Supplementary material for “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

This is a supplementary material for [2], “Phase transition for the smallest eigenvalue of covariance matrices”.

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

In this supplementary material, we shall give the proofs of the following results stated in [2]: Theorem 2.8, Lemma 2.9, Proposition 2.10, Lemma 3.2, Proposition 3.10, Lemma 5.1, and Corollary . We also provide a remark on Theorem 2.11 in [2].

1 Proof of Lemma 2.9

Consider

$$z = (\lambda_-^{\mathsf{mp}} + E) + \mathrm{i}\eta, \quad |E| \le N^{-\varepsilon_1}, \quad N^{-2/3-\varepsilon_2} \le \eta \le \varepsilon_3. \tag{S.1.1}$$

Recall that

$$V_t = \sqrt{t}W + X,$$

where $t = N\mathbb{E}|\mathsf{A}_{ij}|^2$.

By the eigenvalue rigidity (the left edge analog of [6, Theorem 2.13]),

$$|\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t}| \prec N^{-2/3}.$$

As an analog of Lemma 2.6,

$$|\lambda_M(\mathcal{S}(V_t)) - \lambda_-^{\mathsf{mp}}| \prec N^{-2\epsilon_b}.$$

Thus,

$$|\lambda_-^{\mathsf{mp}} - \lambda_{-,t}| \prec N^{-2/3} + N^{-2\epsilon_b} \lesssim N^{-2\varepsilon_1}.$$

We write

$$z = \{\lambda_{-,t} + (\lambda_-^{\mathsf{mp}} - \lambda_{-,t}) + E\} + \mathrm{i}\eta =: (\lambda_{-,t} + E') + \mathrm{i}\eta,$$

where $E' := E + (\lambda_-^{\mathsf{mp}} - \lambda_{-,t})$. Then, with high probability, there exists $\kappa \in \mathbb{R}$ such that

$$z = (\lambda_{-,t} + \kappa) + \mathrm{i}\eta, \quad |\kappa| \le 2N^{-\varepsilon_1}, \quad N^{-2/3-\varepsilon_2} \le \eta \le \varepsilon_3. \tag{S.1.2}$$

Then, the desired result directly follows from the lemma below. Define $b_t \equiv b_t(z) := 1 + c_N t m_t(z)$. Then we have $\zeta_t(z) := z b_t^2 - t b_t (1 - c_N)$.

Lemma S.1.1. *Let z as in (S.1.2). There exist constants $c, C > 0$ such that the following holds:*
(i) For $|\kappa| + \eta \le ct^2(\log N)^{-2C}$,

$$\lambda_M(XX^\top) - \operatorname{Re}\zeta_t(z) \ge ct^2, \quad \operatorname{Im}\zeta_t(z) \ge ctN^{-2/3-\varepsilon_2}.$$

(ii) For $|\kappa| + \eta \ge ct^2(\log N)^{-2C}$,

$$\operatorname{Im}\zeta_t(z) \ge ct^2(\log N)^{-C}.$$

Proof. This lemma is essentially a byprduct of Theorem 2.7 through some elementary calculations. Comparing $\zeta_t(\lambda_{-,t})$ and $\zeta_t(z)$, it boils down to the size of $m_t(\lambda_{-,t}) - m_t(z)$. We shall rely on the square root behavior of ρ_t.

Case (1) $|\kappa| \le 2\eta$. Notice that

$$|m_t(\lambda_{-,t}) - m_t(z)| \le \int_{\lambda_{-,t}}^{\lambda_{+,t}} \frac{3\eta}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda.$$

By the square-root behavior of ρ_t near the left edge,

$$\int_{\lambda_{-,t}}^{\lambda_{-,t}+6\eta} \frac{\eta}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda \lesssim \int_{\lambda_{-,t}}^{\lambda_{-,t}+6\eta} \frac{\eta}{\eta\sqrt{\lambda - \lambda_{-,t}}}d\lambda \lesssim \sqrt{\eta}.$$

If $\lambda \ge \lambda_{-,t} + 6\eta$, we have $\lambda - \lambda_{-,t} - 3\eta \ge (\lambda - \lambda_{-,t})/2$. Thus,

$$\int_{\lambda_{-,t}+6\eta}^{\lambda_{+,t}} \frac{\eta}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda \lesssim \int_{\lambda_{-,t}+6\eta}^{\lambda_{+,t}} \frac{\eta}{(\lambda - \lambda_{-,t})^{3/2}}d\lambda \lesssim \sqrt{\eta}.$$

Case (2) $\kappa > 2\eta$. We need to estimate

$$\int_{\lambda_{-,t}}^{\lambda_{+,t}} \frac{\kappa}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda.$$

Due to the square-root decay,

$$\int_{\lambda_{-,t}}^{\lambda_{-,t}+\eta} \frac{\kappa}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda \lesssim \int_{\lambda_{-,t}}^{\lambda_{-,t}+\eta} \frac{\kappa}{\kappa\sqrt{\lambda - \lambda_{-,t}}}d\lambda \lesssim \sqrt{\eta}.$$

We also observe

$$\int_{\lambda_{-,t}+\eta}^{\lambda_{-,t}+\kappa-\eta} \frac{\kappa}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda \lesssim \int_{\eta}^{\kappa-\eta} \frac{\kappa}{\sqrt{x}(\kappa - x)}dx \lesssim \sqrt{\kappa}\log(\kappa/\eta).$$

If $\lambda \in [\lambda_{-,t} + \kappa - \eta, \lambda_{-,t} + 2\kappa]$, we have $\lambda - \lambda_{-,t} \sim \kappa$, which implies

$$\int_{\lambda_{-,t}+\kappa-\eta}^{\lambda_{-,t}+2\kappa} \frac{\kappa}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda \lesssim \int_0^{\kappa} \frac{\sqrt{\kappa}}{\sqrt{x^2 + \eta^2}}dx \lesssim \sqrt{\kappa}\log(\kappa/\eta).$$

For $\lambda \in [\lambda_{-,t} + 2\kappa, \lambda_{+,t}]$,

$$\int_{\lambda_{-,t}+2\kappa}^{\lambda_{+,t}} \frac{\kappa}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda \lesssim \sqrt{\kappa}$$

Case (3) $\kappa < -2\eta$. By splitting $[\lambda_{-,t}, \lambda_{+,t}]$ into $[\lambda_{-,t}, \lambda_{-,t} + |\kappa|]$ and $[\lambda_{-,t} + |\kappa|, \lambda_{+,t}]$, we find that

$$\int_{\lambda_{-,t}}^{\lambda_{+,t}} \frac{\kappa}{|\lambda - \lambda_{-,t}||\lambda - z|}\rho_t(\lambda)d\lambda \lesssim \sqrt{|\kappa|}.$$

Note $|b_t(\lambda_{-,t})| = O(1) = |b_t(z)|$ due to the fact that $|m_t(u)| \lesssim (t|u|)^{-1/2}$. Thus, for $|\kappa|+\eta \le (\log N)^{-C} t^2$,

$$|\zeta_t(z) - \zeta_t(\lambda_{-,t})| \ll t^2.$$

By Lemma 2.6 and Lemma 3.2,

$$(1-t)\lambda_-^{\mathsf{mp}} - \operatorname{Re}\zeta_t(z) = \big((1-t)\lambda_-^{\mathsf{mp}} - \lambda_M(\mathcal{S}(X))\big) + (\lambda_M(\mathcal{S}(X)) - \zeta_t(\lambda_{-,t})) + \operatorname{Re}[\zeta_t(\lambda_{-,t})) - \zeta_t(z)] \sim t^2.$$

Next, we consider the imaginary part of $\zeta_t(z)$. Setting

$$\Phi(\kappa,\eta) = \begin{cases} \sqrt{\kappa+\eta}, & \kappa \ge 0, \\ \frac{\eta}{\sqrt{|\kappa|+\eta}}, & \kappa < 0, \end{cases}$$

we have $\operatorname{Im}\zeta_t(z) \sim \eta + t\Phi(\kappa,\eta)$, which gives the desired estimates on the imaginary part of $\zeta_t(z)$. □

2 Proof of Proposition 2.10

We estimate the size of $G_{ij}(X,\zeta)$ only. We can bound $G_{ij}(X^\top,\zeta)$ in a similar way. Define $H := X/\sqrt{1-t}$ and denote $\omega := \zeta/(1-t)$. It is enough to find a constant $c = c(\epsilon_a, \epsilon_\alpha, \epsilon_b)$ such that

$$|G_{ij}(H,\omega) - \delta_{ij}\mathsf{m}_{\mathsf{mp}}(\omega)| \prec N^{-c}\mathbf{1}_{i,j\in\mathcal{T}_r} + t^{-2}(1-\mathbf{1}_{i,j\in\mathcal{T}_r}).$$

This can be proved by a minor modication of [11, Section 6]. In light of Lemma 2.6, the following two lemmas are trivial. We may use the rigidiy estimate, Lemma 2.6, to get Lemma S.2.2 below.

Lemma S.2.1 (Crude bound using the imaginary part)**.** *Consider $\omega = E + \mathrm{i}\eta \in \mathbb{C}_+$. If $\eta > C$,*

$$|G_{ij}(H,\omega)| \le C^{-1}.$$

Lemma S.2.2 (Crude bound on the domain D_ζ)**.** *Let $\mathsf{D}_\zeta = \mathsf{D}_\zeta(c_0, C_0)$ be as in Eq. (31). Let $\zeta \in \mathsf{D}_\zeta$. Denote $\omega = \zeta/(1-t)$. Then with high probability,*

$$|G_{ij}(H,\omega)| \lesssim (\log N)^{C_0} t^{-2}.$$

Let us write $H = (h_{ij})$. By Schur complement,

$$G_{ii}(H,\omega) = -\frac{1}{\omega + \frac{\omega}{N}\sum_{k=1}^N G_{kk}((H^{(i)})^\top,\omega) + Z_i} \tag{S.2.1}$$

where we denote by $H^{(i)}$ the matrix obtained from H by removing i-th row and

$$Z_i := \omega \sum_{1\le k,l\le N} h_{ik}h_{il}G_{kl}((H^{(i)})^\top,\omega) - \frac{\omega}{N}\sum_{k=1}^N G_{kk}((H^{(i)})^\top,\omega).$$

We define $\Lambda_d(\omega)$, $\Lambda_o(\omega)$ and $\Lambda(\omega)$ by

$$\Lambda_d(\omega) = \max_{i\in\mathcal{T}_r}|G_{ii}(H,\omega) - \mathsf{m}_{\mathsf{mp}}(\omega)|,\quad \Lambda_o(\omega) = \max_{\substack{i\ne j\\ i,j\in\mathcal{T}_r}}|G_{ij}(H,\omega)|,\quad \Lambda(\omega) = |m_H(\omega) - \mathsf{m}_{\mathsf{mp}}(\omega)|.$$

For $\omega = E + \mathrm{i}\eta$, we define

$$\Phi \equiv \Phi(\omega) \coloneqq \sqrt{\frac{\operatorname{Im} \mathsf{m}_{\mathsf{mp}}(\omega) + \Lambda(\omega)}{N\eta}} + t^{-2}N^{-\epsilon_a/2} + t^{-2}N^{-\epsilon_b}.$$

Define the events $\Omega(\omega, K)$, $\boldsymbol{B}(\omega)$ and $\Gamma(\omega, K)$ for $K > 0$ by

$$\Omega(\omega, K) \coloneqq \Big\{ \max\Big(\Lambda_o(\omega), \max_{i\in\mathcal{T}_r} |G_{ii}(H,\omega) - m_H(\omega)|, \max_{i\in\mathcal{T}_r} |Z_i(\omega)|\Big) \geq K\Phi \Big\},$$

$$\boldsymbol{B}(\omega) \coloneqq \{\Lambda_o(\omega) + \Lambda_d(\omega) > (\log N)^{-1}\}, \quad \Gamma(\omega, K) \coloneqq \Omega^c(\omega, K) \cup \boldsymbol{B}(\omega).$$

We also introduce the logarithmic factor $\varphi \equiv \varphi_N \coloneqq (\log N)^{\log\log N}$.

Lemma S.2.3. *Suppose Ψ is good. Recall $\omega \equiv \omega(\zeta) = \zeta/(1-t)$. There exist a constant $C > 0$ such that the event*

$$\bigcap_{\zeta \in \mathsf{D}_\zeta} \Gamma(\omega, \varphi^C)$$

holds with high probability.

Proof. By a standard lattice argument, it is enough to show that $\Gamma(\omega, \varphi^C)$ holds with with high probability for any $\omega = \omega(\zeta)$ with $\zeta \in \mathsf{D}_\zeta$. Fix $\omega = \omega(\zeta)$ with $\zeta \in \mathsf{D}_\zeta$. We define

$$\begin{aligned}
\Omega_o(\omega, K) &\coloneqq \big\{\Lambda_o(\omega) \geq K\Phi(\omega)\big\}, \\
\Omega_d(\omega, K) &\coloneqq \Big\{ \max_{i\in\mathcal{T}_r} |G_{ii}(H,\omega) - m_H(\omega)| \geq K\Phi(\omega)\Big\}, \\
\Omega_Z(\omega, K) &\coloneqq \Big\{ \max_{i\in\mathcal{T}_r} |Z_i| \geq K\Phi(\omega)\Big\}.
\end{aligned}$$

Since $\Omega = \Omega_o \cup \Omega_d \cup \Omega_Z$, it is sufficient to show $\Omega_o^c \cup \boldsymbol{B}$, $\Omega_d^c \cup \boldsymbol{B}$ and $\Omega_Z^c \cup \boldsymbol{B}$ hold with high probability respectively.

(1) Consider the event $\Omega_o^c \cup \boldsymbol{B}$. Fix $i \neq j$ with $i, j \in \mathcal{T}_r$. On the event $\boldsymbol{B}^c$, we have $|G_{ii}(H,\zeta)| \sim 1$. Then, by the resolvent identity,

$$G_{jj}(H^{(i)}, \omega) = G_{jj}(H, \omega) - \frac{G_{ji}(H,\omega)G_{ij}(H,\omega)}{G_{ii}(H,\omega)}, \tag{S.2.2}$$

it follows that $G_{jj}(H^{(i)}, \omega) \sim 1$ on $\boldsymbol{B}^c$. Thus, we can get

$$\Lambda_o(\omega) \lesssim \max_{\substack{i\neq j \\ i,j\in\mathcal{T}_r}} \left| \sum_{1\leq k,l\leq N} h_{ik}h_{jl}G_{kl}((H^{(ij)})^\top, \omega) \right|,$$

where we denote by $H^{(ij)}$ the matrix obtained from H by removing i-th and j-th rows. Since $i, j \in \mathcal{T}_r$, applying the large deviation estimate [1, Corollary 25], the following estimate holds with high probability:

$$\left| \sum_{1\leq k,l\leq N} h_{ik}h_{jl}G_{kl}((H^{(ij)})^\top, \omega) \right| \leq \varphi^C \left(N^{-\epsilon_b} \max_{k,l} |G_{kl}((H^{(ij)})^\top, \omega)| + \frac{1}{N}\Big(\sum_{k,l} |G_{kl}((H^{(ij)})^\top, \omega)|^2 \Big)^{1/2} \right).$$

Note that

$$\sum_{k,l} |G_{kl}((H^{(ij)})^\top, \omega)|^2 = \frac{\sum_k \operatorname{Im} G_{kk}((H^{(ij)})^\top, \omega)}{\eta}, \tag{S.2.3}$$

and

$$\sum_k G_{kk}((H^{(ij)})^\top, \omega) - \sum_\ell G_{\ell\ell}(H^{(ij)}, \omega) = \frac{O(N)}{\omega}. \tag{S.2.4}$$

Using (S.2.2), (S.2.3) and (S.2.4), together with Lemma S.2.2, we conclude that on the event $\boldsymbol{B}^c$, with high probability, for some constant $C > 0$ large enough,

$$\Lambda_o(\omega) \le \varphi^C \left(t^{-2} N^{-\epsilon_b} + \sqrt{\frac{\operatorname{Im} \mathsf{m}_{\mathsf{mp}} + \Lambda + \Lambda_o^2 + t^{-4} N^{-\epsilon_\alpha}}{N\eta} + \frac{1}{N}} \right),$$

with high probability for some constant $C > 0$ large enough. The event $\Omega_o^c \cap \boldsymbol{B}^c$ holds with high probability.

