LOCAL DENSITY OF THE SPECTRUM ON THE EDGE FOR SAMPLE COVARIANCE MATRICES WITH GENERAL POPULATION

By Zhigang Bao*, Guangming Pan† and Wang Zhou‡

Zhejiang University, Nanyang Technological University and National University of Singapore

In this paper, we will study a class of large dimensional real or complex sample covariance matrices in the form of $\mathcal{W}_N = \Sigma^{1/2} X X^* \Sigma^{1/2}$. Here $X = (x_{ij})_{M,N}$ is an $M \times N$ random matrix with independent entries $x_{ij}, 1 \le i \le M, 1 \le j \le N$ such that $\mathbb{E}x_{ij} = 0$, $\mathbb{E}|x_{ij}|^2 = 1/N$. On dimensions we assume that $M = M(N)$ and $N/M \to d \in (0, \infty)$ as $N \to \infty$. For the deterministic positive definite $M \times M$ population covariance matrix Σ, we will impose a quite general condition which was used by Karoui in [9] on complex Wishart matrices. Such a condition is particularly aimed at the right edge behavior of the spectrum of $\mathcal{W}_N$. In this paper, we will show that under some additional assumptions on the distributions of $(x_{ij})'$s, the so-called *local MP type law* holds on the right edge of the spectrum of $\mathcal{W}_N$. The local density problem was raised and developed by Erdös, Schlein, Yau and Yin etc. in a series work [12]-[21] for Wigner matrices and extended by Pillai and Yin [27] to sample covariance matrices in the null case ($\Sigma = I$), which asserts that the limiting spectral distributions of the above random matrix models even hold in a microscopic regime. The local MP type law will be a crucial input for our subsequent work [7] on establishing the edge universality of $\mathcal{W}_N$. We will essentially pursue the approach developed in [12]-[21] and [27] after deriving the so-called *square root behavior* of the spectrum on the right edge in advance. And we will invoke an argument on stability of the self-consistent equation of Stieltjes transform of the MP-type law raised recently by Erdös and Farrell in [11] for generalized MANOVA matrices.

1. Introduction. As a fundamental object in the theory of multivariate analysis, sample covariance matrix is unremittingly studied by researchers and plays important roles in dealing with large dimensional data arising from various fields such as genomics, image processing, microarray, proteomics and finance, etc. in recent decades. Among numerous topics and methods, the spectral analysis of large dimensional sample covariance matrices via the approaches in the Random Matrix Theory (RMT) has attracted considerable interest among mathematicians, probabilitists and statisticians. The study towards the eigenvalues of sample covariance matrices can date back to the work of Hsu [24], and became flourishing after the seminal work of Marčenko and Pastur [26], in which the authors raised the limiting spectral distribution (*MP type law*) for a class of sample covariance matrices. After that, a lot of researchers took part in developing the asymptotic theory of the empirical spectral distribution of large dimensional sample covariance matrices. One can refer to [6, 8, 32] for instance. In the past few years, in order to tackle some open problems on the local behavior of the eigenvalues for Wigner matrices, Erdös, Schlein and Yau raised the so-called *local semicircle law* in [14]. The local semicircle law has been

*Z.G. Bao was partially supported by NSFC grant 11071213, ZJNSF grant R6090034 and SRFDP grant 20100101110001

†G.M. Pan was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11

‡ W. Zhou was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11, and by a grant R-155-000-116-112 at the National University of Singapore.

MSC 2010 subject classifications: Primary 60B20; secondary 15B52

Keywords and phrases: sample covariance matrices, edge universality, Tracy-Widom law

further improved and developed in the subsequent series work [12]-[21] of Erdös, Schlein, Yau and Yin, etc. Moreover, in [18] and [27] the so-called *local MP law* was established for sample covariance matrices with null population ($\Sigma = I_M$) in different degrees. These local density results have shown to be quite crucial in proving the universality property of various local or global spectral statistics for the corresponding matrix models. One can refer to the following long list of references [13, 17, 18, 19, 20, 27, 29, 30, 31, 33] or the survey paper [10] for further reading. Moreover, the local density result is not only a technical input for establishing the universality property for spectral statistics, it is also of great interest in its own right. Actually, local density result can be viewed as a precise description of the convergence rate of the empirical spectral distribution. The convergence rate issue is relatively more classical in RMT and had been studied before the seminal work of Erdös, Schlein and Yau in [14] under the assumptions in varying degrees. Not trying to be comprehensive, one can see [1, 2, 3, 4, 22, 23] for instance.

Very recently, Erdös and Farrell studied the local eigenvalue density of generalized MANOVA matrices in the bulk case in [11]. As a by-product, the authors in [11] also provided a MP type law for the local eigenvalue density of matrix $T^{1/2}XX^*T^{1/2}$ in the bulk case, where T is specified to be the inverse of another sample covariance matrix which is independent of XX^*. Obviously, the matrix $T^{1/2}XX^*T^{1/2}$ in [11] can also be regarded as a sample covariance matrix with the special random population T. In this sense, [11] shed light on establishing the local MP type law for $\mathcal{W}_N$ under our assumption. In this paper, we will derive a local MP type law on the right edge for the sample covariance matrices with general population. Precisely, we will consider the sample covariance matrix in the form of

$$\mathcal{W}_N = \Sigma^{1/2}XX^*\Sigma^{1/2}, \quad X = (x_{ij})_{M,N}, \tag{1.1}$$

where $\{x_{ij} := x_{ij}(N), 1 \le i \le M := M(N), 1 \le j \le N\}$ is a collection of independent real or complex variables such that

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

To state our results, we introduce some basic notions at first.

1.1. *Basic notions.* In the sequel, we will denote the ordered eigenvalues of an $n\times n$ Hermitian matrix A by

$$\lambda_n(A) \le \ldots \le \lambda_2(A) \le \lambda_1(A). \tag{1.2}$$

Moreover, we call

$$F_A(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\lambda_i(A)\le\lambda\}}$$

the *empirical spectral distribution*(ESD) of A. For ease of presentation, we set

$$d_N := \frac{N}{M} \to d \in (0,\infty), \quad \text{as} \quad N\to\infty.$$

We denote the ESD of Σ by

$$H_N(\lambda) := \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{\{\lambda_i(\Sigma)\le\lambda\}}$$

and that of $\mathcal{W}_N$ by

$$\underline{F}_N(\lambda) := \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{\{\lambda_i(\mathcal{W})\le\lambda\}}.$$

Here $\mathbf{1}_{\mathbb{S}}$ represents the indicator function of the event $\mathbb{S}$. Define the $N \times N$ matrix

$$W_N := X^* \Sigma X$$

which shares the same non-zero eigenvalues with $\mathcal{W}_N$. Denoting the ESD of W_N by F_N, by definition we can see that

$$F_N(\lambda) = d_N^{-1} \underline{F}_N(\lambda) + (1 - d_N^{-1}) \mathbf{1}_{\{\lambda \geq 0\}}.$$

If there is some definite distribution H such that

$$H_N \Longrightarrow H \tag{1.3}$$

as $N \to \infty$, it is well known that there are definite distributions $F_{d,H}$ and $\underline{F}_{d,H}$ such that

$$F_N \Longrightarrow F_{d,H}, \quad \underline{F}_N \Longrightarrow \underline{F}_{d,H}$$

in probability. One can refer to [2] for instance. And we have the relation

$$F_{d,H} = d^{-1} \underline{F}_{d,H} + (1 - d^{-1}) \mathbf{1}_{[0,\infty)}. \tag{1.4}$$

We call $F_{d,H}$ (resp. $\underline{F}_{d,H}$) as the *limiting spectral distribution* (LSD) of W_N (resp. $\mathcal{W}_N$). However, for general Σ, $F_{d,H}$ usually has no closed form expression, so does $\underline{F}_{d,H}$. To define $F_{d,H}$ accurately, we need the theory of Stieltjes transform. For any distribution function D, its Stieltjes transform $m_D(z)$ is defined by

$$m_D(z) = \int \frac{1}{\lambda - z} dD(\lambda)$$

for all $z \in \mathbb{C}^+ := \{\omega \in \mathbb{C}, \Im \omega > 0\}$. And for any square matrix A, its Green function is defined by

$$G_A(z) = (A - zI)^{-1}, \quad z \in \mathbb{C}^+.$$

For simplicity, we denote

$$m(z) := m_{F_{d,H}}(z), \quad \underline{m}(z) := m_{\underline{F}_{d,H}}(z), \quad m_N(z) := m_{F_N}(z), \quad \underline{m}_N(z) := m_{\underline{F}_N}(z).$$

Particularly, we use the notation

$$G(z) = G_N(z) := (W_N - zI)^{-1}, \quad z \in \mathbb{C}^+.$$

By definition we have the elementary relation

$$m_N(z) = \frac{1}{N} TrG(z) = \frac{1}{N} \sum_{i=1}^{N} G_{ii}(z),$$

where we use $G_{ij}(z)$ to denote the (i,j)-th entry of $G(z)$. Fortunately, the Stieltjes transform of $F_{d,H}$ admits a self-consistent equation which is friendly for analysis. Actually $m(z)$ is the unique solution in $\mathbb{C}^+$ of the self-consistent equation

$$m(z) = \frac{1}{-z + d^{-1} \int \frac{t}{tm(z)+1} dH(t)} \tag{1.5}$$

for $z \in \mathbb{C}^+$. One can again refer to [2] for instance. By the well known inverse formula of Stieltjes transform, one can identify a distribution with its Stieltjes transform. We usually call (1.5) the *self-consistent equation* of *MP type law*, which is a generalization of the self-consistent equation of the Stieltjes transform of classical MP law in the null case ($\Sigma = I_M$). Moreover, from (1.4) we also have

$$m(z) = \frac{d^{-1}-1}{z} + d^{-1}\underline{m}(z).$$

However, what we indeed need in the sequel is the non-asymptotic version of $F_{d,H}$ which can be obtained through replacing d and H by d_N and H_N in $F_{d,H}$ and thus will be denoted by F_{d_N,H_N}. More precisely, F_{d_N,H_N} is the corresponding distribution function of the Stieltjes transform $m_{d_N,H_N}(z) := m_{F_{d_N,H_N}}(z) \in \mathbb{C}^+$ satisfying the following self-consistent equation

$$m_{d_N,H_N}(z) = \frac{1}{-z + d_N^{-1}\int \frac{t}{tm_{d_N,H_N}(z)+1} dH_N(t)}, \quad z \in \mathbb{C}^+. \tag{1.6}$$

In other words, by the inverse formula of the Stieltjes transform, we can define

$$F_{d_N,H_N}(x) := \begin{cases} 0, & \text{if} \quad x < 0, \\ \mathbf{1}_{\{d_N \geq 1\}}(1-d_N^{-1}), & \text{if} \quad x = 0, \\ \mathbf{1}_{\{d_N \geq 1\}}(1-d_N^{-1}) + \lim_{\eta\downarrow 0}\frac{1}{\pi}\int_{0^+}^x \Im m_{d_N,H_N}(t+i\eta)dt, & \text{if} \quad x > 0. \end{cases}$$

Correspondingly, we can define the non asymptotic versions of $\underline{F}_{d,H}$ and $\underline{m}(z)$ denoted by $\underline{F}_{d_N,H_N}$ and $\underline{m}_{d_N,H_N}(z)$ respectively. For ease of notation, we will briefly denote

$$m_0(z) := m_{d_N,H_N}(z), \quad \underline{m}_0(z) := \underline{m}_{d_N,H_N}(z), \quad F_0 := F_{d_N,H_N}, \quad \underline{F}_0 := \underline{F}_{d_N,H_N}$$

in the sequel. By definition, we have

$$m_0(z) = \frac{d_N^{-1}-1}{z} + d_N^{-1}\underline{m}_0(z).$$

Moreover, the relation above still holds if we replace $m_0(z)$ and $\underline{m}_0(z)$ by $m_N(z)$ and $\underline{m}_N(z)$ respectively. To state our results, we need to further introduce a crucial parameter $\mathbf{c} =: \mathbf{c}(\Sigma, N, M)$. Let

$$\mathbf{c} := \mathbf{c}(\Sigma, N, M), \quad \mathbf{c} \in [0, 1/\lambda_1(\Sigma)),$$

such that

$$\int \left(\frac{\lambda\mathbf{c}}{1-\lambda\mathbf{c}}\right)^2 dH_N(\lambda) = d_N. \tag{1.7}$$

It is easy to check that the definition of $\mathbf{c}$ is unique. Moreover, we set

$$\lambda_r = \frac{1}{\mathbf{c}}\left(1 + d_N^{-1}\int \frac{\lambda\mathbf{c}}{1-\lambda\mathbf{c}} dH_N(\lambda)\right). \tag{1.8}$$

By the discussions in [28] we can learn that F_0 has a continuous derivative ρ_0 on $\mathbb{R}\backslash\{0\}$. Actually, by Lemma 6.2 of [5], it is not difficult to see the rightmost boundary of the support of ρ_0 is λ_r defined in (1.8), i.e.

$$\lambda_r = \inf\{x \in \mathbb{R} : F_0(x) = 1\}.$$

Moreover, there exists

$$\mathbf{c} = -\lim_{z\in\mathbb{C}^+\to\lambda_r} m_0(z).$$

The existence of $\lim_{z\in\mathbb{C}^+\to x} m_0(z)$ for $x \in \mathbb{R}\setminus\{0\}$ has been proved in [28].

