvalue of covariance matrices", URL http://mazgbao.people.ust.hk

[12] Belinschi S, Dembo A, Guionnet A (2009) Spectral measure of heavy tailed band and covariance random matrices. Comm Math Phys 289(3):1023–1055. https://doi.org/10.1007/s00220-009-0822-4, URL https://doi.org/10.1007/s00220-009-0822-4

[13] Ben Arous G, Guionnet A (2008) The spectrum of heavy tailed random matrices. Comm Math Phys 278(3):715–751. https://doi.org/10.1007/s00220-007-0389-x, URL https://doi.org/10.1007/s00220-007-0389-x

[14] Benaych-Georges F, Guionnet A (2014) Central limit theorem for eigenvectors of heavy tailed matrices. Electron J Probab 19:no. 54, 27. https://doi.org/10.1214/EJP.v19-3093, URL https://doi.org/10.1214/EJP.v19-3093

[15] Benaych-Georges F, Maltsev A (2016) Fluctuations of linear statistics of half-heavy-tailed random matrices. Stochastic Process Appl 126(11):3331–3352. https://doi.org/10.1016/j.spa.2016.04.030, URL https://doi.org/10.1016/j.spa.2016.04.030

[16] Benaych-Georges F, Guionnet A, Male C (2014) Central limit theorems for linear statistics of heavy tailed random matrices. Comm Math Phys 329(2):641–686. https://doi.org/10.1007/s00220-014-1975-3, URL https://doi.org/10.1007/s00220-014-1975-3

[17] Bordenave C, Guionnet A (2013) Localization and delocalization of eigenvectors for heavy-tailed random matrices. Probab Theory Related Fields 157(3-4):885–953. https://doi.org/10.1007/s00440-012-0473-9, URL https://doi.org/10.1007/s00440-012-0473-9

[18] Bordenave C, Guionnet A (2017) Delocalization at small energy for heavy-tailed random matrices. Comm Math Phys 354(1):115–159. https://doi.org/10.1007/s00220-017-2914-x, URL https://doi.org/10.1007/s00220-017-2914-x

[19] Bourgade P, Erdős L, Yau HT, et al (2016) Fixed energy universality for generalized Wigner matrices. Comm Pure Appl Math 69(10):1815–1881. https://doi.org/10.1002/cpa.21624, URL https://doi.org/10.1002/cpa.21624

[20] Diaconu S (2023) More limiting distributions for eigenvalues of Wigner matrices. Ann Probab 51(2):774–804. https://doi.org/10.1214/22-aop1614, URL https://doi.org/10.1214/22-aop1614

[21] Ding X, Yang F (2018) A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. Ann Appl Probab 28(3):1679–1738. https://doi.org/10.1214/17-AAP1341, URL https://doi.org/10.1214/17-AAP1341

[22] Ding X, Yang F (2022) Edge statistics of large dimensional deformed rectangular matrices. J Multivariate Anal 192:Paper No. 105051, 22. https://doi.org/10.1016/j.jmva.2022.105051, URL https://doi.org/10.1016/j.jmva.2022.105051

[23] Ding X, Yang F (2022) Tracy-Widom distribution for heterogeneous Gram matrices with applications in signal detection. IEEE Trans Inform Theory 68(10):6682–6715

[24] Erdős L, Schnelli K (2017) Universality for random matrix flows with time-dependent density. Ann Inst Henri Poincaré Probab Stat 53(4):1606–1656. https://doi.org/10.1214/16-AIHP765, URL https://doi.org/10.1214/16-AIHP765

[25] Erdős L, Yau HT (2015) Gap universality of generalized Wigner and β-ensembles. J Eur Math Soc (JEMS) 17(8):1927–2036. https://doi.org/10.4171/JEMS/548, URL https://doi.org/10.4171/JEMS/548

[26] Erdős L, Yau HT (2017) A dynamical approach to random matrix theory, Courant Lecture Notes in Mathematics, vol 28. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI

[27] Erdős L, Knowles A, Yau HT, et al (2012) Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. Comm Math Phys 314(3):587–640. https://doi.org/10.1007/s00220-012-1527-7, URL https://doi.org/10.1007/s00220-012-1527-7