(2) We claim that $\Omega_Z^c \cup \boldsymbol{B}$ holds with high probability. In fact, the claim directly follows from the large deviation estimate [1, Corollary 25] repeating the same argument we used above; on the event $\boldsymbol{B}^c$, for $i \in \mathcal{T}_r$, we have $|Z_i| \le \varphi^C \Phi$ with high probability for some constant $C > 0$.

(3) We shall prove $\Omega_d^c \cup \boldsymbol{B}$ holds with high probability. For $i \in \mathcal{T}_r$,

$$G_{ii}(H,\omega) - m_H(\omega) \le \max_{j \in \mathcal{T}_r} |G_{ii}(H,\omega) - G_{jj}(H,\omega)| + \varphi^C t^{-2} N^{-\epsilon_\alpha},$$

where we use Lemma S.2.2 to bound G_{jj} with $j \notin \mathcal{T}_r$. For $i, j \in \mathcal{T}_r$ with $i \neq j$, on the event $\boldsymbol{B}^c$, with high probability, we can find that

$$\begin{aligned} |G_{ii}(H,\omega) - G_{jj}(H,\omega)| &\le \left| \frac{1}{\omega + \frac{\omega}{N} \sum_{k=1}^N G_{kk}((H^{(i)})^\top, \omega) + Z_i} - \frac{1}{\omega + \frac{\omega}{N} \sum_{k=1}^N G_{kk}((H^{(j)})^\top, \omega) + Z_j} \right| \\ &\lesssim \max_{i \in \mathcal{T}_r} |Z_i| + \Lambda_o^2 + t^{-4} N^{-\epsilon_\alpha} \end{aligned}$$

where we use

$$\sum_k G_{kk}((H^{(i)})^\top, \omega) - \sum_\ell G_{\ell\ell}(H^{(i)}, \omega) = \frac{M - N + 1}{\omega} \tag{S.2.5}$$

and the estimates we have shown above. The desired result follows. □

Corollary S.2.4. *Suppose Ψ is good. Let $C' > 0$ be a constant. There exist a constant $C > 0$ such that the event $\Omega^c(E + \mathrm{i}\eta, \varphi^C)$ holds with high probability.*

Proof. Recall the argument we used in the proof of the previous lemma. Using the large deviation estimate [1, Corollary 25] with Lemma S.2.1, it is straightforward that Ω_o^c and Ω_Z^c hold with high probability. For Ω_d^c, the desired result follows from the consequence of Cauchy's interlacing theorem, that is,

$$\frac{1}{N} \sum_{k=1}^N G_{kk}((H^{(i)})^\top, \omega) - \frac{1}{N} \sum_{k=1}^N G_{kk}((H^{(j)})^\top, \omega) \lesssim \frac{1}{N\eta}.$$

□

Let us introduce the deviance function $D(u(\omega), \omega)$ by setting

$$D(u(\omega), \omega) := \left(\frac{1}{u(\omega)} + c_N \omega u(\omega) \right) - \left(\frac{1}{\mathsf{m}_{\mathsf{mp}}(\omega)} + c_N \omega \mathsf{m}_{\mathsf{mp}}(\omega) \right).$$

Lemma S.2.5. *On the event $\Gamma(\omega, \varphi^C)$,*

$$|D(m_H(\omega), \omega)| \le O(\varphi^{2C} \Phi^2) + \infty \mathbb{1}_{\boldsymbol{B}(\omega)}.$$

Proof. Recall that $(\mathsf{m}_{\mathsf{mp}})^{-1}(\omega) = -\omega + (1 - c_N) - \omega c_N \mathsf{m}_{\mathsf{mp}}$. Using (S.2.1), (S.2.2) and (S.2.5), on the event $\Omega^c \cap \boldsymbol{B}^c$, we have

$$G_{ii}^{-1}(H,\omega) = (\mathsf{m}_{\mathsf{mp}})^{-1}(\omega) + \omega c_N (\mathsf{m}_{\mathsf{mp}}(\omega) - m_H(\omega)) - Z_i + O(\varphi^{2C} \Phi^2 + t^{-4} N^{-\epsilon_\alpha} + N^{-1}),$$

so it follows that

$$m_H^{-1}(\omega) - G_{ii}^{-1}(H,\omega) = D(m_H(\omega),\omega) + Z_i + O(\varphi^{2C}\Phi^2 + t^{-4}N^{-\epsilon_\alpha} + N^{-1}).$$

Averaging over $i \in \mathcal{T}_r$ yields

$$\frac{1}{|\mathcal{T}_r|}\sum_{i\in\mathcal{T}_r}(m_H^{-1}(\omega) - G_{ii}^{-1}(H,\omega)) = D(m_H(\omega),\omega) + \frac{1}{|\mathcal{T}_r|}\sum_{i\in\mathcal{T}_r} Z_i + O(\varphi^{2C}\Phi^2 + t^{-4}N^{-\epsilon_\alpha} + N^{-1}).$$

Since $\sum_i G_{ii}(H,\omega) - m_H(\omega) = 0$ and

$$m_H^{-1}(\omega) - G_{ii}^{-1}(H,\omega) = \frac{G_{ii}(H,\omega) - m_H(\omega)}{m_H^2(\omega)} - \frac{\big(G_{ii}(H,\omega) - m_H(\omega)\big)^2}{m_H^3(\omega)} + O\Big(\frac{\big(G_{ii}(H,\omega) - m_H(\omega)\big)^3}{m_H^4(\omega)}\Big),$$

we obtain that $|D(m_H(\omega),\omega)| \le O(\varphi^{2C}\Phi^2)$ on the event $\Omega^c \cap \boldsymbol{B}^c$. □

Lemma S.2.6. *Recall $\omega \equiv \omega(\zeta) = \zeta/(1-t)$ and write $\omega = E + \mathrm{i}\eta$. Let $C, C' > 0$ be constants. Consider an event A such that*

$$A \subset \bigcap_{\zeta\in\mathsf{D}_\zeta}\Gamma(\omega,\varphi^C) \cap \bigcap_{\zeta\in\mathsf{D}_\zeta,\eta=C'}\boldsymbol{B}^c(\omega).$$

Suppose that in A, for $\omega = \omega(\zeta)$ with $\zeta \in \mathsf{D}_\zeta$,

$$|D(m_H(\omega),\omega)| \le \mathfrak{d}(\omega) + \infty\mathbb{1}_{B(\omega)},$$

where $\mathfrak{d} : \mathbb{C} \mapsto \mathbb{R}_+$ is a continuous function such that $\mathfrak{d}(E+\mathrm{i}\eta)$ is decreasing in η and $|\mathfrak{d}(z)| \le (\log N)^{-8}$.

Then, for all $\omega \equiv \omega(\zeta)$ with $\zeta \in \mathsf{D}_\zeta$, we have

$$|m_H(\omega) - \mathsf{m}_{\mathsf{mp}}(\omega)| \lesssim \log N \frac{\mathfrak{d}(\zeta)}{\sqrt{|E - \lambda_-^{\mathsf{mp}}| + \eta + \mathfrak{d}(\zeta)}} \quad \text{in } A, \tag{S.2.6}$$

and

$$A \subset \bigcap_{\zeta\in\mathsf{D}_\zeta}\boldsymbol{B}^c(\zeta). \tag{S.2.7}$$

Proof. We follow the proof of [11, Lemma 6.12]. Denote $\omega = \omega(\zeta) = E + \mathrm{i}\eta$ with $\zeta \in \mathsf{D}_\zeta$. For each E, we define

$$I_E := \{\eta : \Lambda_o(E+\mathrm{i}\eta') + \Lambda_d(E+\mathrm{i}\eta') \le (\log N)^{-1} \text{ for all } \eta' \ge \eta \text{ such that } (1-t)\cdot(E+\mathrm{i}\eta') \in \mathsf{D}_\zeta\}.$$

Let m_1 and m_2 be two solutions of equation $D(m(\omega),\omega) = \mathfrak{d}(\omega)$. On $\boldsymbol{B}^c(\omega)$, by assumption, we have

$$|D(m_H(\omega),\omega)| \le \mathfrak{d}(\omega).$$

Then, the estimate (S.2.6) immediately follows from the argument around [11, Eq. (6.45)–Eq. (6.46)].

Next, we will prove the second statement (S.2.7). Due to the case $\eta = C'$, we know $I_E \neq \emptyset$ on A. Let us argue by contradiction. Define

$$\mathcal{D}_E = \{\eta : \omega = E + \mathrm{i}\eta, (1-t)\cdot\omega \in \mathsf{D}_\zeta\}.$$

Assume $I_E \neq \mathcal{D}_E$. Let $\eta_0 = \inf I_E$. For $\omega_0 = E + \mathrm{i}\eta_0$, we have $\Lambda_o(\omega_0) + \Lambda_d(\omega_0) = (\log N)^{-1}$. It also follows

$$\Lambda(\omega_0) \le \Big|\frac{1}{N}\sum_{i\in\mathcal{T}_r}\big(G_{ii}(H,\omega_0) - \mathsf{m}_{\mathsf{mp}}(\omega_0)\big)\Big| + \Big|\frac{1}{N}\sum_{i\notin\mathcal{T}_r}\big(G_{ii}(H,\omega_0) - \mathsf{m}_{\mathsf{mp}}(\omega_0)\big)\Big|$$
$$\le (\log N)^{-1} + \varphi^C t^{-2} N^{-\epsilon_\alpha} \lesssim (\log N)^{-1}.$$

By the first statement we already proved, on the event A, we obtain

$$\Lambda(\omega_0) \lesssim (\log N)^{-3}.$$

Since $\Lambda_o(\omega_0)+\Lambda_d(\omega_0) = (\log N)^{-1}$, we have $A \subset \boldsymbol{B}^c(\omega_0)$ and thus, by the assumption for A, we conclude that $\Lambda_o(\omega_0) + \Lambda_d(\omega_0) \ll (\log N)^{-1}$ on the event A, which makes a contradiction.

□

Proposition S.2.7. *Recall $\omega \equiv \omega(\zeta) = \zeta/(1-t)$ and write $\omega = E + \mathrm{i}\eta$. There exist a constant $C > 0$ such that the following event holds with high probability:*

$$\bigcap_{\zeta \in \mathsf{D}_\zeta} \{\Lambda_o(\omega) + \Lambda_d(\omega) \le \varphi^C (t^{-2}(N\eta)^{-1/2} + t^{-3}N^{-\epsilon_a/2} + t^{-3}N^{-\epsilon_b})\}.$$

Proof. Consider the event

$$A_0 = \bigcap_{\zeta \in \mathsf{D}_\zeta} \Gamma(\omega, \varphi^C).$$

Also we set (for some constant $C' > 1$ and $\omega = E + \mathrm{i}\eta$)

$$A = A_0 \cap \bigcap_{\zeta \in \mathsf{D}_\zeta, \eta = C'} \boldsymbol{B}^c(\omega).$$

By Lemma S.2.3 and Corollary S.2.4, the event A holds with high probability. Using Lemma S.2.2, we observe that for $\omega = \omega(\zeta)$ with $\zeta \in \mathsf{D}_\zeta$,

$$\Phi(\omega) \lesssim \varphi t^{-1}(N\eta)^{-1/2} + t^{-2}N^{-\epsilon_a/2} + t^{-2}N^{-\epsilon_b}.$$

Let us set

$$\mathfrak{d}(\omega) = \varphi^C \big(t^{-1}(N\eta)^{-1/2} + t^{-2}N^{-\epsilon_a/2} + t^{-2}N^{-\epsilon_b}\big).$$

On the event A, for $\omega = \omega(\zeta)$ with $\zeta \in \mathsf{D}_\zeta$, by Lemma S.2.5 and Lemma S.2.6,

$$\Lambda(\omega) \lesssim \frac{\mathfrak{d}(\omega)}{\sqrt{|E - \lambda_-^{\mathsf{mp}}| + \eta}}.$$

Also, by Lemma S.2.6,

$$A \subset \bigcap_{\zeta \in \mathsf{D}_\zeta} \boldsymbol{B}^c(\omega),$$

which means the event A is contained in $\Omega^c(\omega, \varphi^C)$ for any $\omega = \omega(\zeta)$ with $\zeta \in \mathsf{D}_\zeta$. The bound for Λ_d is given by $\max_{k \in \mathcal{T}_r} |G_{kk}(H, \omega) - m_H| + \Lambda$. □

3 Proof of Theorem 2.8

Recall $b_t = 1 + c_N t m_t$ and $\zeta_t = \zeta_t(z) = z b_t^2 - t b_t(1 - c_N)$. We also set

$$\underline{m}_t = c_N m_t - \frac{1 - c_N}{z}, \qquad \underline{\mathsf{m}}_{\mathsf{mp}}^{(t)}(\zeta) = c_N \mathsf{m}_{\mathsf{mp}}^{(t)}(\zeta) - \frac{1 - c_N}{\zeta}.$$

Let us state a left edge analog of [6, Theorem 2.7].

Theorem S.3.1. *Suppose that the assumptions in Theorem 2.8 hold. Then,*

$$|G_{ij}(V_t, z) - b_t G_{ij}(X, \zeta_t(z))| \prec t^{-3}\left(\sqrt{\frac{\operatorname{Im} m_t}{N\eta}} + \frac{1}{N\eta}\right) + \frac{t^{-7/2}}{N^{1/2}},$$

and

$$|G_{ij}(V_t^\top, z) - (1 + t\underline{m}_t)G_{ij}(X^\top, \zeta_t(z))| \prec t^{-3}\left(\sqrt{\frac{\operatorname{Im} m_t}{N\eta}} + \frac{1}{N\eta}\right) + \frac{t^{-7/2}}{N^{1/2}},$$

uniformly in $z \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. *In addition,*

$$|(G(V_t, z)V_t)_{ij} - (G(X, \zeta_t(z))X)_{ij}| \prec t^{-3}\left(\sqrt{\frac{\operatorname{Im} m_t}{N\eta}} + \frac{1}{N\eta}\right) + \frac{t^{-7/2}}{N^{1/2}},$$

and

$$|(V_t^\top G(V_t, z))_{ij} - (X^\top G(X, \zeta_t(z)))_{ij}| \prec t^{-3}\left(\sqrt{\frac{\operatorname{Im} m_t}{N\eta}} + \frac{1}{N\eta}\right) + \frac{t^{-7/2}}{N^{1/2}},$$

uniformly in $z \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$.