1.2. *Main results.* The main condition throughout the paper is as follows.

CONDITION 1.1. *Throughout the paper, we will need the following conditions.*
(i)(On dimensions): We assume that there are some positive constants c_1 and C_1 such that

$$c_1 < d_N < C_1.$$

(ii)(On X): We assume that $\{x_{ij} := x_{ij}(N), 1 \le i \le M, 1 \le j \le N\}$ is a collection of independent real or complex variables such that

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

Moreover, we assume that $\sqrt{N}x_{ij}$'s have a sub-exponential tail, i.e. there exists some positive constant τ_0 independent of i, j, N such that for sufficiently large t, one has

$$\mathbb{P}(|\sqrt{N}x_{ij}| \ge t) \le \tau_0^{-1}\exp(-t^{\tau_0}). \tag{1.9}$$

(iii)(On Σ): We assume that

$$\liminf_N \lambda_M(\Sigma) > 0, \quad \limsup_N \lambda_1(\Sigma) < \infty$$

and

$$\limsup_N \lambda_1(\Sigma)\mathbf{c} < 1. \tag{1.10}$$

REMARK 1.2. *We remind here that (iii) of Condition 1.1 was used by Karoui in [9] on the complex Gaussian sample covariance matrices (i.e. complex Wishart matrices) to guarantee the Tracy-Widom limit of the largest eigenvalue of $\mathcal{W}_N$. Such a condition, especially (1.10) is aimed at the behavior of eigenvalues on the right edge of the spectrum. It will be shown in next section that (1.10) substantially implies a* square root behavior *of the density ρ_0 which will be crucial for our main result.*

Moreover, we need the following terminologies on frequent events.

DEFINITION 1.3. *We say an event $\mathbb{S}$ happens* with overwhelming probability *if*

$$\mathbb{P}(\mathbb{S}) \ge 1 - N^{-\mathbf{A}}$$

for any fixed large constant $\mathbf{A} > 0$ when N is large enough. We say an event $\mathbb{S}$ holds with ζ-high probability *if there is some positive constant C such that*

$$\mathbb{P}(\mathbb{S}) \ge 1 - N^C \exp(-\varphi^\zeta)$$

when N is large enough, where

$$\varphi := \varphi_N = (\log N)^{\log\log N}$$

which will be used as a crucial parameter throughout the paper.

At first, by the definitions of $\mathbf{c}, \lambda_r$ and (iii) of Condition 1.1, it is easy to see

$$c_0 < \mathbf{c}, \lambda_r < C_0$$

for some small positive constant c_0 and large positive constant C_0. Moreover, note that we have the following elementary inequality

$$\lambda_M(\Sigma)\lambda_1(XX^*) \le \lambda_1(\mathcal{W}) \le \lambda_1(\Sigma)\lambda_1(XX^*).$$

Now by the rigidity of the locations of the eigenvalues of XX^* provided in Theorem 3.3 of [27] and Condition 1.1 we know for any fixed positive constant ζ there exists

$$\lambda_M(\Sigma)(1+\sqrt{d})^2/5 \le \lambda_1(\mathcal{W}) \le 5\lambda_1(\Sigma)(1+\sqrt{d})^2 \tag{1.11}$$

with ζ-high probability. Now we set

$$C_l = \lambda_M(\Sigma)(1+\sqrt{d})^2/C_0, \quad C_r = C_0\lambda_1(\Sigma)(1+\sqrt{d})^2.$$

with some sufficiently large positive constant C_0 such that $\lambda_r \in [2C_l, C_r/2]$. Then by (1.11), we can choose C_0 sufficiently large such that

$$C_l \le \lambda_1(\mathcal{W}) \le C_r \tag{1.12}$$

with ζ-high probability. In the sequel, we will always write

$$z = E + i\eta.$$

Then for $\zeta \ge 0$, we define two sets,

$$S(\zeta) := \{z \in \mathbb{C} : C_l \le E \le C_r, \varphi^{\zeta}N^{-1} \le \eta \le 1\},$$

and

$$S_r(\tilde{c}, \zeta) := \{z \in \mathbb{C} : \lambda_r - \tilde{c} \le E \le C_r, \varphi^{\zeta}N^{-1} \le \eta \le 1\},$$

where $\tilde{c}$ is a sufficiently small positive constant. With the above notations we can state our main result which can be viewed as a generalization of the strong local MP law for sample covariance matrices in the null case provided in [27] by Pillai and Yin to a large class of non null cases.

THEOREM 1.4 (Strong local MP type law around λ_r). *Let $z = E + i\eta$. Under Condition 1.1 , for some positive constant C,*

(i):

$$\bigcap_{z\in S_r(\tilde{c},5C)} \left\{ |m_N(z) - m_0(z)| \le \varphi^C \frac{1}{N\eta} \right\} \tag{1.13}$$

holds with overwhelming probability, and

(ii):

$$\bigcap_{z\in S_r(\tilde{c},5C)} \left\{ \max_{j\neq k} |G_{jk}(z)| + \max_i |G_{ii}(z)| \le \varphi^C \left(\sqrt{\frac{\Im m_0(z)}{N\eta}} + \frac{1}{N\eta} \right) \right\} \tag{1.14}$$

holds with overwhelming probability.

Now it has been well understood that the closeness of two Stieltjes transforms with small η is approximately equivalent to the closeness of their corresponding distribution functions in a small scale. In this sense, (1.13) describes the fact that the ESD F_N is well approximated by the LSD F_0 even in a tiny interval on the right edge of F_N. More precisely, we have the following result on the convergence rate of F_N on the right edge.

THEOREM 1.5 (Convergence rate of F_N around λ_r). *Let $\tilde{c}$ be the positive constant in Theorem 1.4. Under Condition 1.1, for any $\zeta > 0$ there exists a constant C_ζ such that the following events hold with overwhelming probability.*

(i): For the largest eigenvalue $\lambda_1(\mathcal{W})$, there exists

$$|\lambda_1(\mathcal{W}) - \lambda_r| \le N^{-2/3}\varphi^{C_\zeta}.$$

(ii): For any

$$E_1, E_2 \in [\lambda_r - \tilde{c}, C_r],$$

there exists

$$|(F_N(E_1) - F_N(E_2)) - (F_0(E_1) - F_0(E_2))| \le \frac{C(\log N)\varphi^{C_\zeta}}{N}. \tag{1.15}$$

REMARK 1.6. *Similar results on the convergence rate of $\mathcal{W}_N$ on the whole real line $\mathbb{R}$ has been given in [25] and [3] recently under weaker moment assumption on the entries of X. However, the best rate in these papers is $O(n^{-1/2})$ which is inadequate to help to establish the edge universality property in our subsequent work [7].*

1.3. *Route of the proof.* Crudely speaking, we will pursue the approach developed in the series work [12]-[21] and [27]. Especially, the main roadmap for the proof will be analogous to that for the null case in [27]. More specifically, we will follow the bootstrap strategy developed in [19, 21, 12, 27] to establish the strong MP type law around λ_r. The word "bootstrap" means that one can provide a weak law of the local eigenvalue density at first (in our case see Theorem 3.3), then the weak law can help to obtain the desired strong law through a bootstrap process. One main technical tool to derive the strong law from the weaker one is an abstract decoupling lemma from [27] (see Lemma 7.3 therein) which can help to bound the summation of a class of weakly dependent random variables. Such a decoupling lemma is similar to Theorem 5.6 of [12] and Lemma 4.1 of [21] but is more general and applicable to our model. However, most parts of the proof require more general treatments and the generality of Σ in our setting produces some additional obstacles.

At first, in the null case, the limiting spectral distribution is well known as the classical MP law which has a closed form. As a consequence, the properties of the Stieltjes transform of MP law can be easily obtained. Actually, these basic properties are crucial inputs for establishing the local MP law in the null case. One can refer to Lemma 6.5 of [27] for instance. However, in the non-null case, to get the corresponding properties around λ_r is not a trivial thing. Actually, the analysis towards the behaviors of ρ_0 and $m_0(z)$ will be our first main task. We will show that ρ_0 admits a square root behavior in an interval $[\lambda_r - \tilde{c}, \lambda_r]$ for some small positive constant $\tilde{c}$. Such a square root behavior is substantially guaranteed by (iii) of Condition 1.1 and will be the basic ingredient to establish the properties of $m_0(z)$ in Lemma 2.3.

Another main difficulty comes from the complexity of the self-consistent equations for the Stieltjes transforms (m_N and m_0) which makes the proof of the strong local MP type law much more cumbersome. For example, in [27], once the closeness of the self-consistent equations of m_N and m_0 is obtained, the difference between m_N and m_0 themselves can be characterized easily with the aid of the closed form of m_0. However, in the non-null case, this step is much more indirect. To overcome this difficulty, we will rely on an argument on the stability of the self-consistent equation of the Stietjes transform for $T^{1/2}XX^*T^{1/2}$ in [11]. Though the argument in [11] was only provided for the bulk case, we find it can be extended to the edge case under our assumptions on $\mathcal{W}_N$.

1.4. *Notation and organization.* Throughout the paper, we will use the notation $O(\cdot)$ and $o(\cdot)$ in the conventional sense. And we will use C, C_0, C_1, C_2, C_3 to denote some positive constants whose values may be different from line to line. We say

$$x \sim y$$

if there exist some positive constants C_1 and C_2 such that

$$C_1|y| \leq |x| \leq C_2|y|.$$

We say two functions $f(z), g(z) : \mathbb{C} \to \mathbb{C}$ have the relation

$$f(z) \sim g(z)$$

if there exist some positive constants C_1 and C_2 independent of z such that

$$C_1|g(z)| \leq |f(z)| \leq C_2|g(z)|.$$

Moreover, we will use $Spec(A)$ to denote the spectrum of a matrix A. And we will denote the operator norm and Hilbert-Schmidt norm of a matrix A by $||A||_{op}$ and $||A||_{HS}$ respectively.

This paper is organized as follows. In Section 2, we will study the properties of $m_0(z)$ and ρ_0 which will be crucial to our further analysis. In Section 3, we will prove that $m_N(z)$ is close to $m_0(z)$ when $\Re z$ is around λ_r, i.e. the strong local MP type law holds around λ_r. In Section 4, we will use the strong local MP type law to study the convergence rate of F_N on the right edge. Section 5 will be denoted to a corollary of our main theorems. Such a corollary will be used in our subsequent work [7] on the edge universality of $\mathcal{W}_N$.

2. Properties of ρ_0 and $m_0(z)$. At first, recall the notation $z = E + i\eta$ and define the parameter

$$\kappa := \kappa(z) = |E - \lambda_r|.$$

And we also need to recall the definitions of C_l, C_r, $S(\zeta)$ and $S_r(\tilde{c}, \zeta)$ in the last section. We will prove the following two lemmas. The first one claims the square root behavior of $\rho_0(x)$ on interval $[\lambda_r - 2\tilde{c}, \lambda_r]$ with some small positive constant $\tilde{c}$.