[28] Erdős L, Yau HT, Yin J (2012) Rigidity of eigenvalues of generalized Wigner matrices. Adv Math 229(3):1435–1515. https://doi.org/10.1016/j.aim.2011.12.010, URL https://doi.org/10.1016/j.aim.2011.12.010

[29] Erdős L, Knowles A, Yau HT (2013) Averaging fluctuations in resolvents of random band matrices. Ann Henri Poincaré 14(8):1837–1926. https://doi.org/10.1007/s00023-013-0235-y, URL https://doi.org/10.1007/s00023-013-0235-y

[30] Feldheim ON, Sodin S (2010) A universality result for the smallest eigenvalues of certain sample covariance matrices. Geom Funct Anal 20(1):88–123. https://doi.org/10.1007/s00039-010-0055-x, URL https://doi.org/10.1007/s00039-010-0055-x

[31] He Y, Knowles A (2021) Fluctuations of extreme eigenvalues of sparse Erdős-Rényi graphs. Probab Theory Related Fields 180(3-4):985–1056. https://doi.org/10.1007/s00440-021-01054-4, URL https://doi.org/10.1007/s00440-021-01054-4

[32] Heiny J, Mikosch T (2018) Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices. Stochastic Processes and their Applications 128(8):2779–2815

[33] Heiny J, Yao J (2022) Limiting distributions for eigenvalues of sample correlation matrices from heavy-tailed populations. Ann Statist 50(6):3249–3280. https://doi.org/10.1214/22-aos2226, URL https://doi.org/10.1214/22-aos2226

[34] Huang J, Landon B (2019) Rigidity and a mesoscopic central limit theorem for Dyson Brownian motion for general β and potentials. Probab Theory Related Fields 175(1-2):209–253. https://doi.org/10.1007/s00440-018-0889-y, URL https://doi.org/10.1007/s00440-018-0889-y

[35] Huang J, Yau HT (2022) Edge universality of sparse random matrices, URL https://arxiv.org/abs/2206.06580, arXiv:2206.06580

[36] Huang J, Landon B, Yau HT (2020) Transition from Tracy-Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős-Rényi graphs. Ann Probab 48(2):916–962. https://doi.org/10.1214/19-AOP1378, URL https://doi.org/10.1214/19-AOP1378

[37] Hwang JY, Lee JO, Schnelli K (2019) Local law and Tracy-Widom limit for sparse sample covariance matrices. Ann Appl Probab 29(5):3006–3036. https://doi.org/10.1214/19-AAP1472, URL https://doi.org/10.1214/19-AAP1472

[38] Johansson K (2000) Shape fluctuations and random matrices. Comm Math Phys 209(2):437–476. https://doi.org/10.1007/s002200050027, URL https://doi.org/10.1007/s002200050027

[39] Johnstone IM (2001) On the distribution of the largest eigenvalue in principal components analysis. Ann Statist 29(2):295–327. https://doi.org/10.1214/aos/1009210544, URL https://doi.org/10.1214/aos/1009210544

[40] Jung P (2018) Lévy-khintchine random matrices and the poisson weighted infinite skeleton tree. Transactions of the American Mathematical Society 370(1):641–668

[41] Knowles A, Yin J (2017) Anisotropic local laws for random matrices. Probab Theory Related Fields 169(1-2):257–352. https://doi.org/10.1007/s00440-016-0730-4, URL https://doi.org/10.1007/s00440-016-0730-4

[42] Landon B, Yau HT (2017) Convergence of local statistics of Dyson Brownian motion. Comm Math Phys 355(3):949–1000. https://doi.org/10.1007/s00220-017-2955-1, URL https://doi.org/10.1007/s00220-017-2955-1

[43] Landon B, Yau HT (2017) Edge statistics of dyson brownian motion, URL https://arxiv.org/abs/1712.03881, arXiv:1712.03881

[44] Landon B, Sosoe P, Yau HT (2019) Fixed energy universality of Dyson Brownian motion. Adv Math 346:1137–1332. https://doi.org/10.1016/j.aim.2019.02.010, URL https://doi.org/10.1016/j.aim.2019.02.010

[45] Lee J (2021) Higher order fluctuations of extremal eigenvalues of sparse random matrices, URL https://arxiv.org/abs/2108.11634, arXiv:2108.11634