Proof. Roughly speaking, the conclusion is a left edge analog of [6, Theorem 2.7]. The proof is nearly the same, and thus we only highlight some differences. We first record the notations from [6, Section B of Supplement]. Due to the rotationally invariant property of Gaussian matrix, we have

$$V_t = X + \sqrt{t}W \overset{d}{=} O_1 \tilde{V}_t O_2^\top, \quad \tilde{V}_t := \tilde{X} + \sqrt{t}W, \tag{S.3.1}$$

where $\tilde{X}$ is a diagonal matrix with diagonal entries being $\lambda_i(\mathcal{S}(X))^{1/2}, i \in [M]$. Recall the notations in Lemma 5.2, and we briefly write $\mathcal{R}(z) = \mathcal{R}(\tilde{V}_t, z)$ in this proof. By (S.3.1), to prove an entrywise local law for $\mathcal{R}(V_t, z)$, it suffices to prove an anisotropic local law for the resolvent $\mathcal{R}(z)$. We further define the asymptotic limit of $\mathcal{R}(z)$ as

$$\Pi^x(z) := \begin{bmatrix} \frac{-(1+c_N t m_t)}{z(1+c_N t m_t)(1+t\underline{m}_t) - \tilde{X}\tilde{X}^\top} & \frac{-z^{-1/2}}{z(1+c_N t m_t)(1+t\underline{m}_t) - \tilde{X}\tilde{X}^\top}\tilde{X} \\ \tilde{X}^\top \frac{-z^{-1/2}}{z(1+c_N t m_t)(1+t\underline{m}_t) - \tilde{X}\tilde{X}^\top} & \frac{-(1+t\underline{m}_t)}{z(1+c_N t m_t)(1+t\underline{m}_t) - \tilde{X}^\top\tilde{X}} \end{bmatrix}.$$

We define the index sets

$$\mathcal{I}_1 := \{1, \cdots, M\}, \quad \mathcal{I}_2 := \{M+1, \cdots, M+N\}, \quad \mathcal{I} := \mathcal{I}_1 \cup \mathcal{I}_2.$$

In the sequel, we use the Latin letter $i, j \in \mathcal{I}_1$, Greek letters $\mu, \nu \in \mathcal{I}_2$, $\mathfrak{a}, \mathfrak{b} \in \mathcal{I}$. For an $\mathcal{I} \times \mathcal{I}$ matrix A and $i, j \in \mathcal{I}_1$, we define the 2×2 minor as

$$A_{[ij]} := \begin{pmatrix} A_{ij} & A_{i\bar{j}} \\ A_{\bar{i}j} & A_{\bar{i}\bar{j}} \end{pmatrix},$$

where $\bar{i} := i + M \in \mathcal{I}_2$. Moreover, for $\mathfrak{a} \in \mathcal{I} \setminus \{i, \bar{i}\}$, we denote

$$A_{[i]\mathfrak{a}} = \begin{pmatrix} A_{i\mathfrak{a}} \\ A_{\bar{i}\mathfrak{a}} \end{pmatrix}, \quad A_{\mathfrak{a}[i]} = (A_{\mathfrak{a}i}, A_{\mathfrak{a}\bar{i}}.)$$

Let the error parameter $\Psi(z)$ be defined as follows,

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_t}{N\eta}} + \frac{1}{N\eta}.$$

Instead of proving [6, Eq. (B.68) in Supplement], which aims at bounding $u^\top(\Pi^x(z))^{-1}[R(z) - \Pi^x(z)](\Pi^x(z))^{-1}v$ for any deterministic unit vector $u, v \in \mathbb{R}^{M+N}$, we shall prove

$$|u^\top[\mathcal{R}(z) - \Pi^x(z)]v| \prec t^{-3}\Psi(z) + \frac{t^{-7/2}}{N^{1/2}}. \tag{S.3.2}$$

We remark here that in [6], it is assumed that all $\lambda_i(\mathcal{S}(X))$'s are $O(1)$. Under this assumption, adding $(\Pi^x(z))^{-1}$ is harmless. However, in our case, $\lambda_i(\mathcal{S}(X))$ could diverge with N. Then, adding the $(\Pi^x(z))^{-1}$ factor which will blow up along with big $\lambda_i(\mathcal{S}(X))$, will complicate the proof of the anisotropic law. On the other hand, (S.3.2) is what we need anyway. Hence, we get rid of the $(\Pi^x(z))^{-1}$ and adapt the proof in [6] to our estimate (S.3.2). Without the $(\Pi^x(z))^{-1}$ factor, the $\mathcal{R}(z)$ and $\Pi^x(z)$ entries are well controlled, and the remaining proof is nearly the same as [6].

We shall first prove an entrywise version of (S.3.2): for any $\mathfrak{a}, \mathfrak{b} \in \mathcal{I}$,

$$|[\mathcal{R}(z) - \Pi^x(z)]_{\mathfrak{a}\mathfrak{b}}| \prec t^{-3}\Psi(z) + \frac{t^{-7/2}}{N^{1/2}}. \tag{S.3.3}$$

The derivation of (S.3.3) follows the same procedure as the proof of [6, Eq. (B.69) in Supplement]. This proof primarily relies on Schur complement, the large deviation of quadratic forms of Gaussian vector, and the fact that $\min_i |\lambda_i(\mathcal{S}(X)) - \zeta_t(z)| \gtrsim t^2$.

Then, for general u, v, analogous to [6, Eq. (B. 72) in Supplement], we have

$$\begin{aligned}|u^\top[\mathcal{R}(z) - \Pi^x(z)]v| \prec t^{-3}\Psi(z) + \frac{t^{-7/2}}{N^{1/2}} + \Big|\sum_{i\neq j} u_{[i]}^\top \mathcal{R}_{[ij]} u_{[j]}\Big| \\ + \Big|\sum_{\mu\neq\nu\geq 2M+1} u_\mu^\top \mathcal{R}_{\mu\nu} u_\nu\Big| + 2\Big|\sum_{i\in\mathcal{I}_1, \mu\geq 2M+1} u_{[i]}^\top \mathcal{R}_{[i]\mu} u_\mu\Big|.\end{aligned}$$

Therefore, it suffices to prove the following high moment bounds, for any $a \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{E}\Big|\sum_{i\neq j} u_{[i]}^\top \mathcal{R}_{[ij]} u_{[j]}\Big|^{2a} \prec \Big(t^{-3}\Psi(z) + \frac{t^{-7/2}}{N^{1/2}}\Big)^{2a},\\ &\mathbb{E}\Big|\sum_{\mu\neq\nu\geq 2M+1} u_\mu^\top \mathcal{R}_{\mu\nu} u_\nu\Big|^{2a} \prec \Big(t^{-3}\Psi(z) + \frac{t^{-7/2}}{N^{1/2}}\Big)^{2a},\\ &\mathbb{E}\Big|\sum_{i\in\mathcal{I}_1, \mu\geq 2M+1} u_{[i]}^\top \mathcal{R}_{[i]\mu} u_\mu\Big|^{2a} \prec \Big(t^{-3}\Psi(z) + \frac{t^{-7/2}}{N^{1/2}}\Big)^{2a}.\end{aligned}$$

The above estimates are proven using a polynomialization method outlined in [5, Section 5], with input from the entrywise estimates (S.3.3) and resolvent expansion (cf. [6, Lemma B.2 in Supplement]). We omit the details. □

Remark 1. *Actually, the estimates in Theorem S.3.1 hold uniformly in z such that*

$$\lambda_{-,t} - \vartheta^{-1}t^2 \leq \operatorname{Re} z \leq \lambda_{-,t} + \vartheta^{-1}, \quad \operatorname{Im} z \cdot \Big(t + \big(|\operatorname{Re} z - \lambda_{-,t}| + \operatorname{Im} z\big)^{1/2}\Big) \geq N^{-1+\vartheta}, \quad \operatorname{Im} z \leq \vartheta^{-1}, \tag{S.3.4}$$

for any $\vartheta > 0$. We can observe that every $z \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ satisfies (S.3.4) if $\epsilon_a, \varepsilon_1, \varepsilon_2$ and ϑ are sufficiently small. Also note that $b_t = \mathcal{O}(1)$ and $1 + t\underline{m}_t = \mathcal{O}(1)$ in the domain $\mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$.

By Theorem S.3.1 and Lemma 2.9, it is enough to analyze $G(X, \zeta)$ and $G(X^\top, \zeta)$ with $\zeta \in \mathsf{D}_\zeta$ in order to get the desired result. This was be done in Proposition 2.10. Together with Proposition S.3.2 and Corollary S.3.3 below, we complete the proof of Theorem 2.8.

Proposition S.3.2. *Suppose that the assumptions in Proposition 2.10 hold. The following estimates hold with respect to the probability measure $\mathbb{P}_\Psi$.*

(i) If $i \in \mathcal{T}_r$, we have

$$|[G(X,\zeta)X]_{ij}| \prec N^{-\epsilon_b/2}.$$

(ii) If $j \in \mathcal{T}_c$, we have

$$|[G(X,\zeta)X]_{ij}| \prec N^{-\epsilon_b/2}.$$

(iii) Otherwise, we have the crude bound

$$|[G(X,\zeta)X]_{ij}| \leq \|G(X,\zeta)X\| \lesssim t^{-2}.$$

Proof. Using Proposition 2.10, it follows from Proposition S.3.4 below. □

With the above bounds, we can further improve the bound of the off-diagonal Green function entries when i or j is typical index.

Corollary S.3.3. *Suppose that the assumptions in Proposition 2.10 hold. The following estimates hold with respect to the probability measure $\mathbb{P}_\Psi$.*

(i) If $i \neq j$ and $i \in \mathcal{T}_r$ (or $j \in \mathcal{T}_r$), there exists a constant $\delta = \delta(\epsilon_a, \epsilon_\alpha, \epsilon_b) > 0$ such that

$$|G_{ij}(X,\zeta)| \prec N^{-\delta}.$$

(ii) If $i \neq j$ and $i \in \mathcal{T}_c$ (or $j \in \mathcal{T}_c$), there exists a constant $\delta = \delta(\epsilon_a, \epsilon_\alpha, \epsilon_b) > 0$ such that

$$|G_{ij}(X^\top,\zeta)| \prec N^{-\delta}.$$

Proof of Corollary S.3.3. We shall give the proof only for the case $i \neq j$ and $i \in \mathcal{T}_r$. The other cases can be proved in the same way. Assume $i \neq j$ and $i \in \mathcal{T}_r$, observe that

$$|G_{ij}(X,\zeta)| = |G_{ii}(X,\zeta)| \cdot \left| \sum_{k,l} x_{ik} G_{kl}((X^{(i)})^\top, \zeta) x_{jl} \right|,$$

where we denote by $X^{(i)}$ the matrix obtained from X by removing i-th row. Note that

$$\sum_l G_{kl}((X^{(i)})^\top,\zeta) x_{jl} = [G((X^{(i)})^\top,\zeta)(X^{(i)})^\top]_{kj}.$$

Since $i \in \mathcal{T}_r$, we apply the large deviation estimates in [1, Corollary 25] to bound

$$\left| \sum_k x_{ik} [G((X^{(i)})^\top,\zeta)(X^{(i)})^\top]_{kj} \right|,$$

where we also use Proposition S.3.4 below to get a high probability bound for $\|G((X^{(i)})^\top,\zeta)(X^{(i)})^\top\|$. □

Proposition S.3.4. *Let $\zeta = E + \mathrm{i}\eta \in \mathbb{C}_+$.*

(i) If $i \in \mathcal{T}_r$, we have

$$\begin{aligned} |[G(X,\zeta)X]_{ij}| \prec & \left(N^{-\epsilon_b} \max_k |G_{kj}((X^{(i)})^\top,\zeta)| + \left(\frac{\operatorname{Im} G_{jj}((X^{(i)})^\top,\zeta)}{N\eta} \right)^{1/2} \right) \\ & \times \left(1 + |\zeta| \cdot |G_{ii}(X,\zeta)| \cdot \left(N^{-\epsilon_b} \max_{k,l} |G_{kl}((X^{(i)})^\top,\zeta)| + \left(\frac{\sum_k \operatorname{Im} G_{kk}((X^{(i)})^\top,\zeta)}{N^2\eta} \right)^{1/2} \right) \right), \end{aligned}$$

where we denote by $X^{(i)}$ the matrix obtained from X by removing i-th row.

(ii) If $j \in \mathcal{T}_c$, we have

$$|[G(X,\zeta)X]_{ij}| \prec \left(N^{-\epsilon_b} \max_k |G_{ik}(X^{[j]},\zeta)| + \left(\frac{\operatorname{Im} G_{ii}(X^{[j]},\zeta)}{N\eta} \right)^{1/2} \right)$$

$$\times\left(1+|\zeta|\cdot|G_{jj}(X^\top,\zeta)|\cdot\left(N^{-\epsilon_b}\max_{k,l}|G_{kl}(X^{[j]},\zeta)|+\left(\frac{\sum_k \operatorname{Im} G_{kk}(X^{[j]},\zeta)}{N^2\eta}\right)^{1/2}\right)\right),$$

where we denote by $X^{[j]}$ the matrix obtained from X by removing j-th column.

(iii) Let $X = UDV$ be a singular value decomposition of X where

$$diag(D) = (d_1, d_2, \cdots, d_p) \equiv \left(\sqrt{\lambda_1(\mathcal{S}(X))}, \sqrt{\lambda_2(\mathcal{S}(X))}, \cdots, \sqrt{\lambda_M(\mathcal{S}(X))}\right).$$

(Here we also assume $M < N$ without loss of generality.) Then,

$$\|G(X,\zeta)X\| \le \max_{1\le i\le p}\left|\frac{d_i}{d_i^2-\zeta}\right|.$$

Proof. (i) Assume $i \in \mathcal{T}_r$. Note that $G(X,\zeta)X = XG(X^\top,\zeta)$. Let $x_{(i)}$ be the i-th row of X. See that

$$X^\top X - \zeta = (X^{(i)})^\top X^{(i)} - \zeta + x_{(i)}^\top x_{(i)}.$$

By the Sherman-Morrison formula,

$$G(X^\top,\zeta) = G((X^{(i)})^\top,\zeta) - \frac{G((X^{(i)})^\top,\zeta)x_{(i)}^\top x_{(i)} G((X^{(i)})^\top,\zeta)}{1+x_{(i)}G((X^{(i)})^\top,\zeta)x_{(i)}^\top}.$$

Since $\big(G_{ii}(X,\zeta)\big)^{-1} = -\zeta\big(1+x_{(i)}G((X^{(i)})^\top,\zeta)x_{(i)}^\top\big)$,

$$G(X^\top,\zeta) = G((X^{(i)})^\top,\zeta) + (\zeta G_{ii}(X,\zeta))\cdot G((X^{(i)})^\top,\zeta)x_{(i)}^\top x_{(i)}G((X^{(i)})^\top,\zeta).$$

We write $[XG(X^\top,\zeta)]_{ij} = x_{(i)}G(X^\top,\zeta)e_j$. Then,

$$x_{(i)}G(X^\top,\zeta)e_j = x_{(i)}G((X^{(i)})^\top,\zeta)e_j + (\zeta G_{ii}(X,\zeta))\cdot(x_{(i)}G((X^{(i)})^\top,\zeta)x_{(i)}^\top)\cdot(x_{(i)}G((X^{(i)})^\top,\zeta)e_j).$$

Since $i \in \mathcal{T}_r$, by the large deviation estimate [1, Corollary 25], the desired result follows.

(ii) Assume $j \in \mathcal{T}_c$. Let $x_{[j]}$ be j-th column of X. See that

$$[G(X,\zeta)X]_{ij} = e_i^\top G(X,\zeta)x_{[j]}.$$

By the Sherman-Morrison formula,

$$G(X,\zeta) = G(X^{[j]},\zeta) + (\zeta G_{jj}(X^\top,\zeta))\cdot G(X^{[j]},\zeta)x_{[j]}x_{[j]}^\top G(X^{[j]},\zeta),$$

where we denote by $X^{[j]}$ the matrix obtained from X by removing j-th column. Then,

$$e_i^\top G(X,\zeta)x_{[j]} = e_i^\top G(X^{[j]},\zeta)x_{[j]} + (\zeta G_{jj}(X^\top,\zeta))\cdot(e_i^\top G(X^{[j]},\zeta)x_{[j]})\cdot x_{[j]}^\top G(X^{[j]},\zeta)x_{[j]}).$$

Using $j \in \mathcal{T}_c$, we get the desired result using the large deviation estimate [1, Corollary 25].