LEMMA 2.1. *Under Condition 1.1, there exists some sufficiently small constant $\tilde{c} > 0$ independent of N such that*

$$\rho_0(x) \sim \sqrt{\lambda_r - x}, \quad \text{for all } x \in [\lambda_r - 2\tilde{c}, \lambda_r]. \tag{2.1}$$

REMARK 2.2. *Note that in [26] and [28] the square root behavior of $\rho_0(x)$ near the boundary of its support has been discussed. However, the results in [26] and [28] does not imply (2.1) since here $m_0(z)$ and $\rho_0(x)$ are N-dependent. For general Σ, it is possible for the square root behavior only holds in an interval (with λ_r being its right end) with length of $o(1)$. For example, when $p = n$, in the 1 spike case $\Sigma = (100, 1\ldots, 1)$ (2.1) does not hold actually.*

The second one collects some crucial properties of $m_0(z)$.

LEMMA 2.3. *Under Condition 1.1, for some sufficiently small positive constant $\tilde{c}$ satisfying (2.1), the following four statements hold.*

(i): For $z \in S(0)$, we have

$$|m_0(z)| \sim 1,$$

(ii): For $z \in S_r(\tilde{c}, 0)$, we have

$$\Im m_0(z) \sim \begin{cases} \frac{\eta}{\sqrt{\kappa+\eta}}, & \text{if } E \geq \lambda_r + \eta \\ \sqrt{\kappa+\eta}, & \text{if } E \in [\lambda_r - \tilde{c}, \lambda_r + \eta) \end{cases}$$

(iii): For $z \in S(0)$, we have

$$\frac{\Im m_0(z)}{N\eta} \geq C\frac{1}{N}, \quad \partial_\eta \frac{\Im m_0(z)}{\eta} \leq 0.$$

for some positive constant C.

(iv): For $z \in S_r(\tilde{c}, 0)$, we have

$$|1 + tm_0(z)| \geq \hat{c}(1 + \lambda_1(\Sigma)m_0(\lambda_r)) \geq c_0, \quad \forall t \in [\lambda_M(\Sigma), \lambda_1(\Sigma)]$$

for some small positive constants $\hat{c}, c_0$. Moreover, there exists a sufficiently small constant $\hat{\eta} := \hat{\eta}(c_0)$, such that when $z \in S_r(\tilde{c}, 0)$ with $\eta \leq \hat{\eta}$, we also have

$$1 + t\Re m_0(z) \geq \hat{c}(1 + \lambda_1(\Sigma)m_0(\lambda_r)) \geq c_0, \quad \forall t \in [\lambda_M(\Sigma), \lambda_1(\Sigma)]. \tag{2.2}$$

Remark 2.4. *Note the second inequality in (iii) of Lemma 2.3 implies that $\Im m_0(z)/\eta$ is decreasing in η.*

At first we come to prove Lemma 2.1, which Lemma 2.3 will heavily rely on.

Proof of Lemma 2.1. For ease of presentation, we denote

$$m_1(z) = \Re m_0(z), \quad m_2(z) = \Im m_0(z).$$

Now let

$$m_0(x) := \lim_{z \in \mathbb{C}^+ \to x} m_0(z), \quad m_1(x) = \Re m_0(x), \quad m_2(x) = \Im m_0(x).$$

By Theorem 1.1 of [28], we know $m_0(x)$ exists for all $x \in \mathbb{R} \setminus \{0\}$. Moreover, $m_0(x)$ is continuous on $\mathbb{R} \setminus \{0\}$. By definition and the inverse formula of the Stieltjes transform we have

$$\rho_0(x) = \frac{1}{\pi} m_2(x), \quad x \in \mathbb{R} \setminus \{0\}.$$

Thus it suffices to prove the square root behavior of $m_2(x)$ when x is to the left of λ_r. Note that by the fact $m_2(\lambda_r) = 0$ we have

$$m_2(x) = \left(-2\int_x^{\lambda_r} m_2 m_2'(t) dt\right)^{1/2}.$$

Thus it suffices to show for some sufficiently small constant $\tilde{c} > 0$ there exists some positive constants $C > C'$ independent of t such that

$$-C \leq m_2 m_2'(t) \leq -C', \quad \text{for all } t \in [\lambda_r - 2\tilde{c}, \lambda_r). \tag{2.3}$$

To verify (2.3), we start from (1.6), which can be rewritten as

$$z = -m_0^{-1}(z) + d_N^{-1} \int \frac{t}{tm_0(z) + 1} dH_N(t). \tag{2.4}$$

When $\Re z \geq c_2$ for any fixed positive number c_2, it is easy to see from (2.4) that

$$m_0(z) \sim 1, \quad \text{if} \quad c_2 \leq \Re z \leq C_r \quad \text{and} \quad 0 < \Im z \leq C \tag{2.5}$$

for some positive constant C. Note that for every fixed N, $H_N(t)$ is a discrete distribution. Therefore, for any i, $\lambda_i(\Sigma)m_0(z) + 1$ does not tend to zero when z tends to some $x \in [c_2, C_r]$. Consequently, it is easy to see that (2.4) and (2.5) also holds for $z = x \in [c_2, C_r]$. That implies

$$\rho_0(x) = \frac{1}{\pi} m_2(x) \leq O(1), \quad c_2 \leq x \leq \lambda_r.$$

Setting $z = x$ in (2.4) and writing down the real and imaginary parts of both two sides we can get

$$\begin{aligned} x &= -\frac{m_1}{m_1^2 + m_2^2} + d_N^{-1} \int \frac{t(1+tm_1)dH_N(t)}{(1+tm_1)^2 + t^2 m_2^2} \\ 0 &= m_2 \left(\frac{1}{m_1^2 + m_2^2} - d_N^{-1} \int \frac{t^2 dH_N(t)}{(1+tm_1)^2 + t^2 m_2^2} \right). \end{aligned} \tag{2.6}$$

When $m_2(x) \neq 0$, i.e. $m_2(x) > 0$, (2.6) implies that

$$\begin{aligned} x &= d_N^{-1} \int \frac{t dH_N(t)}{(1+tm_1)^2 + t^2 m_2^2} \\ 0 &= \frac{1}{m_1^2 + m_2^2} - d_N^{-1} \int \frac{t^2 dH_N(t)}{(1+tm_1)^2 + t^2 m_2^2}. \end{aligned} \tag{2.7}$$

For simplicity, above we have omitted the variable x from the notation $m_1(x)$ and $m_2(x)$. We remind here by continuity, (2.5) and the fact that $1 + \lambda_i(\Sigma)m_0(x) \neq 0$, we can see that (2.7) still holds when $\lambda(\geq c_2)$ is a boundary of the support of $m_2(x)$. Moreover, when λ is a boundary of the support of $m_2(x)$ and $\lambda \geq c_2$ with any fixed positive number c_2, (2.7) can be simplified to

$$0 = \frac{1}{m_1^2(\lambda)} - d_N^{-1} \int \frac{t^2 dH_N(t)}{(1+tm_1(\lambda))^2} \tag{2.8}$$

since $m_2(\lambda) = 0$ in this case. Our analysis below will rely on (2.7). Thus at first we need to guarantee the validity of the following lemma.

Lemma 2.5. *Under Condition 1.1, there exists some positive constant $\tilde{c}$ such that*

$$m_2(x) = \pi \rho_0(x) > 0$$

in $[\lambda_r - 3\tilde{c}, \lambda_r)$.

At first, we proceed to prove Lemma 2.1 assuming the validity of Lemma 2.5 and prove Lemma 2.5 after that. Using (2.6) and Lemma 2.5 we have (2.7). Now taking derivatives with respect to x implicitly on both sides of two equations in (2.7), we can get

$$m_2 m_2' = \frac{m_1(A_2 + m_1 A_3) - m_2^2 A_3}{(A_2 + m_1 A_3)^2 + m_2^2 A_3^2}, \tag{2.9}$$

where

$$A_j = 2 d_N^{-1} \int \frac{t^j dH_N(t)}{((1+tm_1)^2 + t^2 m_2^2)^2}, \quad j = 2, 3.$$

Note that obviously $A_j > 0, j = 2, 3$. Moreover, if

$$\min_i (1 + \lambda_i(\Sigma) m_1(x)) > c', \quad \text{for all} \quad x \in [\lambda_r - 2\tilde{c}, \lambda_r) \tag{2.10}$$

for some positive constant c', we also have in $[\lambda_r - 2\tilde{c}, \lambda_r)$,

$$0 \le m_2^2 A_3, m_2^2 A_3^2 \le C_1 \tag{2.11}$$

and

$$C_2 \le A_2 + m_1 A_3 \le C_2' \tag{2.12}$$

for some positive constants C_1 and $C_2 < C_2'$. If we can take a step further to show

$$-2c_1 \le m_1(x) \le -c_1 \tag{2.13}$$

for some positive constant c_1, then (2.3) immediately follows from (2.9),(2.11),(2.12) and (2.13). Therefore, it remains to show there exists some small positive constant $\tilde{c}$ such that (2.10) and (2.13) holds for all $x \in [\lambda_r - 2\tilde{c}, \lambda_r)$.

Now note that

$$m_1(\lambda_r) = m_0(\lambda_r), \quad m_2(\lambda_r) = 0.$$

At first , by the fact that $m_1(\lambda_r) = -\mathbf{c}$ and (1.7) we see $m_1(\lambda_r) < 0$. Setting $z = \lambda_r$ in (2.4) we can easily see that

$$-\frac{7}{4} c_1 \le m_1(\lambda_r) \le -\frac{5}{4} c_1 \tag{2.14}$$

for some positive constant c_1. And by (iii) of Condition 1.1 and the fact that $m_1(\lambda_r) = -\mathbf{c}$ we also have

$$\min_i (1 + \lambda_i(\Sigma) m_1(\lambda_r)) = 1 + \lambda_1(\Sigma) m_1(\lambda_r) > 2c' \tag{2.15}$$

for some sufficiently small positive constant c'. Now we start from (2.14) and (2.15) to prove (2.10) and (2.13) by continuity. To show (2.10) and (2.13) hold for all $x \in [\lambda_r - 2\tilde{c}, \lambda_r]$ with some sufficiently small positive constant $\tilde{c}$, we need to control the $|m'(x)|$. Differentiating implicitly the first equation in (2.7) with respect to x again and use (2.9), we can get

$$\begin{aligned} |m_1'(x)| &= \left| -\frac{1 + A_3 \frac{m_1(A_2 + m_1 A_3) - m_2^2 A_3}{(A_2 + A_3 m_1)^2 + A_3^2 m_2^2}}{A_2 + m_1 A_3} \right| \\ &\le C \min_i |1 + \lambda_i(\Sigma) m_1(x)|^{-3} \end{aligned} \tag{2.16}$$

which can be easily checked by the definition of $A_j, j = 2, 3$. Now we set

$$\begin{aligned} \lambda_0 := \lambda_0(\tilde{c}) &= \inf \Big\{ x \in [\lambda_r - 3\tilde{c}, \lambda_r) : \frac{1 + \lambda_1(\Sigma) m_1(t)}{1 + \lambda_1(\Sigma) m_1(\lambda_r)} \ge 1/2 \\ &\qquad \text{and } m_1(t) \in [-2c_1, -c_1], \forall x \le t < \lambda_r \Big\}. \end{aligned} \tag{2.17}$$

Now we claim when $\tilde{c}$ is sufficiently small, there exists

$$\lambda_r - \lambda_0 \ge 2\tilde{c}. \tag{2.18}$$

Otherwise, we can assume $\lambda_r - \lambda_0 \leq 2\tilde{c}$ for arbitrary small constant $\tilde{c}$ thus $\lambda_0 > \lambda_r - 3\tilde{c}$. Then by continuity we have

$$1 + \lambda_1(\Sigma)m_1(\lambda_0) = 1/2(1 + \lambda_1(\Sigma)m_1(\lambda_r)), \quad \text{or} \quad m_1(\lambda_0) = -2c_1 \text{ or } -c_1$$

If $1 + \lambda_1(\Sigma)m_1(\lambda_0) = 1/2(1 + \lambda_1(\Sigma)m_1(\lambda_r))$, by using (2.16) we have

$$c' \leq |(1 + \lambda_1(\Sigma)m_1(\lambda_0)) - (1 + \lambda_1(\Sigma)m_1(\lambda_r))| \leq C|1 + \lambda_1(\Sigma)m_1(\lambda_\xi)|^{-3}\tilde{c}$$

for some $\lambda_\xi \in [\lambda_0, \lambda_r]$. However, by the definition of λ_0, we get a contradiction if $\tilde{c}$ is selected to be sufficiently small. If $m_1(\lambda_0) = -2c_1$ or $-c_1$, we see

$$\frac{1}{4}c_1 \leq |m_1(\lambda_r) - m_1(\lambda_0)| \leq C|1 + \lambda_1(\Sigma)m_1(\lambda_\xi)|^{-3}\tilde{c}.$$

Then by definition of λ_0, we also get a contradiction when $\tilde{c}$ is small enough. Thus we conclude the proof. □

Now we come to prove Lemma 2.5.