[46] Lee JO, Schnelli K (2015) Edge universality for deformed Wigner matrices. Rev Math Phys 27(8):1550018, 94. https://doi.org/10.1142/S0129055X1550018X, URL https://doi.org/10.1142/S0129055X1550018X

[47] Lee JO, Schnelli K (2016) Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. Ann Appl Probab 26(6):3786–3839. https://doi.org/10.1214/16-AAP1193, URL https://doi.org/10.1214/16-AAP1193

[48] Lee JO, Schnelli K (2018) Local law and Tracy-Widom limit for sparse random matrices. Probab Theory Related Fields 171(1-2):543–616. https://doi.org/10.1007/s00440-017-0787-8, URL https://doi.org/10.1007/s00440-017-0787-8

[49] Lee JO, Yin J (2014) A necessary and sufficient condition for edge universality of Wigner matrices. Duke Math J 163(1):117–173. https://doi.org/10.1215/00127094-2414767, URL https://doi.org/10.1215/00127094-2414767

[50] Louvaris M (2022) Universality of the least singular value and singular vector delocalisation for lévy non-symmetric random matrices, URL https://arxiv.org/abs/2204.06399, arXiv:2204.06399

[51] Malysheva S (2023) Central limit theorem for traces of the resolvents of half-heavy tailed sample covariance matrices, URL https://arxiv.org/abs/2307.10848, arXiv:2307.10848

[52] Marchenko VA, Pastur LA (1967) Distribution of eigenvalues for some sets of random matrices. Matematicheskii Sbornik 114(4):507–536

[53] Péché S (2009) Universality results for the largest eigenvalues of some sample covariance matrix ensembles. Probab Theory Related Fields 143(3-4):481–516. https://doi.org/10.1007/s00440-007-0133-7, URL https://doi.org/10.1007/s00440-007-0133-7

[54] Pillai NS, Yin J (2014) Universality of covariance matrices. Ann Appl Probab 24(3):935–1001. https://doi.org/10.1214/13-AAP939, URL https://doi.org/10.1214/13-AAP939

[55] Schnelli K, Xu Y (2022) Convergence rate to the Tracy-Widom laws for the largest eigenvalue of Wigner matrices. Comm Math Phys 393(2):839–907. https://doi.org/10.1007/s00220-022-04377-y, URL https://doi.org/10.1007/s00220-022-04377-y

[56] Sodin S (2010) The spectral edge of some random band matrices. Ann of Math (2) 172(3):2223–2251. https://doi.org/10.4007/annals.2010.172.2223, URL https://doi.org/10.4007/annals.2010.172.2223

[57] Soshnikov A (1999) Universality at the edge of the spectrum in Wigner random matrices. Comm Math Phys 207(3):697–733. https://doi.org/10.1007/s002200050743, URL https://doi.org/10.1007/s002200050743

[58] Soshnikov A (2004) Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails. Electron Comm Probab 9:82–91. https://doi.org/10.1214/ECP.v9-1112, URL https://doi.org/10.1214/ECP.v9-1112

[59] Tao T, Vu V (2010) Random matrices: universality of local eigenvalue statistics up to the edge. Comm Math Phys 298(2):549–572. https://doi.org/10.1007/s00220-010-1044-5, URL https://doi.org/10.1007/s00220-010-1044-5

[60] Tikhomirov K (2015) The limit of the smallest singular value of random matrices with i.i.d. entries. Adv Math 284:1–20. https://doi.org/10.1016/j.aim.2015.07.020, URL https://doi.org/10.1016/j.aim.2015.07.020

[61] Tikhomirov KE (2016) The smallest singular value of random rectangular matrices with no moment assumptions on entries. Israel J Math 212(1):289–314. https://doi.org/10.1007/s11856-016-1287-8, URL https://doi.org/10.1007/s11856-016-1287-8

[62] Tracy CA, Widom H (1994) Level-spacing distributions and the Airy kernel. Comm Math Phys 159(1):151–174. URL http://projecteuclid.org/euclid.cmp/1104254495

[63] Tracy CA, Widom H (1996) On orthogonal and symplectic matrix ensembles. Comm Math Phys 177(3):727–754. URL http://projecteuclid.org/euclid.cmp/1104286442

[64] Vallet P, Loubaton P, Mestre X (2012) Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case. IEEE Transactions on Information Theory 58(2):1043–1068