(iii) This is elementary, and thus we omit the details. □

4 Remark on Theorem 2.11

Theorem 2.11 is a version of [7, Theorem V.3] with respect to the left edge. The required modification would be straightforward. Let us summarize the main idea of [7] as follows. Let B_i $(i = 1, \cdots, M)$ be independent standard Brownian motions. We fix two time scales:

$$t_0 = N^{-\frac{1}{3}+\phi_0}, \quad t_1 = N^{-\frac{1}{3}+\phi_1}, \tag{S.4.1}$$

where $\phi_0 \in (\frac{1}{3} - \frac{\epsilon_b}{2}, \frac{1}{3})$ and $0 < \phi_1 < \frac{\phi_0}{100}$.

For time $t \geq 0$, we define the process $\{\lambda_i(t) : 1 \leq i \leq M\}$ as the unique strong solution to the following system of SDEs:

$$\mathrm{d}\lambda_i = 2\lambda_i^{1/2} \frac{\mathrm{dB}_i}{\sqrt{N}} + \left(\frac{1}{N} \sum_{j \neq i} \frac{\lambda_i + \lambda_j}{\lambda_i - \lambda_j} \right) \mathrm{d}t, \quad 1 \leq i \leq M,$$

with initial data $\lambda_i(0) = \lambda_i(\gamma_w \mathcal{S}(V_{t_0}))$ where γ_w is chosen to match the edge eigenvalue gaps of $\mathcal{S}(V_{t_0})$ with those of Wigner matrices. Recall the convention: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$.

Note that the process $\{\lambda_i(t)\}$ has the same joint distribution as the eigenvalues of the matrix

$$\gamma_w \mathcal{S}(V_{t_0 + \frac{t}{\gamma_w}}) = (\gamma_w^{1/2} X + (\gamma_w t_0 + t)^{1/2} W)(\gamma_w^{1/2} X + (\gamma_w t_0 + t)^{1/2} W)^\top.$$

Denote by $\rho_{\lambda,t}$ the asymptotic spectral distribution of $\mathcal{S}(V_{t_0 + \frac{t}{\gamma_w}})$ (in terms of the rectangular free convolution actually). Let $E_\lambda(t)$ be the left edge of $\rho_{\lambda,t}$. Now we introduce a deforemd Wishart matrix $\mathcal{U}\mathcal{U}^\top$. Define $\mathcal{U} := \Sigma^{1/2}\mathcal{X}$ where $\mathcal{X}$ is a $M \times N$ real Gaussian matrix (mean zero and variance N^{-1}) and $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_M)$ is a diagonal population matrix. Let $\rho_{\mu,0}$ be the asymptotic spectral distribution of $\mathcal{U}\mathcal{U}^\top$ (given by the multiplicative free convolution of the MP law and the ESD of Σ). We choose the diagonal population covariance matrix Σ such that $\rho_{\mu,0}$ matches $\rho_{\lambda,0}$ near the left edge $E_\lambda(0)$ (square-root behavior). We write $\mu_i(0) := \mu_i(\mathcal{U}\mathcal{U}^\top)$. Next, define the process $\{\mu_i(t) : 1 \leq i \leq M\}$ through the rectangular DBM with initial data $\{\mu_i(0)\}$. We can show that the edge eigenvalues of $\{\mu_i(t)\}$ are governed by the Tracy-Widom law. We denote by $\rho_{\mu,t}$ the rectangular free convolution of $\rho_{\mu,0}$ with the Marchenko-Pastur (MP) law at time t. Let $E_\mu(t)$ be the left edge of $\rho_{\mu,t}$. We remark that $E_\lambda(0) = E_\mu(0)$. Then, in order to get Theorem 2.11, it is enough to show

$$\left| \big(\lambda_M(t_1) - E_\lambda(t_1)\big) - \big(\mu_M(t_1) - E_\mu(t_1)\big) \right| \prec N^{-2/3-\delta},$$

for $\delta > 0$ sufficiently small. The proof of the above estimate relies on the local equilibrium mechanism of the rectangle DBM, which does not have any difference between the left edge or the right edge of the spectrum, given η_*-regularities of the initial states. Hence, we omit the remaining argument, and refer to [7] for details.

5 Proof of Lemma 3.2

We shall prove Lemma 3.2 in this section.

Proof of Lemma 3.2 (i). The proof is similar to that in [6], we provide proof here completeness. The statement $\zeta_{-,t} - \lambda_M(\mathcal{S}(X)) \leq 0$ follows directly from Lemma 3.1. For the other estimate, by Lemma 3.1, we know that $\Phi_t(\zeta_{-,t})$ is the only local extrema of $\Phi_t(\zeta)$ on the interval $(0, \lambda_M(\mathcal{S}(X)))$. Hence we have $\Phi_t'(\zeta_{-,t}) = 0$, which gives the equation

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

Rearranging the terms, we can get

$$c_N t m_X'(\zeta_{-,t}) = \frac{(1 - c_N t m_X(\zeta_{-,t}))^2}{2\zeta_{-,t} (1 - c_N t m_X(\zeta_{-,t})) + (1 - c_N) t}. \tag{S.5.1}$$

By Lemma 2.1 (iv) and Eq. (22), we have on Ω_Ψ that

$$c_N t m_X(\zeta_{-,t}) = \mathcal{O}(t^{1/2}). \tag{S.5.2}$$

Plugging the above bound back to (S.5.1), we can get $m'_X(\zeta_{-,t}) \sim t^{-1}$. This together with Lemma 2.2 gives $\sqrt{\lambda_M(\mathcal{S}(X)) - \zeta_{-,t}} \sim t$. ☐

Proof of Lemma 3.2 (ii). Since $\mathcal{S}(X)$ is η_*-regular in the sense of Definition 1, the estimates for $|m_X^{(k)}(\zeta)|$ on the event Ω_Ψ is an immediate consequence of Lemmas 2.2 and Lemma 3.2 (i).

We prove the estimate for $|m_X(z) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z)|$ as follows. Recall that $\beta = (\alpha-2)/24$. First, we establish the convergence of Stieltjes transform of a truncated matrix model using the result in [9]. To this end, let us define $\bar{X} = (\bar{x}_{ij}) := (x_{ij}\mathbf{1}_{x_{ij}<N^{-\beta}})$ and $\bar{t} := 1 - N\mathbb{E}|\bar{x}_{ij}|^2$. It is easy to show that $|\bar{t} - t| = \mathfrak{o}(N^{-1})$, and thus we have $|\mathsf{m}_{\mathsf{mp}}^{(t)}(z_1) - \mathsf{m}_{\mathsf{mp}}^{(\bar{t})}(z_1)| \le (N\eta_1)^{-1}$. Then it follows from [9, Theorem 2.7] that for any z_1 such that $|z_1 - \zeta_{-,t}| \le \tau t^2$ and $\eta_1 \equiv \operatorname{Im} z_1 > N^{-1+\delta}$ with $1 > \delta > 0$ to be chosen later,

$$m_{\bar{X}}(z_1) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z_1) \prec \frac{1}{N^\beta} + \frac{1}{N\eta_1}, \tag{S.5.3}$$

We remark here that the local law proved in [9, Theorem 2.7] is for deterministic z. But it is easy to show that the local law holds uniformly in z in the mentioned domain in [9, Theorem 2.7], with high probability, by a simple continuity argument. Hence, as long as z_1 fall in this domain with high probability, even though z_1 might be random, we still have (S.5.3). Using the facts $|\lambda_M(\mathcal{S}(X)) - (1-t)\lambda_-^{\mathsf{mp}}| \lesssim N^{-\epsilon_b}$ and $\lambda_M(\mathcal{S}(X)) - \zeta_{-,t} \sim t^2$ with high probability (cf. Lemmas 2.6 and 3.2 (i)), we have for τ small enough,

$$|z_1 - (1-t)\lambda_-^{\mathsf{mp}}| \ge |\zeta_{-,t} - \lambda_M(\mathcal{S}(X))| - |\lambda_M(\mathcal{S}(X)) - (1-t)\lambda_-^{\mathsf{mp}}| - |z_1 - \zeta_{-,t}| \gtrsim t^2,$$

which gives $|(\mathsf{m}_{\mathsf{mp}}^{(t)})'(z_1)| \lesssim t^{-4}$ with high probability. Also, we have $|m'_{\bar{X}}(z_1)| \lesssim t^{-4}$ with high probability, by the choice of z_1, Eq. (28), and Lemma 3.2 (i). Therefore, for any z_2 satisfying $\operatorname{Re} z_2 = \operatorname{Re} z_1$ and $\eta_2 = \operatorname{Im} z_2 < N^{-1+\delta}$, we have

$$|m_{\bar{X}}(z_2) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z_2)| \lesssim |m_{\bar{X}}(z_1) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z_1)| + t^{-4}|z_1 - z_2| \prec \frac{1}{N^\beta} + \frac{1}{N^{1/2}} + \frac{1}{t^4 N^{1/2}} \lesssim \frac{1}{N^\beta}, \tag{S.5.4}$$

where in the first step we used the fact $|z_i - \zeta_{-,t}| \le \tau t^2, i = 1, 2$, and in the second step we chose $\delta = 1/2$.

Next, we use the rank inequality to compare $m_{\bar{X}}(z)$ with $m_X(z)$. Notice that

$$m_{\bar{X}}(z) - m_X(z) \le \frac{2}{N}\operatorname{Rank}(\bar{X} - X) \cdot (\|(\mathcal{S}(\bar{X}) - z)^{-1}\| + \|(\mathcal{S}(X) - z)^{-1}\|) \prec \frac{\operatorname{Rank}(\bar{X} - X)}{Nt^2}.$$

A similar argument as in the proof of Lemma 2.3 shows that,

$$\operatorname{Rank}(\bar{X} - X) \prec N^{1-(\alpha-2-2\alpha\beta)/4}.$$

Therefore, we can obtain $m_{\bar{X}}(z) - m_X(z) \prec N^{-(\alpha-2-2\alpha\beta)/4}t^{-2}$. Together with the estimate in (S.5.4), we have

$$m_X(z) - \mathsf{m}_{\mathsf{mp}}^{(t)}(z) \prec \frac{1}{N^{(\alpha-2-2\alpha\beta)/4}t^2} + \frac{1}{N^\beta}.$$

The claim now follows by the fact $t \gg N^{(2-\alpha)/16}$ in light of Eq. (3). ☐

Proof of Lemma 3.2 (iii). Repeating the proof of [6, Lemma A.2], we can obtain

$$|\bar{\zeta}_{-,t} - \zeta_{-,t}| \lesssim t^3 |m'_X(\zeta_{-,t}) - (\mathsf{m}_{\mathsf{mp}}^{(t)})'(\zeta_{-,t})|.$$

By the Cauchy integral formula, we have

$$|m'_X(\zeta_{-,t}) - (\mathsf{m}_{\mathsf{mp}}^{(t)})'(\zeta_{-,t})| \lesssim \oint_\omega \frac{|m_X(a) - \mathsf{m}_{\mathsf{mp}}^{(t)}(a)|}{|a-\zeta_{-,t}|^2}\mathrm{d}a, \tag{S.5.5}$$

where $\omega \equiv \{a : |a - \zeta_{-,t}| = \tau t^2\}$ for some small τ. Therefore, we have by Lemma 3.2 (ii),

$$|\bar{\zeta}_{-,t} - \zeta_{-,t}| \lesssim t \sup_{a\in\omega} |m_X(a) - \mathsf{m}_{\mathsf{mp}}^{(t)}(a)| \prec tN^{-\beta},$$

proving the claim. □

6 Proof of Proposition 3.10

In this section, we shall give the proof of Proposition 3.10.

Proof of Proposition 3.10. By a minor process argument, we have with probability at least $1 - N^{-D}$ for arbitrary large D, there exists constant $C_k > 0$, such that

$$\begin{aligned}|\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - \hat{\zeta}_{\mathsf{e}}| &= \left|(1-t)\lambda_-^{\mathsf{mp}} - \bar{\zeta}_{-,t} + \lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - (1-t)\lambda_-^{\mathsf{mp}} + \mathrm{i}N^{-100K} + \bar{\zeta}_{-,t} - \hat{\zeta}_{\mathsf{e}}\right| \\ &\geq \sqrt{c_N}t^2 - |\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - (1-t)\lambda_-^{\mathsf{mp}}| - |\bar{\zeta}_{-,t} - \hat{\zeta}_{\mathsf{e}}| - N^{-100K} \geq C_k t^2. \end{aligned} \tag{S.6.1}$$

Here in the last step, we used Eq. (45) and the fact that $|\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - (1-t)\lambda_-^{\mathsf{mp}}| \prec N^{-\epsilon_b}$. Therefore, for any $k \in [N]$, we can define the event $\Omega_k \equiv \{\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - \bar{\zeta}_{-,t} \geq C_k t^2\}$ with $\mathbb{P}(\Omega_k) \geq 1 - N^{-D}$ for arbitrary large D.

Choosing $\tau \leq \min_k C_k/2$. For any ζ satisfying $|\zeta - \hat{\zeta}_{\mathsf{e}}| \leq \tau t^2$, we define

$$F_k(\zeta) := \log|1 + \tilde{x}_k^\top(G(\tilde{X}^{(k)},\zeta))\tilde{x}_k|^2, \quad \tilde{F}_k(\zeta) := \log|1 + \tilde{x}_k^\top(G(\tilde{X}^{(k)},\zeta))_{\mathrm{diag}}\tilde{x}_k|^2.$$

Since $|\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - \zeta| = |\lambda_M(\mathcal{S}(\tilde{X}^{(k)})) - \hat{\zeta}_{\mathsf{e}}| - |\zeta - \hat{\zeta}_{\mathsf{e}}| \geq C_k t^2/2 > 0$ on Ω_k, we can obtain that $\mathrm{Re}\,(\tilde{x}_k^\top(G(\tilde{X}^{(k)},\zeta))\tilde{x}_k) \vee \mathrm{Re}\,(\tilde{x}_k^\top(G(\tilde{X}^{(k)},\zeta))_{\mathrm{diag}}\tilde{x}_k) \geq 0$. Hence, the functions $F_k(\zeta)$, $\tilde{F}_k(\zeta)$ are well defined on the event Ω_k. For any $\zeta \in \Xi(\tau)$, using Cauchy integral formula with a cutoff of the contour chosen carefully, we can express $Y_k \equiv Y_k(\zeta)$ as

$$Y_k = \frac{t}{2\pi\mathrm{i}N^{1-\alpha/4}}(\mathbb{E}_k - \mathbb{E}_{k-1})\oint_{\omega\cap\gamma}\frac{F_k(z)}{(z-\zeta)^2}\mathrm{d}z + \mathsf{err}_k(\zeta) =: I_k(\zeta) + \mathsf{err}_k(\zeta),$$

with the contour $\omega \equiv \{z \in \mathbb{C} : |z - \zeta| = \tau t^2/10\}$ and $\gamma \equiv \{z \in \mathbb{C} : |\mathrm{Im}\, z| \geq N^{-100}\}$, and err_k collects all the tiny error terms which will not affect our further analysis. Similarly, we can define $\tilde{I}_k(\zeta)$ and $\tilde{\mathsf{err}}_k(\zeta)$ for $\tilde{Y}_k$ in the same manner as shown above. Therefore,

$$\mathbb{E}_{k-1}(Y_kY'_k) - \mathbb{E}_{k-1}(\tilde{Y}_k\tilde{Y}'_k) = \mathbb{E}_{k-1}((I_k(\zeta)I_k(\zeta')) - \mathbb{E}_{k-1}((\tilde{I}_k(\zeta)\tilde{I}_k(\zeta')) + \mathsf{HOT},$$

where HOT collects terms containing $\mathsf{err}_k(\zeta)$ or $\tilde{\mathsf{err}}_k(\zeta)$, which are irrelevant in our analysis. For the leading term, since $F_k(z)$, $\tilde{F}_k(z)$,$\tilde{F}_k(z)$,$\tilde{F}_k(z')$ are uniformly bounded on $z \in \omega\cap\gamma$ and $z' \in \omega'\cap\gamma$, we may commute the conditional expectation and the integral to obtain

$$\mathbb{E}_{k-1}((I_k(\zeta)I_k(\zeta')) - \mathbb{E}_{k-1}((\tilde{I}_k(\zeta)\tilde{I}_k(\zeta')) = -\frac{t^2}{4\pi^2N^{2-\alpha/2}}\oint_{\omega\cap\gamma}\oint_{\omega'\cap\gamma}\frac{\varphi_k(z,z') - \tilde{\varphi}_k(z,z')}{(z-\zeta)^2(z'-\zeta')^2}\mathrm{d}z'\mathrm{d}z, \tag{S.6.2}$$

where

$$\begin{aligned}\varphi_k(z,z') &:= \mathbb{E}_{k-1}\big((\mathbb{E}_k - \mathbb{E}_{k-1})F_k(z)(\mathbb{E}_k - \mathbb{E}_{k-1})F_k(z')\big) \\ \tilde{\varphi}_k(z,z') &:= \mathbb{E}_{k-1}\big((\mathbb{E}_k - \mathbb{E}_{k-1})\tilde{F}_k(z)(\mathbb{E}_k - \mathbb{E}_{k-1})\tilde{F}_k(z')\big),\end{aligned}$$

and $\omega' := \{z \in \mathbb{C} : |z - \zeta'| = at^2\}$ with a small constant a.