Proof of Lemma 2.5. We define the largest endpoint of the support of ρ_0 smaller than λ_r by λ_{r-}. It suffices to show that there exists a sufficiently small constant $\tilde{c}$ such that

$$\lambda_r - \lambda_{r-} \geq 4\tilde{c}. \tag{2.19}$$

Note that by (2.8), we know when λ is an endpoint of the support of ρ_0 and $\lambda \geq c_2$ with any fixed positive number c_2, there must be

$$\int \frac{t^2 m_1^2(\lambda)}{(1 + tm_1(\lambda))^2} dH_N(t) = d_N, \quad m_2(\lambda) = 0. \tag{2.20}$$

In the sequel we assume $\lambda_{r-} \geq c_2$ for some fixed positive number c_2, otherwise (2.19) holds naturally. We already know that $m_1(\lambda_r)$ is the unique solution of the equation

$$\int \frac{t^2 x^2}{(1 + tx)^2} dH_N(t) = d_N$$

in $(-1/\lambda_1(\Sigma), 0)$. Thus we have

$$m_1(\lambda_{r-}) \in \mathbb{R} \setminus (-1/\lambda_1(\Sigma), 0). \tag{2.21}$$

By (iii) of Condition 1.1, we have

$$-1/\lambda_1(\Sigma) - c'' \leq m_1(\lambda_r) \leq -\frac{5}{4}c_1 \leq -c'' \tag{2.22}$$

for some small positive constant c''. Here the upper bound in (2.22) follows from (2.14). Hence, (2.21) and (2.22) imply

$$|m_1(\lambda_r) - m_1(\lambda_{r-})| \geq c''. \tag{2.23}$$

Now use (2.4) we have

$$\lambda_r - \lambda_{r-} = \frac{|(m_1(\lambda_r) - m_1(\lambda_{r-}))|}{|m_1(\lambda_r)m_1(\lambda_{r-})|} \left|1 - d_N^{-1} \int \frac{t^2 m_1(\lambda_r)m_1(\lambda_{r-})dH_N(t)}{(1 + tm_1(\lambda_r))(1 + tm_1(\lambda_{r-}))}\right| \tag{2.24}$$

Note that from (2.5) we have

$$m_1(\lambda_{r-}), m_1(\lambda_r) \sim 1.$$

Then by (2.23) it suffices show that there exists some positive constant c''' such that

$$\left|1 - d_N^{-1} \int \frac{t^2 m_1(\lambda_r) m_1(\lambda_{r-}) dH_N(t)}{(1+tm_1(\lambda_r))(1+tm_1(\lambda_{r-}))}\right| \geq c'''. \tag{2.25}$$

Now we come to verify (2.25). Note that by (2.20) we have

$$\begin{aligned}
&\left|1 - d_N^{-1} \int \frac{t^2 m_1(\lambda_r) m_1(\lambda_{r-}) dH_N(t)}{(1+tm_1(\lambda_r))(1+tm_1(\lambda_{r-}))}\right| \\
= \; & d_N^{-1} \left| \int \frac{t^2 m_1^2(\lambda_{r-}) dH_N(t)}{(1+tm_1(\lambda_{r-}))^2} - \int \frac{t^2 m_1(\lambda_r) m_1(\lambda_{r-}) dH_N(t)}{(1+tm_1(\lambda_r))(1+tm_1(\lambda_{r-}))} \right| \\
= \; & d_N^{-1} \left| \frac{m_1(\lambda_{r-}) - m_1(\lambda_r)}{m_1(\lambda_{r-})} \int \frac{t^2 m_1^2(\lambda_{r-}) dH_N(t)}{(1+tm_1(\lambda_{r-}))^2 (1+tm_1(\lambda_r))} \right|.
\end{aligned}$$

Moreover, by assumption we have

$$1 + tm_1(\lambda_r) \geq c', \quad \forall t \in Spec(\Sigma).$$

Combining these facts with (2.23) and (2.20) for $m_1(\lambda_{r-})$, we obtain (2.25) with some sufficiently small $c''' > 0$. Inserting (2.23) and (2.25) into (2.24) we obtain that (2.19) holds with some positive constant $\tilde{c} := \tilde{c}(c'', c''')$, which implies Lemma 2.5. □

With the aid of Lemma 2.1, now we can start to prove Lemma 2.3.

Proof of Lemma 2.3. Note that by definition,

$$\frac{\Im m_0(z)}{\eta} = \int \frac{1}{(x-E)^2 + \eta^2} dF_0(x),$$

Thus the second inequality of (iii) follows directly. It suffices to verify (i), (ii), (iv) and the first inequality of (iii) in the sequel. At first, we come to show (i). Note that when $z \in S(0)$,

$$C_l \leq E \leq C_r.$$

Then it is easy to obtain (i) by (2.4).

Now we come to verify (ii). The proof of (ii) relies on Lemma 2.1. By definition we have

$$m_0(z) = \int \frac{1}{x-z} dF_0(x).$$

At first we deal with the case of

$$E \geq \lambda_r + \eta.$$

We do it as follows.

$$\begin{aligned}
\Im m_0(z) &= \int_0^{\lambda_r} \frac{\eta}{(x-E)^2 + \eta^2} dF_0(x) \\
&= \int_{C_l/2}^{\lambda_r} \frac{\eta}{(x-E)^2 + \eta^2} \rho_0(x) dx + O(\eta)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{C_l/2}^{\lambda_r-\tilde{c}} \frac{\eta}{(x-E)^2+\eta^2}\rho_0(x)dx + \int_{\lambda_r-\tilde{c}}^{\lambda_r} \frac{\eta}{(x-E)^2+\eta^2}\rho_0(x)dx + O(\eta) \\
&\sim \eta + \int_{\lambda_r-\tilde{c}}^{\lambda_r} \frac{\eta}{(\lambda_r-x)^2+(\kappa+\eta)^2}\sqrt{\lambda_r-x}dx \\
&= \eta + \int_0^{\tilde{c}} \frac{\eta}{t^2+(\kappa+\eta)^2}\sqrt{t}dt
\end{aligned}
$$

Now if $\tilde{c} \geq \kappa+\eta$, we have

$$
\begin{aligned}
\int_0^{\tilde{c}} \frac{\eta}{t^2+(\kappa+\eta)^2}\sqrt{t}dt &= \int_0^{\kappa+\eta} \frac{\eta}{t^2+(\kappa+\eta)^2}\sqrt{t}dt + \int_{\kappa+\eta}^{\tilde{c}} \frac{\eta}{t^2+(\kappa+\eta)^2}\sqrt{t}dt \\
&\sim \int_0^{\kappa+\eta} \frac{\eta}{(\kappa+\eta)^2}\sqrt{t}dt + \int_{\kappa+\eta}^{\tilde{c}} \frac{\eta}{t^2}\sqrt{t}dt \\
&\sim \frac{\eta}{\sqrt{\kappa+\eta}}.
\end{aligned}
$$

If $\tilde{c} < \kappa+\eta$, then we have

$$
\int_0^{\tilde{c}} \frac{\eta}{t^2+(\kappa+\eta)^2}\sqrt{t}dt \sim \int_0^{\kappa+\eta} \frac{\eta}{(\kappa+\eta)^2}\sqrt{t}dt \sim \frac{\eta}{\sqrt{\kappa+\eta}}.
$$

Now we come to deal with the case of

$$
E \in [\lambda_r-\tilde{c}, \lambda_r+\eta).
$$

Note that our discussion for the case of $E \geq \lambda_r+\eta$ can be extended to the case of $E \geq \lambda_r$. Actually, in the region $E \in [\lambda_r, \lambda_r+\eta]$, $\kappa \leq \eta$, thus one has

$$
\frac{\eta}{\sqrt{\kappa+\eta}} \sim \sqrt{\kappa+\eta}.
$$

Therefore, it suffices to handle the case of $E \in [\lambda_r-\tilde{c}, \lambda_r)$. Note

$$
\begin{aligned}
\Im m_0(z) &= \int_0^{\lambda_r} \frac{\eta}{(x-E)^2+\eta^2}dF_0(x) \\
&= \int_{C_l/2}^{\lambda_r} \frac{\eta}{(x-E)^2+\eta^2}\rho_0(x)dx + O(\eta) \\
&= \int_{C_l/2}^{\lambda_r-2\tilde{c}} \frac{\eta}{(x-E)^2+\eta^2}\rho_0(x)dx + \int_{\lambda_r-2\tilde{c}}^{\lambda_r} \frac{\eta}{(x-E)^2+\eta^2}\rho_0(x)dx + O(\eta) \\
&\sim \eta + \int_{\lambda_r-2\tilde{c}}^{\lambda_r} \frac{\eta}{(x-E)^2+\eta^2}(\sqrt{|x-E|}+\sqrt{\kappa})dx.
\end{aligned}
$$

By splitting the integral region $[\lambda_r-2\tilde{c}, \lambda_r]$ into two parts by $|x-E| \geq \kappa$ or $|x-E| < \kappa$, it is easy to see

$$
\Im m_0(z) = O(\sqrt{\kappa+\eta}).
$$

Thus we proved (ii).

Now we start to show the first inequality of (iii) whose proof will also relies on Lemma 2.1. Note that by discussions above we have

$$
\Im m_0(z) \geq \int_{C_l/2}^{\lambda_r} \frac{\eta}{(x-E)^2+\eta^2}\rho_0(x)dx \geq \int_{\lambda_r-2\tilde{c}}^{\lambda_r} \frac{\eta}{(x-E)^2+\eta^2}\rho_0(x)dx.
$$

Then it is obvious that

$$\Im m_0(z) \geq C\eta, \tag{2.26}$$

since

$$\rho_0(x) \sim \sqrt{\lambda_r - x} \sim 1, \quad x \in [\lambda_r - 2\tilde{c}, \lambda_r - \tilde{c}].$$

At the end, we come to prove (iv). At first, we claim that when $\tilde{c}$ is small enough, there exists

$$\inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]} |1 + tm_0(x)| \geq 1/2(1 + \lambda_1(\Sigma)m_0(\lambda_r)), \quad \forall x \in [\lambda_r - \tilde{c}, C_r]. \tag{2.27}$$

To see (2.27), we split the interval into $(\lambda_r, C_r]$ and $[\lambda_r - \tilde{c}, \lambda_r]$. For the first case, it is not difficult to see $m_0(x) = m_1(x)$ is negative and increasing. Thus

$$\begin{aligned} \inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]} |1 + tm_0(x)| &= \inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]} (1 + tm_0(x)) \\ &= 1 + \lambda_1(\Sigma)m_0(x) \geq 1 + \lambda_1(\Sigma)m_0(\lambda_r) \end{aligned} \tag{2.28}$$

when $x \geq \lambda_r$. For the second case, we recall λ_0 defined in (2.17) and the inequality (2.18). Then it is obvious that

$$\begin{aligned} \inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]} |1 + tm_0(x)| &\geq \inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]} (1 + tm_1(x)) \\ &= 1 + \lambda_1(\Sigma)m_1(x) \geq \frac{1}{2}(1 + \lambda_1(\Sigma)m_0(\lambda_r)) \end{aligned} \tag{2.29}$$

when $x \in [\lambda_r - \tilde{c}, \lambda_r]$. Thus (2.27) follows. Moreover, (2.28) and (2.29) also imply that

$$\inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]} (1 + tm_1(x)) \geq \frac{1}{2}(1 + \lambda_1(\Sigma)m_0(\lambda_r)), \quad x \in [\lambda_r - \tilde{c}, C_r]. \tag{2.30}$$

Now we extend (2.27) and (2.30) from the real line to the full region $S_r(\tilde{c}, 0)$. At first, we use an elementary inequality which can be found in [5] (see the proof of Lemma 6.10 therein),

$$|(m(z)x + 1)^{-1}| \leq \max\left(\frac{4x}{\eta}, 2\right), \tag{2.31}$$

where $m(z) := m(E + i\eta)$ can be the Stieltjes transform of arbitrary probability measure and x can be any positive number. By (2.31), we know that

$$\inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]} |1 + tm_0(z)| \geq \min\left(\frac{\eta}{4\lambda_1(\Sigma)}, \frac{1}{2}\right).$$

Therefore, it suffices to show (2.2) when η is sufficiently small. To this end, we will combine (2.30) and a bound of derivative of $m_1(z)$ with respect to η which can be obtained as follows. By definition, we have for any $\eta > 0$,

$$m_1(z) = \int \frac{x - E}{(x - E)^2 + \eta^2} dF_0(x), \quad \partial_\eta m_1(z) = -2\int \frac{(x - E)\eta}{((x - E)^2 + \eta^2)^2} dF_0(x).$$

Now let α be a small positive constant. At first, we handle the case of $E \in [\lambda_r - \tilde{c}, \lambda_r]$. We split the estimation of $\partial_\eta m_1(z)$ into two cases: $\kappa \leq 2\eta^{1-\alpha}$ and $\kappa \geq 2\eta^{1-\alpha}$. Note that for the first case, we have

$$|\partial_\eta m_1(z)| = 2\left|\left(\int_{C_l/2}^{E-\eta^{1-\alpha}} + \int_{E-\eta^{1-\alpha}}^{\lambda_r}\right) \frac{(x - E)\eta}{((x - E)^2 + \eta^2)^2}\rho_0(x)dx\right| + O(\eta)$$

$$\begin{aligned}
&\le C\left(\int_{C_l/2}^{E-\eta^{1-\alpha}} \frac{\eta}{(E-x)^3}dx + \int_{E-\eta^{1-\alpha}}^{\lambda_r} \eta^{2-\alpha}\cdot\eta^{-4}\cdot\eta^{(1-\alpha)/2}dx\right) + O(\eta) \\
&= C(\eta^{-1+2\alpha} + \eta^{-1/2-5\alpha/2}) \\
&= O(\eta^{-1+2\alpha})
\end{aligned}$$

when α is chosen to be sufficiently small.

When $\kappa \ge 2\eta^{1-\alpha}$, we have

$$\begin{aligned}
|\partial_\eta m_1(z)| &= 2\left|\left(\int_{C_l/2}^{E-\eta^{1-\alpha}} + \int_{E+\eta^{1-\alpha}}^{\lambda_r} + \int_{E-\eta^{1-\alpha}}^{E+\eta^{1-\alpha}}\right)\frac{(x-E)\eta}{((x-E)^2+\eta^2)^2}\rho_0(x)dx\right| + O(\eta) \\
&\le 2\left|\int_{E-\eta^{1-\alpha}}^{E+\eta^{1-\alpha}} \frac{(x-E)\eta}{((x-E)^2+\eta^2)^2}\rho_0(x)dx\right| + C\eta^{-1+2\alpha} \\
&= 2\left|\int_{E-\eta^{1-\alpha}}^{E+\eta^{1-\alpha}} \frac{(x-E)\eta}{((x-E)^2+\eta^2)^2}(\rho_0(E)+\rho_0'(\xi_{\{E,x\}})(x-E))dx\right| + C\eta^{-1+2\alpha},
\end{aligned}$$

by the mean value theorem, where $\xi_{\{E,x\}}$ is some real number between E and x. Note the analyticity of $\rho_0(x)$ in its support has been proved in [28]. Moreover, we have

$$|\rho_0'(t)| \sim (\lambda_r - t)^{-1/2}, \quad t \in [\lambda_r - \tilde{c}, \lambda_r)$$

which is implied by (2.3) and Lemma 2.1. Thus by the assumption of $\kappa \ge 2\eta^{1-\alpha}$ we have

$$\begin{aligned}
|\partial_\eta m_1(z)| &\le 2\int_{E-\eta^{1-\alpha}}^{E+\eta^{1-\alpha}} \frac{(x-E)^2\eta}{((x-E)^2+\eta^2)^2}|\rho_0'(\xi_{\{E,x\}})|dx + C\eta^{-1+2\alpha} \\
&\le C\left(\int_{E-\eta^{1-\alpha}}^{E+\eta^{1-\alpha}} \eta^{3-2\alpha}\cdot\eta^{-4}\cdot\eta^{-1/2+\alpha/2}dx + \eta^{-1+2\alpha}\right) \\
&\le O(\eta^{-1+2\alpha})
\end{aligned}$$

when α is chosen to be sufficiently small.

The case of $E \in (\lambda_r, C_r)$ is similar and simpler, thus we omit the details. Actually, for $E \in [\lambda_r - \tilde{c}, C_r]$, we have

$$|\partial_\eta m_1(z)| \le O(\eta^{-1+2\alpha})$$

for all $\eta \le \hat{\eta}$ with some some sufficiently small constant $\hat{\eta}$.

Now we recall (2.27) and fix an $E \in [\lambda_r - \tilde{c}, C_r]$. Since $m_0(E+i\eta)$ is continuous for $\eta \ge 0$, there exists some small positive η_0 such that

$$1 + tm_1(E+i\eta_0) \ge \frac{1}{2}(1+tm_1(E)) \ge \frac{1}{4}(1+\lambda_1(\Sigma)m_1(\lambda_r)), \quad t \in [\lambda_M(\Sigma), \lambda_1(\Sigma)].$$

Now consider $E + i\eta \in S_r(\tilde{c}, 0)$. Note that we have mentioned above it suffices to consider the case where η is sufficiently small. Observe that we have

$$|m_1(E+i\eta) - m_1(E+i\eta_0)| = \left|\int_{\eta_0}^{\eta} \partial_t m_1(E+it)dt\right| \le C\left|\int_{\eta_0}^{\eta} t^{-1+2\alpha}dt\right| \le C\eta^{2\alpha}.$$

Therefore, when η is sufficiently small, we have

$$\inf_{t\in[\lambda_M(\Sigma),\lambda_1(\Sigma)]}(1+tm_1(E+i\eta)) \ge \hat{c}(1+\lambda_1(\Sigma)m_1(\lambda_r)) \ge c_0$$

for some small positive constant $\hat{c}, c_0$. Therefore, we conclude the proof. □

3. Asymptotic analysis of $m_N(z)$. In this section, we will prove the strong local MP type law around λ_r. To this end, we will derive a self-consistent equation for $m_N(z)$, which is quite close to that of $m_0(z)$ (see (1.6)). Then we can figure out the closeness of $m_N(z)$ and $m_0(z)$ through studying the stability of the self-consistent equation of $m_0(z)$ via pursuing the argument in [11] with slight modification. The main proof route in this section is parallel to that in [27]. However, owing to the generality of Σ, most parts require more general treatments.

To simplify some discussions in the sequel, we will truncate and renormalize $\sqrt{N}x_{ij}$ at first. Let C_0' be some sufficiently large positive constant. We set

$$\hat{x}_{ij} = x_{ij}\mathbf{1}_{\{|\sqrt{N}x_{ij}|\le(\log N)^{C_0'}\}}, \quad \tilde{x}_{ij} = \frac{\hat{x}_{ij}-\mathbb{E}\hat{x}_{ij}}{\sqrt{Var(\hat{x}_{ij})}}.$$

Now set $\hat{X} := (\hat{x}_{ij})_{M,N}$ and $\tilde{X} := (\tilde{x}_{ij})_{M,N}$. Correspondingly, let

$$\hat{\mathcal{W}}_N := \Sigma^{1/2}\hat{X}\hat{X}^*\Sigma^{1/2}, \quad \tilde{\mathcal{W}}_N := \Sigma^{1/2}\tilde{X}\tilde{X}^*\Sigma^{1/2}.$$

Moreover, by (1.9) we see

$$\mathcal{W}_N = \hat{\mathcal{W}}_N \tag{3.1}$$

with probability larger than

$$1 - O\left(N^2\exp(-(\log N)^{\tau_0 C_0'})\right) \gg 1-\exp(-(\log N)^2) \tag{3.2}$$

if we choose C_0' to be sufficiently large. Moreover, note that

$$\tilde{x}_{ij} = \hat{x}_{ij} + O(\exp(-(\log N)^{CC_0'})).$$

Thus by using basic perturbation theory of eigenvalues such as Weyl's inequality, we have

$$\max_i |\lambda_i(\mathcal{W}_N) - \lambda_i(\tilde{\mathcal{W}}_N)| \le N^{O(1)}\exp(-(\log N)^{CC_0'}). \tag{3.3}$$

Consequently, by (3.1) and (3.3) we can work on $\tilde{\mathcal{W}}_N$ instead of $\mathcal{W}_N$. For ease of presentation, we recycle the notation X and $\mathcal{W}_N$ to denote $\tilde{X}$ and $\tilde{\mathcal{W}}_N$, and we also denote the truncated and renormalized variable $\tilde{x}_{ij}$ by x_{ij} in the sequel. Thus without loss of generality, we will assume

$$\max_{i,j}|\sqrt{N}x_{ij}| \le (\log N)^{C_0'}. \tag{3.4}$$

Now we introduce some notation. We denote the Green functions of W_N and $\mathcal{W}_N$ respectively by

$$G_N(z) := (W_N - z)^{-1}, \quad \mathcal{G}_N(z) = (\mathcal{W}_N - z)^{-1},$$

where $z \in \mathbb{C}^+$. For ease of presentation, when there is no confusion we will omit the subscript N or variable z from the above notation. Note that by definition we have

$$m_N(z) = \frac{1}{N}TrG(z), \quad \underline{m}_N(z) = \frac{1}{M}Tr\mathcal{G}(z),$$

and

$$TrG(z) - Tr\mathcal{G}(z) = \frac{M-N}{z}.$$

Furthermore, we use $\mathbf{x}_i$ to denote the i-th column of X, and introduce the notation $X^{(\mathbb{T})}$ to denote the $M \times (N - |\mathbb{T}|)$ minor of X obtained by deleting $\mathbf{x}_i$ from X if $i \in \mathbb{T}$. For convenience, we will briefly write $(\{i\})$ and $(\{i,j\})$ as (i) and (ij) respectively. Correspondingly, we denote

$$W^{(\mathbb{T})} = X^{(\mathbb{T})*}\Sigma X^{(\mathbb{T})}, \quad \mathcal{W}^{(\mathbb{T})} = \Sigma^{1/2}X^{(\mathbb{T})}X^{(\mathbb{T})*}\Sigma^{1/2}$$

and

$$G^{(\mathbb{T})}(z) = (W^{(\mathbb{T})} - z)^{-1}, \quad \mathcal{G}^{(\mathbb{T})}(z) = (\mathcal{W}^{(\mathbb{T})} - z)^{-1}.$$
$$m_N^{(\mathbb{T})}(z) = \frac{1}{N}trG^{(\mathbb{T})}(z), \quad \underline{m}_N^{(\mathbb{T})}(z) = \frac{1}{M}Tr\mathcal{G}^{(\mathbb{T})}(z).$$

In the sequel, we will keep the names of indices of X for $X^{(\mathbb{T})}$. That means

$$X_{ij}^{(\mathbb{T})} = \mathbf{1}_{\{j \notin \mathbb{T}\}}X_{ij}.$$

Correspondingly, we will denote (i,j)-th entry of $G^{(\mathbb{T})}(z)$ by $G_{ij}^{(\mathbb{T})}(z)$ for all $i,j \notin \mathbb{T}$. Similarly, we remind here the index (i,j) is not in the conventional sense. Note that by definition, $G^{(\mathbb{T})}$ is an $(N - |\mathbb{T}|) \times (N - |\mathbb{T}|)$ matrix. However, here we use the index set $\{1,\ldots,N\} \setminus \mathbb{T}$ instead of $\{1,\ldots,N - |\mathbb{T}|\}$. Set

$$\mathbf{r}_i = \Sigma^{1/2}\mathbf{x}_i.$$

At first, we state the following Lemma which collects some basic formulas on the entries of Green functions.