In view of (S.6.2), it suffices to prove that uniformly on $z \in \omega \cap \gamma$ and $z' \in \omega' \cap \gamma$, $\varphi_k - \tilde{\varphi}_k \equiv \varphi_k(z, z') - \tilde{\varphi}_k(z, z') \ll t^2 N^{1-\alpha/2}$. In the sequel, we write $F_k = F_k(z)$, $\tilde{F}_k = \tilde{F}_k(z)$, $F_k' = F_k(z')$, and $\tilde{F}_k' = \tilde{F}_k(z')$ for simplicity. Let

$$\eta_k = \eta_k(z) := \tilde{x}_k^\top (G(X^{(k)}, z))\tilde{x}_k - \tilde{x}_k^\top (G(\tilde{X}^{(k)}, z))_{\mathrm{diag}} \tilde{x}_k = \sum_{i \neq j} [G(\tilde{X}^{(k)}, z)]_{ij} \tilde{x}_{ik} \tilde{x}_{jk},$$

and

$$\varepsilon_k = \varepsilon_k(z) := F_k - \tilde{F}_k = \log |1 + \eta_k (1 + \tilde{x}_k^\top (G(\tilde{X}^{(k)}, z))_{\mathrm{diag}} \tilde{x}_k)^{-1}|^2.$$

We also write $\eta_k' \equiv \eta_k(z')$ and $\varepsilon_k' \equiv \varepsilon_k(z')$. Using the following elementary identity,

$$\mathbb{E}_{k-1}\big((\mathbb{E}_k - \mathbb{E}_{k-1})(A)(\mathbb{E}_k - \mathbb{E}_{k-1})(B)\big) = \mathbb{E}_{k-1}\big(\mathbb{E}_k(A)\mathbb{E}_k(B)\big) - \mathbb{E}_{k-1}(A)\mathbb{E}_{k-1}(B),$$

we may rewrite φ_k and $\tilde{\varphi}_k$ as

$$\begin{aligned}
\varphi_k &= \mathbb{E}_{k-1}\big(\mathbb{E}_k(F_k)\mathbb{E}_k(F_k')\big) - \mathbb{E}_{k-1}(F_k)\mathbb{E}_{k-1}(F_k'),\\
\tilde{\varphi}_k &= \mathbb{E}_{k-1}\big(\mathbb{E}_k(\tilde{F}_k)\mathbb{E}_k(\tilde{F}_k')\big) - \mathbb{E}_{k-1}(\tilde{F}_k)\mathbb{E}_{k-1}(\tilde{F}_k').
\end{aligned}$$

Therefore, let $E_{\tilde{x}_k}$ denote the expectation with respect to the randomness of k-th column of $\tilde{X}$, we have by the definitions of ε_k, ε_k',

$$\begin{aligned}
\varphi_k - \tilde{\varphi}_k &= \mathbb{E}_{\tilde{x}_k}\big(\mathbb{E}_k(\tilde{F}_k)\mathbb{E}_k(\varepsilon_k')\big) + \mathbb{E}_{\tilde{x}_k}\big(\mathbb{E}_k(\tilde{F}_k')\mathbb{E}_k(\varepsilon_k)\big) + \mathbb{E}_{\tilde{x}_k}\big(\mathbb{E}_k(\varepsilon_k)\mathbb{E}_k(\varepsilon_k')\big)\\
&\quad - \mathbb{E}_k\mathbb{E}_{\tilde{x}_k}(\tilde{F}_k)\mathbb{E}_k\mathbb{E}_{\tilde{x}_k}(\varepsilon_k') - \mathbb{E}_k\mathbb{E}_{\tilde{x}_k}(\tilde{F}_k')\mathbb{E}_k\mathbb{E}_{x_k}(\varepsilon_k) - \mathbb{E}_k\mathbb{E}_{\tilde{x}_k}(\varepsilon_k')\mathbb{E}_k\mathbb{E}_{\tilde{x}_k}(\varepsilon_k)\\
&\equiv T_1 + T_2 + T_3 + T_4 + T_5 + T_6.
\end{aligned}$$

Before bounding T_i's, $1 \le i \le 6$, we introduce some shorthand notation for simplicity. Let

$$J_k = J_k(z) := \frac{1}{1 + \tilde{x}_k^\top G(\tilde{X}^{(k)}, z)\tilde{x}_k}, \quad J_{k,\mathrm{diag}} = J_{k,\mathrm{diag}}(z) := \frac{1}{1 + \tilde{x}_k^\top (G(\tilde{X}^{(k)}, z))_{\mathrm{diag}}\tilde{x}_k},$$

and $J_k' = J_k(z')$, $J_{k,\mathrm{diag}}' = J_{k,\mathrm{diag}}(z')$. Further set

$$J_{k,\mathrm{Tr}} := \frac{1}{1 + \frac{\sigma_N^2}{N}\mathrm{Tr} G(\tilde{X}^{(k)}, z)}, \qquad \mathcal{E} := \tilde{x}_k^\top (G(\tilde{X}^{(k)}, z))_{\mathrm{diag}}\tilde{x}_k - \frac{\sigma_N^2}{N}\mathrm{Tr} G(\tilde{X}^{(k)}, z).$$

This gives $J_{k,\mathrm{diag}} = J_{k,\mathrm{Tr}} - \mathcal{E} J_{k,\mathrm{Tr}} J_{k,\mathrm{diag}}$. We may now establish an upper bound for $\mathbb{E}_{\tilde{x}_k}(\varepsilon_k)$ as follows:

$$\begin{aligned}
\mathbb{E}_{\tilde{x}_k}(\varepsilon_k) &= \mathbb{E}_{\tilde{x}_k} \log |1 + \eta_k J_{k,\mathrm{diag}}|^2 \overset{\text{(i)}}{\le} \log \mathbb{E}_{\tilde{x}_k} |1 + \eta_k J_{k,\mathrm{diag}}|^2\\
&= \log \mathbb{E}_{\tilde{x}_k} (1 + 2\mathrm{Re}\,(\eta_k J_{k,\mathrm{Tr}} - \eta_k E J_{k,\mathrm{Tr}} J_{k,\mathrm{diag}}) + |\eta_k J_{k,\mathrm{diag}}|^2)\\
&\overset{\text{(ii)}}{\le} \log\big(1 + \mathcal{O}(\mathbb{E}_{\tilde{x}_k}(|\eta_k||\mathcal{E}|)) + \mathcal{O}(\mathbb{E}_{\tilde{x}_k}(|\eta_k|^2))\big),
\end{aligned}$$

where, in (i), Jensen's inequality is applied, and in (ii), we used the fact that $J_{k,\mathrm{Tr}}$ and $J_{k,\mathrm{diag}}$ are uniformly bounded for $\zeta \in \Xi$ on the event Ω_k. Similarly, using the identity $|1 + \eta_k J_{k,\mathrm{diag}}||1 - \eta_k J_k| = 1$, we have

$$\begin{aligned}
\mathbb{E}_{\tilde{x}_k}(-\varepsilon_k) &= \mathbb{E}_{\tilde{x}_k} \log |1 - \eta_k J_k|^2 = \mathbb{E}_{\tilde{x}_k} \log |1 - \eta_k J_{k,\mathrm{Tr}} - \eta_k(\eta_k + \mathcal{E}) J_{k,\mathrm{diag}} J_k|^2\\
&\le \log\big(1 + \mathcal{O}(\mathbb{E}_{\tilde{x}_k}(|\eta_k||\mathcal{E}|)) + \mathcal{O}(\mathbb{E}_{\tilde{x}_k}(|\eta_k|^2))\big).
\end{aligned}$$

By the Cauchy-Schwarz inequality and Lemma S.6.1 and Lemma S.6.2,

$$\mathbb{E}_{\tilde{x}_k}(|\eta_k||\mathcal{E}|) \le \sqrt{\mathbb{E}_{\tilde{x}_k}(|\eta_k|^2) \cdot \mathbb{E}_{\tilde{x}_k}(|\mathcal{E}|^2)} \lesssim N^{-1/2} t^{-2} N^{\vartheta(2-\alpha/2)-1/2} \|G(\tilde{X}^{(k)}, z)\|$$

Since $\vartheta = 1/4 + 1/\alpha + \epsilon_\vartheta > 1/4 + 1/\alpha$, and recall that $\|G(\tilde{X}^{(k)}, z)\| \le |\lambda_1(\mathcal{S}(\tilde{X}^{(k)})) - z|^{-1} \lesssim t^{-2}$ on Ω_k, the above bound can be further simplified as

$$\mathbb{E}_{\tilde{x}_k}(|\eta_k||\mathcal{E}|) \lesssim N^{1-\alpha/2} \cdot N^{2/\alpha+3\alpha/8+\epsilon_\vartheta(4-\alpha)/2-2} t^{-4}.$$

By the facts $\epsilon_\vartheta < (3\alpha-5)/(4\alpha)$ and $t \gg N^{(\alpha-4)/48}$, it can be verified that $\mathbb{E}_{\tilde{x}_k}(|\eta_k||\mathcal{E}|) \ll t^2 N^{1-\alpha/2}$. Therefore, we can conclude that $|\mathbb{E}_{\tilde{x}_k}(\varepsilon_k)| \ll t^2 N^{1-\alpha/2}$. This shows $|T_6| \ll t^2 N^{1-\alpha/2}$. Together with the crude bound $\tilde{F}_k \le \log|1 + N^{2\vartheta}\|G(\tilde{X}^{(k)}, z)\||^2 \lesssim \log N$, we have $|T_4|, |T_5| \ll t^2 N^{1-\alpha/2}$.

For $|T_3|$, by Cauchy-Schwarz inequality, it suffices to give a bound on $\mathbb{E}_{\tilde{x}_k}\big(|\mathbb{E}_k(\varepsilon_k)|^2\big)$. By Jensen's inequality,

$$\mathbb{E}_{\tilde{x}_k}\big(|\mathbb{E}_k(\varepsilon_k)|^2\big) \le \mathbb{E}_k \mathbb{E}_{\tilde{x}_k}(|\varepsilon_k|^2).$$

Using again the identity $|1 + \eta_k J_{k,\mathrm{diag}}||1 - \eta_k J_k| = 1$,

$$\begin{aligned}
&|\log|1 + \eta_k J_{k,\mathrm{diag}}|^2| = \mathbf{1}_{\{|1+\eta_k J_{k,\mathrm{diag}}|\ge 1\}} \log|1 + \eta_k J_{k,\mathrm{diag}}|^2 + \mathbf{1}_{\{|1-\eta_k J_k|>1\}} \log|1 - \eta_k J_k|^2 \\
&= \mathbf{1}_{\{|1+\eta_k J_{k,\mathrm{diag}}|>1\}} \log(1 + 2\mathrm{Re}\,(\eta_k J_{k,\mathrm{diag}}) + |\eta_k J_{k,\mathrm{diag}}|^2) + \mathbf{1}_{\{|1-\eta_k J_k|>1\}} \log(1 - 2\mathrm{Re}\,(\eta_k J_k) + |\eta_k J_k|^2) \\
&\le \mathbf{1}_{\{|1+\eta_k J_{k,\mathrm{diag}}|>1\}} \big(2\mathrm{Re}\,(\eta_k J_{k,\mathrm{diag}}) + |\eta_k J_{k,\mathrm{diag}}|^2\big) + \mathbf{1}_{\{|1-\eta_k J_k|>1\}} \big(- 2\mathrm{Re}\,(\eta_k J_k) + |\eta_k J_k|^2\big).
\end{aligned}$$

Therefore, with the fact that $|\eta_k J_{k,\mathrm{diag}}| \le N^C$ for some $C > 0$,

$$\mathbb{E}_{\tilde{x}_k}|\log|1 + \eta_k J_{k,\mathrm{diag}}|^2|^2 \lesssim \log N \cdot \mathbb{E}_{\tilde{x}_k} \log|1 + \eta_k J_{k,\mathrm{diag}}|^2 \lesssim \log N \cdot \mathbb{E}_{\tilde{x}_k}(|\eta_k|^2) \lesssim N^{-1} t^{-5},$$

which gives $|T_3| \ll t^2 N^{1-\alpha/2}$ by the fact $t \gg N^{-2/7+\alpha/14}$.