LEMMA 3.1. *Under the above notation, we have*
(i):

$$G_{ii}(z) = -\frac{1}{z + z\mathbf{r}_i^*\mathcal{G}^{(i)}(z)\mathbf{r}_i}.$$

(ii):

$$G_{ij}(z) = zG_{ii}(z)G_{jj}^{(i)}(z)\mathbf{r}_i^*\mathcal{G}^{(ij)}(z)\mathbf{r}_j, \quad i \neq j.$$

(iii):

$$G_{ij}(z) = G_{ij}^{(k)}(z) + \frac{G_{ik}(z)G_{kj}(z)}{G_{kk}(z)}, \quad i,j \neq k.$$

PROOF. One can refer to Lemma 2.3 of [27] or Lemma 3.2 of [21] for instance. Actually, if we regard $\mathbf{r}_i$ as $\mathbf{x}_i$ in [27], then G has the same structure as that in the null case in [27], so does $\mathcal{G}$. □

By Lemma 3.1 we see that

$$m_N(z) = \frac{1}{N}\sum_{i=1}^{N} G_{ii}(z) = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z + z\mathbf{r}_i^*\mathcal{G}^{(i)}(z)\mathbf{r}_i}. \tag{3.5}$$

Now we write

$$m_N(z) = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z - \frac{1}{N}Tr(m_N(z)\Sigma + I)^{-1}\Sigma + Y_i}, \tag{3.6}$$

where

$$Y_i = z\mathbf{r}_i^* \mathcal{G}^{(i)}(z)\mathbf{r}_i + \frac{1}{N} Tr(m_N(z)\Sigma + I)^{-1}\Sigma.$$

Observe that

$$\frac{1}{N} Tr(m_N(z)\Sigma + I)^{-1}\Sigma = d_N^{-1} \int \frac{t}{1 + tm_N(z)} dH_N(t).$$

Below we will use the decomposition

$$Y_i = z(T_i + U_i + V),$$

where

$$\begin{aligned} T_i := T_i(z) &= \mathbf{r}_i^* \mathcal{G}^{(i)}(z)\mathbf{r}_i - \frac{1}{N} Tr\mathcal{G}^{(i)}(z)\Sigma, \\ U_i := U_i(z) &= \frac{1}{N} Tr\mathcal{G}^{(i)}(z)\Sigma - \frac{1}{N} Tr\mathcal{G}(z)\Sigma, \\ V := V(z) &= \frac{1}{N} Tr\mathcal{G}(z)\Sigma - \frac{1}{N} Tr(-zm_N(z)\Sigma - zI)^{-1}\Sigma. \end{aligned} \tag{3.7}$$

Note that by definition

$$\mathcal{W} = \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^*, \quad \mathcal{G}(z) = (\sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^* - zI)^{-1}.$$

Thus

$$\begin{aligned} & \mathcal{G}(z) - (-zm_N(z)\Sigma - zI)^{-1} \\ = \quad & -(-zm_N(z)\Sigma - zI)^{-1} \left[\sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^* - (-zm_N(z))\Sigma \right] \mathcal{G}(z). \end{aligned}$$

Now using the Sherman-Morrison formula

$$\mathbf{r}^*(\mathbf{C} + \mathbf{r}\mathbf{r}^*)^{-1} = \frac{1}{1 + \mathbf{r}^*\mathbf{C}^{-1}\mathbf{r}} \mathbf{r}^*\mathbf{C}^{-1}$$

for any invertable matrix $\mathbf{C}$, we obtain

$$\mathbf{r}_i \mathbf{r}_i^* \mathcal{G}(z) = \frac{1}{1 + \mathbf{r}_i^* \mathcal{G}^{(i)}(z)\mathbf{r}_i} \mathbf{r}_i \mathbf{r}_i^* \mathcal{G}^{(i)}(z).$$

Therefore, we have

$$\begin{aligned} & -(-zm_N(z)\Sigma - zI)^{-1} \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^* \mathcal{G}(z) \\ = \quad & \sum_{i=1}^N \frac{1}{z(1 + \mathbf{r}_i^* \mathcal{G}^{(i)}(z)\mathbf{r}_i)} \cdot (m_N(z)\Sigma + I)^{-1} \mathbf{r}_i \mathbf{r}_i^* \mathcal{G}^{(i)}(z). \end{aligned} \tag{3.8}$$

Taking (3.5) into account we obtain

$$\begin{aligned} & -(-zm_N(z)\Sigma - zI)^{-1}(-zm_N(z))\Sigma\mathcal{G}(z) \\ = \quad & \sum_{i=1}^N \frac{1}{z(1 + \mathbf{r}_i^* \mathcal{G}^{(i)}(z)\mathbf{r}_i)} \cdot \frac{1}{N} (m_N(z)\Sigma + I)^{-1} \Sigma\mathcal{G}(z). \end{aligned} \tag{3.9}$$

Combining (3.8) and (3.9) we obtain

$$\begin{aligned}
&\mathcal{G}(z)-(-zm_N(z)\Sigma-zI)^{-1}\\
=&\sum_{i=1}^N\frac{1}{z(1+\mathbf{r}_i^*\mathcal{G}^{(i)}(z)\mathbf{r}_i)}\cdot\left[(m_N(z)\Sigma+I)^{-1}\mathbf{r}_i\mathbf{r}_i^*\mathcal{G}^{(i)}(z)-\frac{1}{N}(m_N(z)\Sigma+I)^{-1}\Sigma\mathcal{G}(z)\right]
\end{aligned}$$

Therefore, we can further decompose $V(z)$ defined in (3.7) into four parts.

$$V(z):=v_1(z)+v_2(z)+v_3(z)+v_4(z),$$

where

$$\begin{aligned}
v_1(z)&:=-\frac{1}{N}\sum_{i=1}^N G_{ii}\left[\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma+I)^{-1}\mathbf{r}_i-\frac{1}{N}Tr(m_N^{(i)}(z)\Sigma+I)^{-1}\Sigma\mathcal{G}^{(i)}\Sigma\right],\\
v_2(z)&:=-\frac{1}{N}\sum_{i=1}^N G_{ii}\left[\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N(z)\Sigma+I)^{-1}\mathbf{r}_i-\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma+I)^{-1}\mathbf{r}_i\right],\\
v_3(z)&:=-\frac{1}{N}\sum_{i=1}^N G_{ii}\left[\frac{1}{N}Tr(m_N^{(i)}(z)\Sigma+I)^{-1}\Sigma\mathcal{G}^{(i)}\Sigma-\frac{1}{N}Tr(m_N^{(i)}(z)\Sigma+I)^{-1}\Sigma\mathcal{G}\Sigma\right],\\
v_4(z)&:=-\frac{1}{N}\sum_{i=1}^N G_{ii}\left[\frac{1}{N}Tr(m_N^{(i)}(z)\Sigma+I)^{-1}\Sigma\mathcal{G}\Sigma-\frac{1}{N}Tr(m_N(z)\Sigma+I)^{-1}\Sigma\mathcal{G}\Sigma\right].
\end{aligned}$$

Moreover, we will denote

$$v_{1i}:=v_{1i}(z)=\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma+I)^{-1}\mathbf{r}_i-\frac{1}{N}Tr(m_N^{(i)}(z)\Sigma+I)^{-1}\Sigma\mathcal{G}^{(i)}\Sigma, \tag{3.10}$$

thus

$$v_1(z)=-\frac{1}{N}\sum_{i=1}^N G_{ii}v_{1i}.$$

In the sequel, we will also encounter the quantity

$$\begin{aligned}
R_i(z)\ :=\ &\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma+I)^{-1}\Sigma(m_0(z)\Sigma+I)^{-1}\mathbf{r}_i\\
&-\frac{1}{N}Tr\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma+I)^{-1}\Sigma(m_0(z)\Sigma+I)^{-1}\Sigma.
\end{aligned}$$

Observe that if we can show that Y_i's are small enough, (3.6) turns out to be close to (1.6). Roughly speaking, our main task in the sequel is to bound the quantities

$$|Y_i(z)|,\quad i=1,\ldots,N$$

in some region we are interested in. Then we will take a step further to figure out the closeness of $m_N(z)$ and $m_0(z)$. Specifically, We will provide bounds for the following quantities

$$\Lambda_d(z):=\max_i|G_{ii}(z)-m_0(z)|,\quad \Lambda_o(z):=\max_{i\neq j}|G_{ij}(z)|,\quad \Lambda(z):=|m_N(z)-m_0(z)|$$

for all $z\in S_r(\tilde{c},C)$ with some positive constant C. Our target in this section is to prove the following theorem which is a slight modification of Theorem 1.4 under the additional condition (3.4).

THEOREM 3.2 (Strong local MP type law around λ_r for truncated matrix). *Under Condition 1.1 and (3.4), for any $\zeta > 0$ there exists some constant C_ζ such that*

(i):

$$\bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \left\{ \Lambda(z) \le \varphi^{C_\zeta} \frac{1}{N\eta} \right\} \tag{3.11}$$

holds with ζ-high probability, and

(ii):

$$\bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \left\{ \Lambda_o(z) + \Lambda_d(z) \le \varphi^{C_\zeta} \left(\sqrt{\frac{\Im m_0(z)}{N\eta}} + \frac{1}{N\eta} \right) \right\} \tag{3.12}$$

holds with ζ-high probability.

PROOF OF THEOREM 1.4 ASSUMING THEOREM 3.2. Note that by the definition of $m_N(z)$, the fact of $\eta \gg N^{-1}$, (3.2) and (3.3) we see that the difference between $m_N(z)$ of the original matrix and that of the truncated one is smaller that $O(N^{-C})$ with overwhelming probability for any positive constant C when N is large enough.Then we can easily recover (1.13) from (3.11). Analogously, by spectral decomposition we can also easily see that (1.14) can be recovered from (3.12). Thus we complete the proof of Theorem 1.4 assuming its truncated version Theorem 3.2. □

Roughly speaking, in this part, we will adopt the proof route of that for Theorem 3.1 of [27]. At first, we will provide a weak bound for the strong local MP type law around the right edge. And then we use the weak bound to get the strong bound. As mentioned in Introduction, such a bootstrap strategy was developed in a series of work [19, 21, 12, 27]. However, since the generality of our setting on Σ, most parts of the proof need new techniques thus the details are relatively different from those of the null case in [27]. Especially, in the discussion of the stability of the self-consistent equation of $m_0(z)$, we will extend an idea from [11] to our case.

3.1. *Weak local MP type law around λ_r.* In this subsection, we will prove the following weak local MP type law in the region around λ_r.

THEOREM 3.3. *Under Condition 1.1 and (3.4), for any $\zeta > 0$, there exists some positive constant C_ζ such that the event*

$$\bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \left\{ \Lambda_d(z) + \Lambda_o(z) \le \varphi^{C_\zeta} \frac{1}{(N\eta)^{1/4}} \right\}.$$

holds with ζ-high probability.

At first, it follows from the definitions that

$$\Lambda(z) \le \Lambda_d(z) \le \max_i |G_{ii}(z) - m_N(z)| + \Lambda(z). \tag{3.13}$$

For $z \in S_r(\tilde{c}, 5C_\zeta)$ and positive number K we define the event

$$\Omega(z, K) := \left\{ \max \left\{ \Lambda_o(z), \max_i |G_{ii}(z) - m_N(z)|, \max_i |T_i(z)|, \max_i |R_i(z)|, \max_i |v_{1i}(z)| \right\} \ge K\Psi(z) \right\}$$

$$\bigcup\left\{\max\left\{\max_i|U_i(z)|,|v_2(z)|,|v_3(z)|,|v_4(z)|\right\}\geq K^2\Psi^2(z)\right\},$$

with

$$\Psi(z):=\sqrt{\frac{\Im m_0(z)+\Lambda(z)}{N\eta}}.$$

Moreover, we set the events

$$\begin{aligned}&\Xi(z):=\{\Lambda_o(z)+\Lambda_d(z)>(\log N)^{-1}\},\\ &\Gamma(z,K):=\Omega(z,K)^c\cup\Xi(z).\end{aligned}\tag{3.14}$$

In the sequel, we will frequently use the following large deviation estimates whose proof can be found in [20].

LEMMA 3.4. *Let $\mathbf{x}_i,\mathbf{x}_j, i\neq j$ be two columns of the matrix X satisfying (ii) of Condition 1.1. Then for any $M\times 1$ vector $\mathbf{b}$ and $M\times M$ matrix $\mathbf{C}$ independent of $\mathbf{x}_i$ and $\mathbf{x}_j$, the following three inequalities hold with ζ-high probability*

(i):

$$|\mathbf{x}_i^*\mathbf{C}\mathbf{x}_i-\frac{1}{N}Tr\mathbf{C}|\leq\frac{\varphi^{\tau\zeta}}{N}||\mathbf{C}||_{HS}.$$

(ii):

$$|\mathbf{x}_i^*\mathbf{C}\mathbf{x}_j|\leq\frac{\varphi^{\tau\zeta}}{N}||\mathbf{C}||_{HS}.$$

(iii)

$$|\mathbf{b}^*\mathbf{x}_i|\leq\frac{\varphi^{\tau\zeta}}{\sqrt{N}}||\mathbf{b}||.$$

Here $\tau:=\tau(\tau_0)>1$ is some positive constant.

PROOF. See Appendix B of [20] for instance. □

To prove Theorem 3.3, we will provide the desired bound for the case of $\eta\sim 1$ at first. Then we extend it to the full region $S_r(\tilde{c},5C_\zeta)$. To fulfill the first step, by (3.13) it suffices to prove the following two lemmas.

LEMMA 3.5. *Under Condition 1.1 and (3.4), for any $\zeta>0$, there exists a constant C_ζ such that the event*

$$\bigcap_{z\in S_r(\tilde{c},5C_\zeta),\eta\sim 1}\Omega(z,\varphi^{C_\zeta})^c$$

holds with ζ-high probability.

LEMMA 3.6. *Under Condition 1.1 and (3.4), for any $\zeta>0$ there exists a constant C_ζ such that*

$$\bigcap_{z\in S_r(\tilde{c},5C_\zeta),\eta\sim 1}\{\Lambda(z)\leq\varphi^{C_\zeta}\frac{1}{(N\eta)^{1/2}}\}$$

holds with ζ-high probability.

At first, we need the following bounds on the elements of Green functions to verify Lemmas 3.5 and 3.6.

LEMMA 3.7. *When $z \in S_r(\tilde{c}, 5C_\zeta)$ with $\eta \sim 1$, for any $\mathbb{T} \subset \{1, \dots, N\}$ with $|\mathbb{T}| = O(1)$ we have*

$$G_{ii}^{(\mathbb{T})}, m_N^{(\mathbb{T})} \sim 1, \quad \Lambda_o^{(\mathbb{T})} \leq O(1), \quad \frac{1}{N} Tr|\mathcal{G}^{(\mathbb{T})}| \leq O(1) \tag{3.15}$$

with ζ-high probability.

PROOF. By spectral decomposition, we have

$$G_{ii}^{(\mathbb{T})}(z) = \sum_{k=1}^{N-|\mathbb{T}|} \frac{1}{\lambda_k(W^{(\mathbb{T})}) - z} |\mathbf{u}_{ki}(W^{(\mathbb{T})})|^2.$$

Here $\lambda_k(W^{(\mathbb{T})})$ is the k-th largest eigenvalue of $W^{(\mathbb{T})}$ by the notation in (1.2) and

$$\mathbf{u}_k(W^{(\mathbb{T})}) := (\mathbf{u}_{k1}(W^{(\mathbb{T})}), \dots, \mathbf{u}_{k,N-|\mathbb{T}|}(W^{(\mathbb{T})}))^T$$

is its corresponding unit eigenvector. Now similar to (1.12), we see $\lambda_1(W^{(\mathbb{T})})$ is bounded with ζ high probability. Thus when $z \in S_r(\tilde{c}, 5C_\zeta)$ with $\eta \sim 1$ we have

$$|G_{ii}^{(\mathbb{T})}(z)| \geq \Im G_{ii}^{(\mathbb{T})}(z) = \frac{\eta}{(\lambda_1(W^{(\mathbb{T})}) - E)^2 + \eta^2} \sum |\mathbf{u}_{ki}(W^{(\mathbb{T})})|^2 \geq C^{-1}$$

for some positive constant C with ζ-high probability. It is similar to show that $m_N^{(\mathbb{T})} \sim 1$ with ζ-high probability. Moreover, we also have the definite upper bound

$$|G_{ij}^{(\mathbb{T})}(z)| = \left| \sum_{k=1}^{N-|\mathbb{T}|} \frac{1}{\lambda_k(W^{(\mathbb{T})}) - z} \mathbf{u}_{ki}(W^{(\mathbb{T})}) \bar{\mathbf{u}}_{kj}(W^{(\mathbb{T})}) \right| \leq \frac{1}{\eta} \leq C$$

with some positive constant C. And we also have

$$\frac{1}{N} Tr|\mathcal{G}^{(\mathbb{T})}(z)| = \frac{1}{N} \sum_{l=1}^{M} \frac{1}{|\lambda_l(\mathcal{W}^{(\mathbb{T})}) - z|} \leq O(\eta^{-1}) \sim 1.$$

Thus we conclude the proof. □

Now we come to verify Lemma 3.5.

PROOF OF LEMMA 3.5. At first, we claim it suffices to show that for any fixed $z \in S_r(\tilde{c}, 5C_\zeta)$ with $\eta \sim 1$, there is some positive constant C_ζ independent of z such that

$$\max \left\{ \Lambda_o(z), \max_i |G_{ii}(z) - m_N(z)|, \max_i |T_i(z)|, \max_i |R_i(z)|, \max_i |v_{1i}(z)| \right\} < \varphi^{C_\zeta} \Psi(z) \tag{3.16}$$

and

$$\max \left\{ \max_i |U_i(z)|, |v_2(z)|, |v_3(z)|, |v_4(z)| \right\} < \varphi^{C_\zeta} \Psi^2(z) \tag{3.17}$$

hold with ζ-high probability. At first, by (2.31) we have

$$||(m_N^{(i)}(z)\Sigma+I)^{-1}||_{op}, ||m_N(z)\Sigma+I||_{op} \le \max\left(\frac{4||\Sigma||_{op}}{\eta}, 2\right) \sim 1 \tag{3.18}$$

when $\eta \sim 1$. Moreover, we have already truncated the variables $\sqrt{N}x_{ij}$ at $(\log N)^{C_0'}$. Then it is easy to check the derivatives of the quantities $G_{ij}(z)$,$T_i(z)$,$U_i(z)$, $R_i(z)$, $v_{1i}(z)$, $v_2(z)$, $v_3(z)$, $v_4(z)$ with respect to z are all bounded by η^{-A_0} in magnitude with some positive constant A_0. Now we can assign an ε-net on the region $S_r(\tilde{c}, 5C_\zeta)$ with $\varepsilon = N^{-100A_0}$ (say) . Then it suffices to show (3.16) and (3.17) for all z in this ε-net. By the definition of ζ-high probability, it suffices to prove (3.16) and (3.17) for any fixed z.

Note that when $\eta \sim 1$, by using (3.15) we have

$$G_{ii}(z) \sim 1, \quad m_N(z) \sim 1 \tag{3.19}$$

with ζ-high probability. Using Lemma 2.3 and (3.19) we have

$$\Psi(z) = O(N^{-1/2}), \quad \eta \sim 1.$$

By the definition of $\Omega(z, K)$, we should bound the following quantities

$$\Lambda_o(z), \max_i |G_{ii}(z) - m_N(z)|, \max_i |T_i(z)|, \max_i |R_i(z)|,$$
$$\max_i |v_{1i}(z)|, \max_i |U_i(z)|, |v_2(z)|, |v_3(z)|, |v_4(z)|$$

when $\eta \sim 1$ one by one. We do it as follows.

At first, we remind here the basic inequality

$$||AB||_{HS} \le ||A||_{op}||B||_{HS} \tag{3.20}$$

for any two matrices A and B. Moreover, we have the following basic bound

$$TrG(z) - TrG^{(i)}(z) \le O(\eta^{-1}). \tag{3.21}$$

To see (3.21), we denote the ESD of $W_N^{(i)}$ by $F_N^{(i)}$. Then by Cauchy interlacing property we know

$$\lambda_N(W_N) \le \lambda_{N-1}(W_N^{(i)}) \le \cdots \le \lambda_2(W_N) \le \lambda_1(W_N^{(i)}) \le \lambda_1(W_N),$$

which implies

$$\sup_{x\in\mathbb{R}} |F_N(x) - F_N^{(i)}(x)| \le \frac{1}{N}.$$

Therefore,

$$\begin{aligned} |TrG(z) - TrG^{(i)}(z)| &\le N\int \frac{1}{|\lambda - z|} d(F_N(\lambda) - F_N^{(i)}(\lambda)) \\ &\le \eta^{-1}\int \frac{\eta}{(\lambda-E)^2+\eta^2} d\lambda = \pi\eta^{-1}. \end{aligned} \tag{3.22}$$

By using formula (ii) of Lemma 3.1, Lemma 3.4 and Lemma 3.7, we can get that with ζ-high probability,

$$|\Lambda_o(z)| \le C\max_{i\ne j} |\mathbf{r}_i^*\mathcal{G}^{(ij)}(z)\mathbf{r}_j| \le \max_{i\ne j} \frac{\varphi^{C_\zeta}}{N} ||\Sigma^{1/2}\mathcal{G}^{(ij)}(z)\Sigma^{1/2}||_{HS}$$

$$
\begin{aligned}
&\leq C\frac{\varphi^{C_\zeta}}{N}\max_{i\neq j}||\mathcal{G}^{(ij)}(z)||_{HS} = C\varphi^{C_\zeta}\max_{i\neq j}\sqrt{\frac{\Im Tr\mathcal{G}^{(ij)}(z)}{N^2\eta}}\\
&= C\varphi^{C_\zeta}\max_{i\neq j}\sqrt{\frac{\Im TrG^{(ij)}(z)}{N^2\eta}+O(\frac{1}{N})} = C\varphi^{C_\zeta}\sqrt{\frac{\Im m_N(z)}{N\eta}+O(\frac{1}{N})}.
\end{aligned} \tag{3.23}
$$

Here in the last step we have used the fact that $\eta \sim 1$ and

$$
\left|TrG(z)-TrG^{(ij)}(z)\right| \leq \left|TrG(z)-TrG^{(i)}(z)\right| + \left|TrG^{(i)}(z)-TrG^{(ij)}(z)\right| \leq O(\eta^{-1}).
$$

Now by (3.23) and the fact that $\eta \sim 1$ and thus

$$
\Im m_N(z) \sim 1
$$

with ζ-high probability, we have

$$
|\Lambda_o(z)| \leq C\frac{\varphi^{C_\zeta}}{N}\max_{i\neq j}||\mathcal{G}^{(ij)}(z)||_{HS} \leq O(\varphi^{C_\zeta}\Psi(z)) \tag{3.24}
$$

with ζ-high probability. Similarly, we also have

$$
\frac{\varphi^{C_\zeta}}{N}\max_i ||\mathcal{G}^{(i)}(z)||_{HS} \leq O(\varphi^{C_\zeta}\Psi(z))
$$

with ζ-high probability.