To evaluate $|T_2|$, we start by expressing it as follows:

$$\begin{aligned}
T_2 &= \mathbb{E}_{\tilde{x}_k}\big(\mathbb{E}_k(\varepsilon_k)\mathbb{E}_k(\log|1 + N^{-1}\sigma_N^2 \mathrm{Tr} G(\tilde{X}^{(k)}, z') + \mathcal{E}|^2)\big) \\
&= \mathbb{E}_{\tilde{x}_k}\big(\mathbb{E}_k(\varepsilon_k)\mathbb{E}_k(\log|1 + N^{-1}\sigma_N^2 \mathrm{Tr} G(\tilde{X}^{(k)}, z')|^2)\big) + \mathbb{E}_{\tilde{x}_k}\big(\mathbb{E}_k(\varepsilon_k)\mathbb{E}_k(\log|1 + \mathcal{E} J_{k,\mathrm{Tr}}|^2)\big).
\end{aligned}$$

First, we use the fact that $\log|1 + N^{-1}\sigma_N^2 \mathrm{Tr} G(\tilde{X}^{(k)}, z)|^2$ is independent of $\tilde{x}_k$ and that $\mathbb{E}_{\tilde{x}_k}(\varepsilon_k) = 0$ to obtain the inequality

$$T_2 \lesssim \mathbb{E}_{\tilde{x}_k}\big(|\mathbb{E}_k(\varepsilon_k)| \cdot \mathbb{E}_k|\mathcal{E}|\big).$$

Next, we apply the Cauchy-Schwarz inequality to obtain

$$T_2 \le \sqrt{\mathbb{E}_{\tilde{x}_k}\big(|\mathbb{E}_k(\varepsilon_k)|^2\big)\mathbb{E}_{\tilde{x}_k}\big(|\mathbb{E}_k(|\mathcal{E}|)|^2\big)}.$$

Finally, by Jensen's inequality, we have

$$T_2 \le \sqrt{\mathbb{E}_k \mathbb{E}_{\tilde{x}_k}\big(|\varepsilon_k|^2\big)\mathbb{E}_k \mathbb{E}_{\tilde{x}_k}\big(|\mathcal{E}|^2\big)} \le N^{1-\alpha/2} \cdot N^{2/\alpha+3\alpha/8+\epsilon_\vartheta(4-\alpha)/2-2} t^{-9/2}.$$

The bound $|T_2| \ll t^2 N^{1-\alpha/2}$ follows by the facts $\epsilon_\vartheta < (3\alpha-5)/(4\alpha)$ and $t \gg N^{(\alpha-4)/56}$. The same bound holds for $|T_1|$. Therefore, we can obtain that for any $z \in \omega \cap \gamma$ and $z' \in \omega' \cap \gamma$, $|\varphi_k - \tilde{\varphi}_k| \ll t^2 N^{1-\alpha/2}$, which conludes the proof. □

Lemma S.6.1 ([4], Lemma 4.1)**.** *Let $a \equiv (a_1, \cdots, a_N)^\top$ be a column vector whose entries are i.i.d. centered and satisfy (ii) and (iii) in Lemma 3.12. Then for deterministic matrix G, the random variables*

$$X \equiv \sum_{i \ne j} G_{ij} a_i a_j, \quad E \equiv \sum_i G_{ii} a_i^2 - \frac{1}{N}\mathrm{Tr} G$$

satisfy

$$\mathbb{E}|X|^2 \le 2N^{-1}\|G\|^2, \quad \mathbb{E}|E|^2 \le 10C(\|G\|^2 + 1)N^{\vartheta(4-\alpha)-1}.$$

The following lemma is a directly consequence of Lemma S.6.1.

Lemma S.6.2. *Fix $C>0$. For any $\zeta\in\{\xi\in\mathbb{C}:|\xi-\bar{\zeta}_{-,t}|\le Ct^2\}$, we have there exist constant $\tau=\tau(C)$ such that $\mathbb{E}_{\tilde{x}_k}(|\eta_k|^2)\le\tau^{-2}N^{-1}t^{-4}$ on the event $\Omega_k=\{\lambda_1(\mathcal{S}(\tilde{X}^{(k)}))-\bar{\zeta}_{-,t}\ge\tau t^2\}$.*

7 Proof of Lemma 5.1

We need the following lemma on the monotonicity of the Green function to the linearization of $\mathcal{S}(Y^\gamma)$.

Lemma S.7.1 ([3], Lemma 2.1)**.** *For deterministic matrix $A\in\mathbb{R}^{M\times N}$, let $\mathcal{L}(A)$ be defined as in Eq. (76) Further define $\Gamma(z):=\max_{i,j\in[M+N]}[(\mathcal{L}(A)-z)^{-1}]_{ij}\vee 1$. We have for any $L>1$ and $z\in\mathbb{C}^+$, we have $\Gamma(E+\mathrm{i}\eta/L)\le L\Gamma(E+\mathrm{i}\eta)$.*

Recall that for any $\delta>0$, $z=E+\mathrm{i}\eta\in\mathsf{D}$,

$$\mathfrak{P}_0(\delta,z,\Psi)=\mathbb{P}_\Psi\Big(\sup_{\substack{a,b\in[M]\\0\le\gamma\le1}}|z^{1/2}\mathfrak{X}_{ab}[G^\gamma(z)]_{ab}|>N^\delta\Big),$$

$$\mathfrak{P}_1(\delta,z,\Psi)=\mathbb{P}_\Psi\Big(\sup_{\substack{u,v\in[N]\\0\le\gamma\le1}}|z^{1/2}\mathfrak{Y}_{uv}[\mathcal{G}^\gamma(z)]_{uv}|>N^\delta\Big),$$

$$\mathfrak{P}_2(\delta,z,\Psi)=\mathbb{P}_\Psi\Big(\sup_{\substack{a\in[M],u\in[N]\\0\le\gamma\le1}}|\mathfrak{Z}_{au}[G^\gamma(z)Y^\gamma]_{au}|>N^\delta\Big).$$

Now let us give the proof of Lemma 5.1.

Proof of Lemma 5.1. Let p be any sufficiently large (but fixed) integer, and $F_p(x):=|x|^{2p}+1$. It can be easily verified that there exists a constant C_p, only depends on p such that $|F_p^{(a)}(x)|\le C_pF_p(x)$, for all $x\in\mathbb{R}$ and $a\in\mathbb{Z}^+$. Recall Theorem 4.2, and we will focus on the case when $(\#_1,\#_2,\#_3)=(\mathfrak{X}_{ab}\mathrm{Im}\,[G^\gamma(z)]_{ab},\ \mathfrak{X}_{ab}\mathrm{Im}\,[G^0(z)]_{ab},\ \mathfrak{I}_{0,ab})$ therein. Applying Theorem 4.2 with $F(x)=F_p(x)$, we have for any $a,b\in[M]$, there exists constant $C_1>0$ such that,

$$\mathbb{E}_\Psi\big(F_p(\mathfrak{X}_{ab}\mathrm{Im}\,[G^\gamma(z)]_{ab})\big)-\mathbb{E}_\Psi\big(F_p(\mathfrak{X}_{ab}\mathrm{Im}\,[G^0(z)]_{ab})\big)<C_1N^{-\omega}(\mathfrak{I}_{p,0}+1)+C_1Q_0N^{C_1},$$

where $\mathfrak{I}_{p,0}\equiv\sup_{i,j\in[M],0\le\gamma\le1}\mathbb{E}_\Psi\big(\big|F_p(\mathfrak{X}_{ij}\mathrm{Im}\,[G^\gamma(z)]_{ij})\big|\big)$. Taking supremum over $a,b\in[M]$ and $0\le\gamma\le 1$ yields

$$(1-C_1N^{-\omega})\mathfrak{I}_{p,0}\le\max_{i,j\in[M]}\mathbb{E}_\Psi\big(F_p(\mathfrak{X}_{ij}\mathrm{Im}\,[G^0(z)]_{ij})\big)+C_1N^{-\omega}+3C_1N^{C_1}\max_{k\in[0:2]}\mathfrak{P}_k(\varepsilon,z,\Psi).$$

Applying Lemma S.7.1 on $\mathcal{R}(Y^\gamma,z)=z^{-1/2}(\mathcal{L}(Y^\gamma)-z^{1/2})^{-1}$ with $z^{1/2}=\tilde{E}+\mathrm{i}\tilde{\eta}$, we have,

$$\max_{i,j\in[M+N]}|z^{1/2}[\mathcal{R}(Y^\gamma,z)]_{ij}|\vee 1\le L\Big(\max_{i,j\in[M+N]}|(z')^{1/2}[\mathcal{R}(Y^\gamma,z')]_{ij}|\vee 1\Big),$$

for any $L>0$ and $z'\in\mathbb{C}^+$ satisfies $(z')^{1/2}=\tilde{E}+\mathrm{i}L\tilde{\eta}$. Let $L\equiv N^{\varepsilon/6}$ and thus $(z')^{1/2}\equiv\tilde{E}+\mathrm{i}N^{\varepsilon/6}\tilde{\eta}$, to obtain

$$\max_{i,j\in[M]}|z^{1/2}[G^\gamma(z)]_{ij}|\vee 1,\ \max_{i,j\in[N]}|z^{1/2}[\mathcal{G}^\gamma(z)]_{ij}|\vee 1,\ \max_{i\in[M],j\in[N]}|[G^\gamma(z)Y^\gamma]_{ij}|\le\mathfrak{S},$$

where

$$\mathfrak{S}\equiv N^{\varepsilon/6}\Big(\max_{i,j\in[M]}|(z')^{1/2}[G^\gamma(z')]_{ij}|\vee\max_{i,j\in[N]}|(z')^{1/2}[\mathcal{G}^\gamma(z')]_{ij}|\vee\max_{i\in[M],j\in[N]}|[G^\gamma(z')Y^\gamma]_{ij}|\vee 1\Big).$$

This implies that

$$\max_{k\in[0:2]} \mathfrak{P}_k(\varepsilon, z, \Psi) \le \max_{k\in[0:2]} \mathfrak{P}_k(\varepsilon/2, z', \Psi). \tag{S.7.1}$$

For any $z_0 = E_0 + \mathrm{i}\eta_0 \in \mathsf{D}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, we have $\mathbb{E}_\Psi\big(F_p(\mathfrak{X}_{ij}\mathrm{Im}\,[G^0(z_0)]_{ij})\big) \lesssim N$ (cf. Theorem 2.8). Then there exists some large constant $C_2 > 0$ such that

$$\mathfrak{I}_{p,0} \le C_2 N + C_2 N^{C_2} \max_{k\in[0:2]} \mathfrak{P}_k(\varepsilon, z_0, \Psi).$$

Using (S.7.1) by setting $z \equiv z_0$, we have for $z_1 = E_1 + \mathrm{i}\eta_1$ where (E_1, η_1) are defined through (73),

$$\mathfrak{I}_{p,0} \le C_2 N + C_2 N^{C_2} \max_{k\in[0:2]} \mathfrak{P}_k(\varepsilon/2, z_1, \Psi).$$

For any $a, b \in [M]$, and $0 \le \gamma \le 1$, applying Markov's inequality with the fact that $p\delta > D + 100$, we have that there exists some large constant $C_3 > 0$ such that

$$\begin{aligned}\mathbb{P}_\Psi\Big(|z_0^{1/2}\mathfrak{X}_{ab}[\mathrm{Im}\,G^\gamma(z_0)]_{ab}| > N^\delta\Big) &\le \frac{|z_0|^{p/2}\mathbb{E}_\Psi\big(\big|F_p(\mathfrak{X}_{ab}\mathrm{Im}\,[G^\gamma(z_0)]_{ab})\big|\big)}{N^{p\delta}} \le \frac{|z_1|^{p/2}\mathfrak{I}_{p,0}}{N^{p\delta}} \\ &\le C_3 N^{-D-90} + C_3 N^{C_2} \max_{k\in[0:2]} \mathfrak{P}_k(\varepsilon/2, z_1, \Psi),\end{aligned}$$

where in the last step we used the fact that $|z_0|$ is bounded. Similar bound holds when Im is replaced by Re, we omit the details. Now we may apply union bounds on $i, j \in [M]$ and an ϵ-net argument on γ with the following deterministic bounds

$$\left|\frac{\partial [G^\gamma(z)]_{ab}}{\partial \gamma}\right| \lesssim \frac{\|A\| + \gamma\|t^{1/2}W\|}{\eta^2},$$

$\eta > N^{-1}$, $\|A\| \le N^{1/2}$ and $\mathbb{P}(\|t^{1/2}W\| > 2) < N^{-D}$, to obtain that

$$\begin{aligned}\mathfrak{P}_0(\delta, z_0, \Psi) =& \mathbb{P}_\Psi\Big(\sup_{\substack{a,b\in[M]\\ 0\le\gamma\le 1}} |z_0^{1/2}\mathfrak{X}_{ab}[G^\gamma(z_0)]_{ab}| > N^\delta\Big) \\ &\le C_4 N^{-D-50} + C_4 N^{C_4} \max_{k\in[0:2]} \mathfrak{P}_k(\varepsilon/2, z_1, \Psi),\end{aligned}$$

for some large constant $C_4 > 0$. Repeating the above procedure for all $\mathfrak{P}_k(\delta, \eta, \Psi), k = 1, 2$ proves the claim. □

8 Proof of Corollary

We prove this corollary using a similar argument as in [Section 4, [11]] or [Section 4, [8]]. The key inputs are the rigidity estimate in Theorem 4.4 and the Green function comparison in Theorem 4.5.

Proof of Corollary . Let us first define for any E,

$$\mathcal{N}(E) := \big|\{i : \lambda_i(\mathcal{S}(Y)) \le \lambda_{-,t} + E\}\big|.$$

For any $\epsilon > 0$, we take $\ell = N^{-2/3-\epsilon/3}$ and $\eta = N^{-2/3-\epsilon}$. Recall from Theorem 4.4 that $\lambda_M(\mathcal{S}(Y)) \ge \lambda_{-,t} - N^{-2/3+\epsilon}$ holds with high probability. We further define

$$\begin{aligned}\chi_E(x) &:= \mathbf{1}_{[-N^{-2/3+\epsilon}, E]}(x - \lambda_{-,t}), \\ \theta_\eta(x) &:= \frac{\eta}{\pi(x^2+\eta^2)} = \frac{1}{\pi}\mathrm{Im}\,\frac{1}{x - \mathrm{i}\eta}.\end{aligned}$$

Then following the same arguments as in [Lemma 2.7, [10]], we can obtain that for $|E| \le N^{-2/3+\epsilon}$, the following holds with high probability:

$$\mathrm{Tr}(\chi_{E-\ell} * \theta_\eta)(\mathcal{S}(Y)) - N^{-\epsilon/9} \le \mathcal{N}(E) \le \mathrm{Tr}(\chi_{E+\ell} * \theta_\eta)(\mathcal{S}(Y)) + N^{-\epsilon/9}.$$

Let $K(x) : \mathbb{R} \to [0,1]$ be a smooth monotonic increasing function such that

$$K(x) = 1 \quad \text{if} \quad x \ge 2/3, \quad K(x) = 0 \quad \text{if} \quad x \le 1/3.$$

Therefore, we have with high probability that

$$\begin{aligned} K(\mathrm{Tr}(\chi_{E-\ell} * \theta_\eta)(\mathcal{S}(Y))) + \mathcal{O}(N^{-\epsilon/9}) \le K(\mathcal{N}(E)) = \mathbf{1}_{\mathcal{N}(E)\ge 1} \\ \le K(\mathrm{Tr}(\chi_{E+\ell} * \theta_\eta)(\mathcal{S}(Y))) + \mathcal{O}(N^{-\epsilon/9}). \end{aligned}$$

Taking expectation on the above inequality, we have for $|s| \le N^{\epsilon}/2$ that

$$\begin{aligned} &\mathbb{E}\left[K\left(\mathrm{Im}\left[\frac{N}{\pi}\int_{-N^{-2/3+\epsilon}}^{sN^{-2/3}-\ell} m^1(\lambda_{-,t}+y+\mathrm{i}\eta)\right]\mathrm{d}y\right)\right] + \mathcal{O}(N^{-\epsilon/9}) \\ &\le \mathbb{P}\Big(N^{2/3}(\lambda_M(\mathcal{S}(Y)) - \lambda_{-,t}) \le s\Big) = \mathbb{E}\Big[\mathbf{1}_{\mathcal{N}(sN^{-2/3})\ge 1}\Big] \\ &\le \mathbb{E}\left[K\left(\mathrm{Im}\left[\frac{N}{\pi}\int_{-N^{-2/3+\epsilon}}^{sN^{-2/3}+\ell} m^1(\lambda_{-,t}+y+\mathrm{i}\eta)\right]\mathrm{d}y\right)\right] + \mathcal{O}(N^{-\epsilon/9}). \end{aligned} \tag{S.8.1}$$

Similarly, repeating the above arguments with $\mathcal{S}(Y)$ replaced by $\mathcal{S}(V_t)$, we can also have

$$\begin{aligned} &\mathbb{E}\left[K\left(\mathrm{Im}\left[\frac{N}{\pi}\int_{-N^{-2/3+\epsilon}}^{sN^{-2/3}-\ell} m^0(\lambda_{-,t}+y+\mathrm{i}\eta)\right]\mathrm{d}y\right)\right] + \mathcal{O}(N^{-\epsilon/9}) \\ &\le \mathbb{P}\Big(N^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t}) \le s\Big) \\ &\le \mathbb{E}\left[K\left(\mathrm{Im}\left[\frac{N}{\pi}\int_{-N^{-2/3+\epsilon}}^{sN^{-2/3}+\ell} m^0(\lambda_{-,t}+y+\mathrm{i}\eta)\right]\mathrm{d}y\right)\right] + \mathcal{O}(N^{-\epsilon/9}). \end{aligned} \tag{S.8.2}$$

Note that the conditional expectation $\mathbb{E}_\Psi$ in (68) can be replaced by $\mathbb{E}$ using the law of total expectation together with the fact that Ω_Ψ holds with high probability. Therefore, we can combine (S.8.1) and (S.8.2) with (68) to obtain that

$$\begin{aligned} \mathbb{P}\Big(N^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t}) \le s - 2\ell N^{-2/3}\Big) + \mathcal{O}(N^{-\epsilon/9}) \le \mathbb{P}\Big(N^{2/3}(\lambda_M(\mathcal{S}(Y)) - \lambda_{-,t}) \le s\Big) \\ \le \mathbb{P}\Big(N^{2/3}(\lambda_M(\mathcal{S}(V_t)) - \lambda_{-,t}) \le s + 2\ell N^{-2/3}\Big) + \mathcal{O}(N^{-\epsilon/9}). \end{aligned}$$

Now (69) follows by the fact that $\ell N^{-2/3} \ll 1$. For (70), we first note by Theorem 2.12 that

$$|\lambda_{-,t} - \lambda_{\mathsf{shift}}| \le N^{-2/3+\epsilon} \tag{S.8.3}$$

holds in probability. This together with Theorem 4.4 implies that

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

also holds in probability. Then we may proceed similar to the proof of (69), but with all high probability estimates replaced by in probability estimates. It's worth noting that during the derivation of (S.8.1) and (S.8.2), the error term $\mathcal{O}(N^{-\epsilon/9})$ will become $\mathfrak{o}(1)$ because we lack an polynomial bound for the failure probability of (S.8.3). Finally, we can conclude the proof of (70) by using Theorem 4.5. □

9 Proof of Theorem 4.5

Proof. To ease presentation, we show the proof of the following comparison instead: for any $|E| \le N^{-2/3+\epsilon}$,

$$\Big|\mathbb{E}_\Psi\Big(F(N\eta_0 \mathrm{Im}\, m^1(\lambda_{-,t}+E+\mathrm{i}\eta_0))\Big) - \mathbb{E}_\Psi\Big(F(N\eta_0 \mathrm{Im}\, m^0(\lambda_{-,t}+E+\mathrm{i}\eta_0))\Big)\Big| \le CN^{-\delta_1}. \tag{S.9.1}$$

The proof of (68) is similar, and thus we omit it. Using the same notation as in the proof of Theorem 4.3 and further defining $h_{\gamma,(ij)}(\lambda,\beta) \equiv \eta_0 \sum_a f_{\gamma,(aa),(ij)}(\lambda,,\beta)$, we have

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

with

$$(I_1)_{ij} \equiv \mathbb{E}_\Psi\Big[A_{ij}F'\Big(h_{\gamma,(ij)}\big([Y^\gamma]_{ij},X_{ij}\big)\Big)g_{(ij)}\big([Y^\gamma]_{ij},X_{ij}\big)\Big],$$
$$(I_2)_{ij} \equiv \frac{\gamma t^{1/2}}{(1-\gamma^2)^{1/2}}\mathbb{E}_\Psi\Big[w_{ij}F'\Big(h_{\gamma,(ij)}\big([Y^\gamma]_{ij},X_{ij}\big)\Big)g_{(ij)}\big([Y^\gamma]_{ij},X_{ij}\big)\Big].$$

We first consider the estimation for $(I_1)_{ij}$. Notice that $(I_1)_{ij}$ can be further decomposed as

$$(I_1)_{ij} = (I_1)_{ij}\cdot \mathbf{1}_{\psi_{ij}=0} + (I_1)_{ij}\cdot \mathbf{1}_{\psi_{ij}=1} = (I_1)_{ij}\cdot \mathbf{1}_{\psi_{ij}=0},$$

where in the last step we used the fact that $A_{ij}\cdot \mathbf{1}_{\psi_{ij}=1} = 0$. Therefore, we only need to consider the case when $\psi_{ij}=0$, and $(I_1)_{ij}$ can be rewritten as

$$(I_1)_{ij} = \mathbb{E}_\Psi\Big[(1-\chi_{ij})a_{ij}F'\Big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\Big)g_{(ij)}(d_{ij},\chi_{ij}b_{ij})\Big]\cdot \mathbf{1}_{\psi_{ij}=0}.$$

By Taylor expansion, for an $s_1>0$ to be chosen later, there exists $\tilde{d}_{ij}\in[0,d_{ij}]$ such that,

$$\begin{aligned}(I_1)_{ij} &= \sum_{k_1=0}^{s_1}\frac{1}{k_1!}\mathbb{E}_\Psi\Big[(1-\chi_{ij})a_{ij}d_{ij}^{k_1}g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij})F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\Big]\cdot\mathbf{1}_{\psi_{ij}=0}\\ &+\frac{1}{(s_1+1)!}\mathbb{E}_\Psi\Big[(1-\chi_{ij})a_{ij}d_{ij}^{s_1+1}g_{(ij)}^{(s_1+1,0)}(\tilde{d}_{ij},\chi_{ij}b_{ij})F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\Big]\cdot\mathbf{1}_{\psi_{ij}=0}\\ &\equiv \sum_{k_1=0}^{s_1}(I_1)_{ij,k_1} + \mathsf{Rem}_1.\end{aligned}$$

Using (96)-(98), ,the perturbation argument as in (80), and the fact that $\mathrm{Im}\, m^\gamma(z_t)\prec 1$, we have for any (small)$\epsilon>0$ and (large)$D>0$,

$$\mathbb{P}_\Psi\Bigg(\Omega_{\epsilon,1} := \Big\{\big|g_{(ij)}^{(s_1+1,0)}(\tilde{d}_{ij},\chi_{ij}b_{ij})F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\big|\cdot\mathbf{1}_{\psi_{ij}=0} < t^{-s_1-2}N^\epsilon\Big\}\Bigg) \ge 1-N^{-D}.$$

Further, by the Gaussianity of w_{ij}, we have

$$\mathbb{P}_\Psi\Bigg(\Omega_{\epsilon,2} := \Big\{\max_{i\in[M],j\in[N]}|t^{1/2}w_{ij}| < N^{-1/2+\epsilon}\Big\}\Bigg) \ge 1-N^{-D}.$$

Let $\Omega_\epsilon := \Omega_{\epsilon,1}\cap\Omega_{\epsilon,2}$. Then

$$|\mathsf{Rem}_1| \lesssim \mathbb{E}_\Psi\Big[|(1-\chi_{ij})a_{ij}d_{ij}^{s_1+1}|\cdot\big|g_{(ij)}^{(s_1+1,0)}(\tilde{d}_{ij},\chi_{ij}b_{ij})F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\big|\cdot\mathbf{1}_{\Omega_\epsilon}\Big]\cdot\mathbf{1}_{\psi_{ij}=0}$$

$$
\begin{aligned}
&+\mathbb{E}_\Psi\left[\left|(1-\chi_{ij})a_{ij}d_{ij}^{s_1+1}\right|\cdot\left|g_{(ij)}^{(s_1+1,0)}(\tilde d_{ij},\chi_{ij}b_{ij})F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\right|\cdot\mathbf{1}_{\Omega_\epsilon^c}\right]\cdot\mathbf{1}_{\psi_{ij}=0}\\
&\overset{\text{(i)}}{\lesssim}\mathbb{E}_\Psi\left[\left|(1-\chi_{ij})a_{ij}d_{ij}^{s_1+1}\right|\cdot\left|g_{(ij)}^{(s_1+1,0)}(\tilde d_{ij},\chi_{ij}b_{ij})F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\right|\cdot\mathbf{1}_{\Omega_\epsilon}\right]\cdot\mathbf{1}_{\psi_{ij}=0}\\
&+N^{-D+C_1+2(s_1+3)}\\
&\overset{\text{(ii)}}{\lesssim}\frac{N^\epsilon}{N^{1/2+\epsilon_b(s_1+1)}t^{s_1+2}},
\end{aligned}
\tag{S.9.2}
$$

where in (i) we used the deterministic bound $\left|g_{(ij)}^{(s_1+1,0)}(\tilde d_{ij},\chi_{ij}b_{ij})F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\right|\le N^{C_1+2(s_1+3)}$ when $\eta\ge N^{-2}$, and (ii) is a consequence of the definition of Ω_ϵ. Choosing s_1 sufficiently large, i.e., $s_1>4/\epsilon_b$, and $t\gg N^{-\epsilon_b/2}$ we can obtain

$$
|\mathsf{Rem}_1|\lesssim N^{-5/2}.
$$

For $(I_1)_{ij,k_1}$, we need to further expand $F'\big(h_{\gamma,(ij)}(d_{ij})\big)$ as follows:

$$
F'\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)=\sum_{k=0}^{s_2}\frac{d_{ij}^k}{k!}\frac{\partial^k F'}{\partial d_{ij}^k}\big(h_{\gamma,(ij)}(0,\chi_{ij}b_{ij})\big)+\frac{d_{ij}^{s_2+1}}{(s_2+1)!}\frac{\partial^k F'}{\partial d_{ij}^k}\big(h_{\gamma,(ij)}(\hat d_{ij},\chi_{ij}b_{ij})\big),
$$

where s_2 is a positive integer to be chosen later, and $\hat d_{ij}\in[0,d_{ij}]$. Then $(I_1)_{ij,k_1}$ can be rewritten as,

$$
\begin{aligned}
(I_1)_{ij,k_1}&=\sum_{k_2=0}^{s_2}\frac{1}{k_1!k_2!}\mathbb{E}_\Psi\left[(1-\chi_{ij})a_{ij}d_{ij}^{k_1+k_2}g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij})\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,\chi_{ij}b_{ij})\big)\right]\cdot\mathbf{1}_{\psi_{ij}=0}\\
&+\frac{1}{k_1!(s_2+1)!}\mathbb{E}_\Psi\left[(1-\chi_{ij})a_{ij}d_{ij}^{k_1+s_2+1}g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij})\frac{\partial^{s_2+1}F'}{\partial d_{ij}^{s_2+1}}\big(h_{\gamma,(ij)}(\hat d_{ij},\chi_{ij}b_{ij})\big)\right]\cdot\mathbf{1}_{\psi_{ij}=0}\\
&\equiv\sum_{k_2=0}^{s_2}(I_1)_{ij,k_1k_2}+\mathsf{Rem}_2.
\end{aligned}
$$

By Faà di Bruno's formula, we have for any integer $n>0$,

$$
\begin{aligned}
\frac{\partial^n F'}{\partial d_{ij}^n}\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)=\sum_{(m_1,\cdots,m_n)}\frac{n!}{m_1!m_2!\cdot m_n!}\cdot F^{(m_1+\cdots+m_n+1)}\big(h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\big)\\
\times\prod_{\ell=1}^n\left(\frac{h_{\gamma,(ij)}^{(\ell)}(d_{ij},\chi_{ij}b_{ij})}{\ell!}\right)^{m_\ell}
\end{aligned}
\tag{S.9.3}
$$

Considering (S.9.3), (96)-(98), and using the perturbation argument as described in (80), we arrive at the following result:

$$
\frac{\partial^{s_2+1}F'}{\partial d_{ij}^{s_2+1}}\big(h_{\gamma,(ij)}(\hat d_{ij},\chi_{ij}b_{ij})\big)\prec\prod_{\ell=1}^n t^{-(\ell+1)m_\ell}\le t^{-2n}.
\tag{S.9.4}
$$

Moreover, taking into account the fact that $g_{(ij)}^{(k_1)}(0)\prec t^{-(k_1+1)}$, we can deduce that:

$$
|\mathsf{Rem}_2|\lesssim\frac{N^\epsilon}{N^{1/2+\epsilon_b(k_1+s_2+1)}t^{k_1+2(s_2+1)}}\lesssim N^{-5/2},
$$

where, for the final step, we have chosen $s_2\ge 4/\epsilon_b$ and $t\gg N^{-\epsilon_b/4}$. Next, we estimate $(I_1)_{ij,k_1k_2}$ in different cases.

Case 1: $k_1 + k_2$ is even. By the law of total expectation,

$$\begin{aligned}
&(I_1)_{ij,k_1k_2} \\
&= \frac{\mathbf{1}_{\psi_{ij}=0}}{k_1!k_2!}\sum_{n=0}^{1}\mathbb{E}_\Psi\left[(1-\chi_{ij})a_{ij}d_{ij}^{k_1+k_2}g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij})\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,\chi_{ij}b_{ij})\big)\Big|\chi_{ij}=n\right]\mathbb{P}(\chi_{ij}=n) \\
&= \frac{\mathbf{1}_{\psi_{ij}=0}}{k_1!k_2!}\mathbb{E}_\Psi\left[a_{ij}d_{ij}^{k_1+k_2}g_{(ij)}^{(k_1,0)}(0,0)\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,0)\big)\Big|\chi_{ij}=0\right]\mathbb{P}(\chi_{ij}=0) \\
&= \frac{\mathbf{1}_{\psi_{ij}=0}=0}{k_1!k_2!}\mathbb{E}_\Psi\left[a_{ij}d_{ij}^{k_1+k_2}\Big|\chi_{ij}=0\right]\mathbb{E}_\Psi\left[g_{(ij)}^{(k_1,0)}(0,0)\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,0)\big)\right]\mathbb{P}(\chi_{ij}=0),
\end{aligned} \tag{S.9.5}$$

where the last step follows from the symmetry condition.

Case 2: $k_1 + k_2$ is odd and $k_1 + k_2 \geq 5$. Similar to (S.9.5), we have

$$\begin{aligned}
|(I_1)_{ij,k_1k_2}| &\lesssim \left|\mathbb{E}_\Psi\left[a_{ij}d_{ij}^{k_1+k_2}\Big|\chi_{ij}=0\right]\right|\mathbb{E}_\Psi\left[|g_{(ij)}^{(k_1,0)}(0,0)|\left|\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,0)\big)\right|\Big|\chi_{ij}=0\right]\mathbb{P}(\chi_{ij}=0)\mathbf{1}_{\psi_{ij}=0} \\
&\lesssim \frac{1}{N^{2+2\epsilon_a+(k_1+k_2-3)\epsilon_b}}\mathbb{E}_\Psi\left[|g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij})|\left|\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,\chi_{ij}b_{ij})\big)\right|\right]\mathbf{1}_{\psi_{ij}=0,\chi_{ij}=0}.
\end{aligned}$$

We may again obtain the bound $|g_{(ij)}^{(k_1)}(0,\chi_{ij}b_{ij})|\cdot\mathbf{1}_{\psi_{ij}=0,\chi_{ij}=0} \prec t^{-(k_1+1)}$ by (96)-(98), and the perturbation argument as described in (80). Using (i)equation (S.9.3) with d_{ij} replaced by 0, and (ii)the following rank inequality,

$$|h_{\gamma,(ij)}(0,\chi_{ij}b_{ij}) - h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})|\cdot\mathbf{1}_{\psi_{ij}=0,\chi_{ij}=0} \leq 2\eta_0\big(\|G_{(ij)}^{\gamma,d_{ij}}(z_t)\| + \|G_{(ij)}^{\gamma,0}(z_t)\|\big)\mathbf{1}_{\psi_{ij}=0,\chi_{ij}=0} \leq 2, \tag{S.9.6}$$

with the fact that $h_{\gamma,(ij)}(d_{ij},\chi_{ij}b_{ij})\cdot\mathbf{1}_{\psi_{ij}=0,\chi_{ij}=0} \prec 1$, we can obtain that

$$\left|\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,\chi_{ij}b_{ij})\big)\right|\cdot\mathbf{1}_{\psi_{ij}=0,\chi_{ij}=0} \prec t^{-2k_2}. \tag{S.9.7}$$

Combining the above estimates and choosing $t \gg N^{-\epsilon_b/8}$, we arrive at

$$|(I_1)_{ij,k_1k_2}| \lesssim \frac{N^\epsilon}{N^{2+2\epsilon_a+(k_1+k_2-3)\epsilon_b}t^{k_1+1+2k_2}} \lesssim \frac{1}{N^{2+2\epsilon_a}}.$$

Case 3: $k_1 + k_2 = 3$. The estimation in this case is similar to Case 2 above, but we need to use the bound $g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij}) \prec 1$ when $i \in \mathcal{T}_r$ and $j \in \mathcal{T}_c$. Recall that $|\mathcal{D}_r| \vee |\mathcal{D}_c| \leq N^{1-\epsilon_d}$. Then we have

$$\begin{aligned}
|(I_1)_{ij,k_1k_2}| &\lesssim \frac{N^\epsilon}{N^{2+2\epsilon_a}}\cdot\mathbf{1}_{\psi_{ij}=0}\cdot\mathbf{1}_{i\in\mathcal{T}_r,j\in\mathcal{T}_c} + \frac{1}{N^{2-\epsilon_d}}\cdot\frac{N^\epsilon}{N^{2\epsilon_a+\epsilon_d}t^{k_1+1+2k_2}}\cdot\mathbf{1}_{\psi_{ij}=0}\cdot(1-\mathbf{1}_{i\in\mathcal{T}_r,j\in\mathcal{T}_c}) \\
&\lesssim \frac{1}{N^{2+\epsilon_a}}\cdot\mathbf{1}_{\psi_{ij}=0}\cdot\mathbf{1}_{i\in\mathcal{T}_r,j\in\mathcal{T}_c} + \frac{1}{N^{2-\epsilon_d+\epsilon_a}}\cdot\mathbf{1}_{\psi_{ij}=0}\cdot(1-\mathbf{1}_{i\in\mathcal{T}_r,j\in\mathcal{T}_c}),
\end{aligned}$$

where in the last step, we used the fact $t \gg N^{-\epsilon_d/8}$.

Case 4: $k_1 + k_2 = 1$. In this case, using (S.9.5) we may compute that

$$(I_1)_{ij,k_1k_2} = \mathbb{E}_\Psi\big[\gamma a_{ij}^2\big]\cdot\mathbb{E}_\Psi\left[g_{(ij)}^{(k_1,0)}(0,0)\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,0)\big)\Big|\chi_{ij}=0\right]\cdot\mathbb{P}(\chi_{ij}=0)\cdot\mathbf{1}_{\psi_{ij}=0}.$$

We note that there will be corresponding terms in $(I_2)_{ij}$, and these terms will cancel out with the ones described above.

Combining the estimates in the above cases, we can obtain that there exists some constant $\delta_1 = \delta_1(\epsilon_a)$ such that

$$\sum_{i,j}(I_1)_{ij} = \sum_{i,j}\sum_{\substack{k_1,k_2\geq 0,\\ k_1+k_2=1}} \mathbb{E}_\Psi\left[\gamma a_{ij}^2\right]\cdot \mathbb{E}_\Psi\left[g_{(ij)}^{(k_1,0)}(0,0)\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,0)\big)\right] \times \mathbb{P}(\chi_{ij}=0)\cdot \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\delta_1}). \tag{S.9.8}$$

Next, we consider the estimation for $(I_2)_{ij}$. When $\psi_{ij}=1$, we can apply Gaussian integration by parts to obtain that

$$|(I_2)_{ij}\cdot \mathbf{1}_{\psi_{ij}=1}| \lesssim \frac{t^{1/2}}{N}\mathbb{E}_\Psi\left[\left|\partial_{w_{ij}}\left\{g_{(ij)}(e_{ij},c_{ij})F'\big(h_{\gamma,(ij)}(e_{ij},c_{ij})\big)\right\}\right|\right]\cdot \mathbf{1}_{\psi_{ij}=1} \lesssim \frac{N^\epsilon}{Nt}\cdot \mathbf{1}_{\psi_{ij}=1},$$

where the last step follows from (96)-(98). The estimation for $(I_2)_{ij}\cdot \mathbf{1}_{\psi_{ij}=0}$ is similar to those of $(I_1)_{ij}$, we omit repetitive details. In summary, with the independence between z_t and w_{ij}, we have by possibly adjusting δ_1,

$$\begin{aligned}\sum_{i,j}(I_2)_{ij} &= \sum_{i,j}(I_2)_{ij}\cdot \mathbf{1}_{\psi_{ij}=0} + \sum_{i,j}(I_2)_{ij}\cdot \mathbf{1}_{\psi_{ij}=1}\\ &= \sum_{i,j}\sum_{\substack{k_1,k_2\geq 0,\\ k_1+k_2=1}} \mathbb{E}_\Psi\left[\gamma t w_{ij}^2\right]\mathbb{E}_\Psi\left[g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij})\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,\chi_{ij}b_{ij})\big)\right]\cdot \mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\delta_1}).\end{aligned} \tag{S.9.9}$$

Note by (87) and the choices of ϵ_a and ϵ_b, we have

$$\mathbb{E}_\Psi\left[\gamma a_{ij}^2\right]\mathbb{P}(\chi_{ij}=0) - \mathbb{E}_\Psi\left[\gamma t w_{ij}^2\right] = \mathcal{O}\left(\frac{t}{N^{2+2\epsilon_b}}\right).$$

This together with the t dependent bounds for $g_{(ij)}^{(k_1,0)}$ and $\partial^{k_2}F'/(\partial d_{ij}^{k_2})$ implies that it suffices to bound the following quantity:

$$\mathsf{G} := \left(\mathbb{E}_\Psi\left[g_{(ij)}^{(k_1,0)}(0,\chi_{ij}b_{ij})\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,\chi_{ij}b_{ij})\big)\right] - \mathbb{E}_\Psi\left[g_{(ij)}^{(k_1,0)}(0,0)\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,0)\big)\right]\right)\cdot \mathbf{1}_{\psi_{ij}=0}$$

To provide a more precise distinction between (S.9.8) and (S.9.9), we let

$$\mathsf{F}_{k_1,k_2}(z_t^{(ij)}(\beta)) := g_{(ij)}^{(k_1)}(0,\beta)\frac{\partial^{k_2}F'}{\partial d_{ij}^{k_2}}\big(h_{\gamma,(ij)}(0,\beta)\big).$$

Therefore,

$$\mathsf{G} = \left(\mathbb{E}_\Psi\left[\mathsf{F}_{k_1,k_2}\big(z_t(\chi_{ij}b_{ij})\big)\right] - \mathbb{E}_\Psi\left[\mathsf{F}_{k_1,k_2}\big(z_t(0)\big)\right]\right)\cdot \mathbf{1}_{\psi_{ij}=0}.$$

We may apply Taylor expansion to obtain that

$$\begin{aligned}&\left(\mathbb{E}_\Psi\left[\mathsf{F}_{k_1,k_2}\big(z_t(\chi_{ij}b_{ij})\big)\right] - \mathbb{E}_\Psi\left[\mathsf{F}_{k_1,k_2}\big(z_t(0)\big)\right]\right)\cdot \mathbf{1}_{\psi_{ij}=0}\\ &= \mathbb{E}_\Psi\left[\chi_{ij}^2 b_{ij}^2 \mathsf{F}'_{k_1,k_2}\big(z_t(b)\big)\cdot \frac{\partial^2\lambda_{-,t}}{\partial B_{ij}^2}(b)\right]\cdot \mathbf{1}_{\psi_{ij}=0} + \mathbb{E}_\Psi\left[\chi_{ij}^2 b_{ij}^2 \mathsf{F}''_{k_1,k_2}\big(z_t(b)\big)\cdot \left(\frac{\partial\lambda_{-,t}}{\partial B_{ij}}(b)\right)^2\right]\cdot \mathbf{1}_{\psi_{ij}=0},\end{aligned} \tag{S.9.10}$$

with $b \in [0, B_{ij}]$. Here the first oder term disappeared due to symmetry. To bound the above terms we need to first verify that $z_t(b)$ still lies inside D (w.h.p). This can be done by noting that for the replacement matrix $X_{(ij)}(b)$ which replace the B_{ij} by b in X still satisfies the η^*- regularity. Therefore by Weyl's inequality,

$$\begin{aligned}|\lambda_{-,t}(\chi_{ij}b_{ij}) - \lambda_{-,t}(b)| \prec |\lambda_{-,t}(\chi_{ij}b_{ij}) - \lambda_M(\mathcal{S}(X))| + |\lambda_M(\mathcal{S}(X)) - \lambda_M(\mathcal{S}(X_{(ij)}(b)))| \\ + |\lambda_M(\mathcal{S}(X_{(ij)}(b))) - \lambda_{-,t}(b)| \prec N^{-2/3} + N^{-\epsilon_b} + N^{-2/3} \prec N^{-\epsilon_b}. \qquad \text{(S.9.11)}\end{aligned}$$

Applying the perturbation argument as in (80) to relate $g^{(k_1)}_{(ij)}(0, b)$ back to $g^{(k_1)}_{(ij)}(d_{ij}, b)$, and then using (S.9.11) to verify that $z^{(ij)}_t(b) \in \mathsf{D}$, we can see that the bound $g^{(k_1)}_{(ij)}(0, b) \prec t^{-(k_1+1)}$ still holds. Similarly, we can also obtain $h^{(k_2)}_{\gamma,(ij)}(0, b) \prec t^{-k_2}$ for $k_2 \geq 1$. For the case when $k_2 = 0$, we may use (S.9.6) and the fact that $N\eta_0 \mathrm{Im}\, m^\gamma(z^{(ij)}_t(b)) \prec 1$ to conclude that $h_{\gamma,(ij)}(0, b) \prec 1$. Combining the above bounds with a Cauchy integral argument, we have

$$\mathsf{F}'_{k_1,k_2}\big(z_t(b)\big) \prec \frac{1}{\eta_0 t^2}, \quad \mathsf{F}''_{k_1,k_2}\big(z_t(b)\big) \prec \frac{1}{\eta_0^2 t^2}.$$

Further using Lemma 5.4, we have for arbitrary (small)$\epsilon > 0$ and (large)$D > 0$,

$$\mathbb{P}\bigg(\Omega := \Big\{\Big|\mathsf{F}'_{k_1,k_2}\big(z_t(b)\big) \cdot \frac{\partial^2 \lambda_{-,t}}{\partial B_{ij}^2}(b)\Big| \leq \frac{N^\epsilon}{N\eta_0 t^7}\Big\} \bigcap \Big\{\Big|\mathsf{F}''_{k_1,k_2}\big(z_t(b)\big) \cdot \Big(\frac{\partial \lambda_{-,t}}{\partial B_{ij}}(b)\Big)^2\Big| \leq \frac{N^\epsilon}{N^2\eta_0^2 t^8}\Big\}\bigg) \geq 1 - N^{-D}.$$

Since

$$\begin{aligned}&\chi_{ij}^2 b_{ij}^2\Big(\mathsf{F}'_{k_1,k_2}\big(z_t(b)\big) \cdot \frac{\partial^2 \lambda_{-,t}}{\partial B_{ij}^2}(b) + \mathsf{F}''_{k_1,k_2}\big(z_t(b)\big) \cdot \Big(\frac{\partial \lambda_{-,t}}{\partial B_{ij}}(b)\Big)^2\Big) \cdot \mathbf{1}_{\psi_{ij}=0} \\ &= \Big(\mathsf{F}_{k_1,k_2}\big(z_t(\chi_{ij}b_{ij})\big) - \mathsf{F}_{k_1,k_2}\big(z_t(0)\big)\Big) \cdot \mathbf{1}_{\psi_{ij}=0} - \Big(\chi_{ij}b_{ij}F'_{k_1,k_2}(z_t(0)) \cdot \frac{\partial \lambda_{-,t}}{\partial B_{ij}}(0)\Big) \cdot \mathbf{1}_{\psi_{ij}=0},\end{aligned}$$

the deterministic upper bound for the left hand side of the above equation follows from (89) in Lemma 5.4 and the fact that $\mathrm{Im}\, z_t \geq N^{-1}$. Then we may follow the steps as in (S.9.2) to obtain that

$$\mathbb{E}_\Psi\Big[\chi_{ij}^2 b_{ij}^2\Big(\mathsf{F}'_{k_1,k_2}\big(z_t(b)\big) \cdot \frac{\partial^2 \lambda_{-,t}}{\partial B_{ij}^2}(b) + \mathsf{F}''_{k_1,k_2}\big(z_t(b)\big) \cdot \Big(\frac{\partial \lambda_{-,t}}{\partial B_{ij}}(b)\Big)^2\Big)\Big] \cdot \mathbf{1}_{\psi_{ij}=0} \lesssim \frac{N^\epsilon}{N^2\eta_0 t^7}$$

Therefore, with the fact that $\mathbb{E}_\Psi[\gamma a_{ij}^2]\mathbb{P}(\chi_{ij} = 0) \sim t\mathbb{E}_\Psi[\gamma w_{ij}^2] = \gamma t/N$, we have by possibly adjusting δ_1,

$$\Big|\sum_{i,j}(I_1)_{ij} - (I_2)_{ij}\Big| = \sum_{i,j}\frac{\gamma t}{N}\sum_{\substack{k_1,k_2\geq 0,\\ k_1+k_2=1}}\Big|\mathbb{E}_\Psi\Big[\mathsf{F}_{k_1,k_2}\big(z_t(\chi_{ij}b_{ij})\big)\Big] - \mathbb{E}_\Psi\Big[\mathsf{F}_{k_1,k_2}\big(z_t(0)\big)\Big]\Big|\mathbf{1}_{\psi_{ij}=0} + \mathcal{O}(N^{-\delta_1}) = \mathcal{O}(N^{-\delta_1}).$$

This together with the arguments as in (115)-(116) completes the proof of (S.9.1). The proof for the case $\alpha = 8/3$ closely parallels, and is in fact simpler, primarily due to the absence of randomness in λ_{shift}. Thus we omit the details. This concludes the proof. □

References

[1] 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

[2] Bao Z, Lee J, Xu X (2023) Central limit theorem for traces of the resolvents of half-heavy tailed sample covariance matrices, URL https://arxiv.org/abs/2308.09581, arXiv:2308.09581

[3] Bauerschmidt R, Knowles A, Yau HT (2017) Local semicircle law for random regular graphs. Comm Pure Appl Math 70(10):1898–1960. https://doi.org/10.1002/cpa.21709, URL https://doi.org/10.1002/cpa.21709

[4] 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

[5] Bloemendal A, Erdős L, Knowles A, et al (2014) Isotropic local laws for sample covariance and generalized Wigner matrices. Electron J Probab 19:no. 33, 53. https://doi.org/10.1214/ejp.v19-3054, URL https://doi.org/10.1214/ejp.v19-3054

[6] 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

[7] 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

[8] 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

[9] 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

[10] Knowles A, Yin J (2013) Eigenvector distribution of Wigner matrices. Probab Theory Related Fields 155(3-4):543–582. https://doi.org/10.1007/s00440-011-0407-y, URL https://doi.org/10.1007/s00440-011-0407-y

[11] 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