For Λ_d, we start from the observation

$$
\max_i |G_{ii}-m_N| \leq \max_{i\neq j}|G_{ii}-G_{jj}|, \tag{3.25}
$$

while

$$
\begin{aligned}
|G_{ii}-G_{jj}| &= \left|\frac{1}{-z-z\mathbf{r}_i^*\mathcal{G}^{(i)}\mathbf{r}_i} - \frac{1}{-z-z\mathbf{r}_j^*\mathcal{G}^{(j)}\mathbf{r}_j}\right|\\
&\leq |G_{ii}G_{jj}|\left(|T_i-T_j| + \frac{1}{N}|Tr\mathcal{G}^{(i)}\Sigma - Tr\mathcal{G}^{(j)}\Sigma|\right),
\end{aligned}
$$

where T_i is defined in (3.7). Note that by using Lemma 3.4 again, we have

$$
|T_i(z)| \leq \varphi^{C_\zeta}\frac{1}{N}||\mathcal{G}^{(i)}(z)\Sigma||_{HS} \leq C\varphi^{C_\zeta}\frac{1}{N}||\mathcal{G}^{(i)}(z)||_{HS} = O(\varphi^{C_\zeta}\Psi(z)) \tag{3.26}
$$

with ζ-high probability. Here the last step can be obtained by a calculation similar to that for (3.24). Moreover, we have

$$
\frac{1}{N}|Tr\mathcal{G}^{(i)}(z)\Sigma - Tr\mathcal{G}^{(j)}(z)\Sigma| \leq |U_i(z)| + |U_j(z)| \tag{3.27}
$$

Therefore, we come to estimate $U_i(z)$ below. Now using the Sherman-Morrison formula

$$
(A+\mathbf{r}_i\mathbf{r}_j^*)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{r}_i\mathbf{r}_j^*A^{-1}}{1+\mathbf{r}_j^*A^{-1}\mathbf{r}_i} \tag{3.28}
$$

for any $M\times M$ invertible matrix A, we have

$$
|U_i(z)| = \frac{1}{N}|Tr\left(\mathcal{G}^{(i)}(z)-\mathcal{G}(z)\right)\Sigma| = \frac{1}{N}\left|\frac{\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma\mathcal{G}^{(i)}\mathbf{r}_i}{1+\mathbf{r}_i^*\mathcal{G}^{(i)}\mathbf{r}_i}\right| = \frac{1}{N}|zG_{ii}(z)|\left|\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma\mathcal{G}^{(i)}\mathbf{r}_i\right|
$$

$$\leq \quad C\frac{1}{N}\left(\left|\frac{1}{N}Tr\mathcal{G}^{(i)}\Sigma\mathcal{G}^{(i)}\Sigma\right| + \frac{\varphi^{C_\zeta}}{N}||\mathcal{G}^{(i)}\Sigma\mathcal{G}^{(i)}\Sigma||_{HS}\right)$$

with ζ-high probability, where the last step above follows from Lemma 3.4.

Observe that

$$\left|Tr\mathcal{G}^{(i)}\Sigma\mathcal{G}^{(i)}\Sigma\right| \leq Tr\mathcal{G}^{(i)}\Sigma^2(\mathcal{G}^{(i)})^* \leq C||\mathcal{G}^{(i)}||_{HS}^2.$$

and

$$||\mathcal{G}^{(i)}\Sigma\mathcal{G}^{(i)}\Sigma||_{HS} \leq C||(\mathcal{G}^{(i)})^2||_{HS} \leq C||\mathcal{G}^{(i)}||_{HS}^2.$$

Therefore we have

$$|U_i(z)| \leq C\varphi^{C_\zeta}\frac{1}{N^2}||\mathcal{G}^{(i)}||_{HS}^2 = \varphi^{C_\zeta}\frac{\Im Tr\mathcal{G}^{(i)}}{N^2\eta} = O(\varphi^{C_\zeta}\Psi^2(z)) \tag{3.29}$$

with ζ-high probability. Thus (3.25)-(3.29) imply that

$$\max_i |G_{ii}(z) - m_N(z)| = O(\varphi^{C_\zeta}\Psi(z))$$

with ζ-high probability.

Hence, it remains to bound

$$R_i(z), \quad v_{1i}(z), \quad i = 1, \cdots, N, \quad \text{and} \quad v_k(z), \quad k = 2, 3, 4.$$

To bound these quantities, we recall (3.18). For $v_{1i}(z)$, we use Lemma 3.4 again. Thus we have

$$\begin{aligned}
\max_i |v_{1i}(z)| &\leq \max_i |\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma + I)^{-1}\mathbf{r}_i - \frac{1}{N}Tr(m_N^{(i)}(z)\Sigma + I)^{-1}\Sigma\mathcal{G}^{(i)}(z)\Sigma| \\
&\leq \max_i \frac{\varphi^{C_\zeta}}{N}||(m_N^{(i)}(z)\Sigma + I)^{-1}\Sigma\mathcal{G}^{(i)}(z)\Sigma||_{HS} \\
&\leq C\max_i \frac{\varphi^{C_\zeta}}{N}||\mathcal{G}^{(i)}||_{HS} = O(\varphi^{C_\zeta}\Psi(z))
\end{aligned}$$

with ζ-high probability, where the last inequality above follows from (3.18). Similarly we can get that

$$|R_i(z)| = O(\varphi^{C_\zeta}\Psi(z))$$

with ζ-high probability.

For $v_2(z)$, we have

$$\begin{aligned}
|v_2(z)| &\leq C\max_i \left|\mathbf{r}_i^*\mathcal{G}^{(i)}(z)\Sigma\left[(m_N(z)\Sigma + I)^{-1} - (m_N^{(i)}(z)\Sigma + I)^{-1}\right]\mathbf{r}_i\right| \\
&= \max_i \left|\mathbf{r}_i^*\mathcal{G}^{(i)}(z)\Sigma(m_N(z)\Sigma + I)^{-1}(m_N(z) - m_N^{(i)}(z))\Sigma(m_N^{(i)}(z)\Sigma + I)^{-1}\mathbf{r}_i\right| \\
&\leq \max_i |m_N(z) - m_N^{(i)}(z)|\Big(\left|Tr\mathcal{G}^{(i)}(z)\Sigma(m_N(z)\Sigma + I)^{-1}\Sigma(m_N^{(i)}(z)\Sigma + I)^{-1}\Sigma\right| \\
&\quad +\varphi^{C_\zeta}||\mathcal{G}^{(i)}(z)\Sigma(m_N(z)\Sigma + I)^{-1}\Sigma(m_N^{(i)}(z)\Sigma + I)^{-1}\Sigma||_{HS}\Big) \\
&\leq C\max_i |m_N(z) - m_N^{(i)}(z)|\left(\frac{1}{N}Tr|\mathcal{G}^{(i)}(z)| + \frac{\varphi^{C_\zeta}}{N}||\mathcal{G}^{(i)}(z)||_{HS}\right) \\
&\leq C|\Lambda_o(z)|^2 = O(\varphi^{C_\zeta}\Psi^2(z)). \qquad (3.30)
\end{aligned}$$

with ζ-high probability. Here in the last inequality above we have used the fact that

$$|m_N(z) - m_N^{(i)}(z)| = O((\Lambda_o(z))^2)$$

which is implied by (3.15) and (iii) of Lemma 3.1.

For $v_3(z)$, we can use the same approach as that we have used to bound $U_i(z)$ in (3.29). Actually by using (3.28) and (3.15) again, we have

$$\begin{aligned}
|v_3(z)| &\leq C\max_i \left|\frac{1}{N}Tr\left(\mathcal{G}^{(i)}(z) - \mathcal{G}(z)\right)\Sigma(m_N^{(i)}(z)\Sigma + I)^{-1}\Sigma\right| \\
&= C\max_i \frac{1}{N}|zG_{ii}(z)|\left|\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma + I)^{-1}\Sigma\mathcal{G}^{(i)}\mathbf{r}_i\right| \\
&\leq C\varphi^{C_\zeta}\max_i \frac{1}{N^2}||\mathcal{G}^{(i)}||_{HS}^2 = \varphi^{C_\zeta}\max_i \frac{\Im Tr\mathcal{G}^{(i)}}{N^2\eta} = O(\varphi^{C_\zeta}\Psi^2(z)).
\end{aligned}$$

Moreover, the estimate of $v_4(z)$ is similar to that of $v_2(z)$. Actually, we have

$$\begin{aligned}
|v_4(z)| &\leq \max_i \left|\frac{1}{N}Tr(m_N^{(i)}(z)\Sigma + I)^{-1}\Sigma\mathcal{G}(z)\Sigma - \frac{1}{N}Tr(m_N(z)\Sigma + I)^{-1}\Sigma\mathcal{G}(z)\Sigma\right| \\
&\leq C\max_i |m_N(z) - m_N^{(i)}(z)| = O(\Lambda_o^2) = O(\varphi^{C_\zeta}\Psi^2(z))
\end{aligned}$$

with ζ-high probability. Thus we conclude the proof of Lemma 3.5. □

With the aid of Lemma 3.5, we can prove Lemma 3.6 below.

Proof of Lemma 3.6. Similar to the discussion in the proof of Lemma 3.5, it suffices to estimate $\Lambda(z)$ for a fixed z. At first, we pursue the idea in [27] to introduce the function

$$\begin{aligned}
\mathcal{D}(u)(z) &= \left((u(z))^{-1} - d_N^{-1}\int \frac{t}{tu(z)+1}dH_N(t)\right) \\
&\quad - \left((m_0(z))^{-1} - d_N^{-1}\int \frac{t}{tm_0(z)+1}dH_N(t)\right).
\end{aligned}$$

Note that by the fact

$$\sum_{i=1}^N (G_{ii}(z) - m_N(z)) = 0,$$

it is not difficult to see that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \left((G_{ii}(z))^{-1} - (m_N(z))^{-1}\right) \\
&= \frac{1}{N}\sum_{i=1}^N \frac{m_N(z) - G_{ii}(z)}{m_N^2(z)} + O\left(\frac{1}{N}\sum_{i=1}^N \frac{(G_{ii}(z) - m_N(z))^2}{G_{ii}(z)m_N^2(z)}\right) \\
&= O\left(\frac{1}{N}\sum_{i=1}^N \frac{(G_{ii}(z) - m_N(z))^2}{G_{ii}(z)m_N^2(z)}\right) = O\left(\varphi^{C_\zeta}\Psi^2(z)\right) \qquad (3.31)
\end{aligned}$$

with ζ-high probability by using Lemma 3.5 and Lemma 3.7. By the fact that

$$(G_{ii}(z))^{-1} = -z + d_N^{-1}\int \frac{t}{tm_N(z)+1}dH_N(t) - Y_i,$$

(3.31) and (1.6) we have

$$|\mathcal{D}(m_N)(z)| \le |[Y]| + \varphi^{C_\zeta}\Psi^2(z)$$

with ζ-high probability, where

$$[Y] = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

Therefore by the definition of $\mathcal{D}(u)(z)$ and the bounds for $|Y_i|$, we know that in the case of $\eta \sim 1$, there exists

$$\left|((m_N(z))^{-1} - (m_0(z))^{-1}) + d_N^{-1}\int \left(\frac{t}{tm_0(z)+1} - \frac{t}{tm_N(z)+1}\right) dH_N(t)\right| \le \varphi^{C_\zeta}\frac{1}{(N\eta)^{1/2}}$$

with ζ-high probability. Taking the fact that $m_0(z), m_N(z) \sim 1$ into account we obtain

$$(m_0(z) - m_N(z))\left[1 - d_N^{-1}\int \frac{t^2 m_N(z)m_0(z)}{(tm_N(z)+1)(tm_0(z)+1)}dH_N(t)\right] = \delta_0(z) \tag{3.32}$$

with some function $\delta_0(z)$ satisfying

$$|\delta_0(z)| \le \varphi^{C_\zeta}\frac{1}{(N\eta)^{1/2}} \tag{3.33}$$

with ζ-high probability. Now we need to estimate

$$1 - d_N^{-1}\int \frac{t^2 m_N(z)m_0(z)}{(tm_N(z)+1)(tm_0(z)+1)}dH_N(t)$$

when $\eta \sim 1$.

Note that by Cauchy-Schwarz inequality, we have

$$\begin{aligned} &\left|d_N^{-1}\int \frac{t^2 m_N(z)m_0(z)dH_N(t)}{(tm_N(z)+1)(tm_0(z)+1)}\right| \\ \le\ &\left(\int \frac{d_N^{-1}t^2|m_N(z)|^2}{|tm_N(z)+1|^2}dH_N(t)\right)^{1/2}\cdot\left(\int \frac{d_N^{-1}t^2|m_0(z)|^2}{|tm_0(z)+1|^2}dH_N(t)\right)^{1/2}. \end{aligned}$$

Now by (1.6)

$$1 + zm_0(z) - d_N^{-1}\int \frac{tm_0(z)}{tm_0(z)+1}dH_N(t) = 0,$$

we have

$$\overline{m_0(z)} + z|m_0(z)|^2 - d_N^{-1}|m_0(z)|^2\int \frac{t}{tm_0(z)+1}dH_N(t) = 0.$$

Taking the imaginary part of the above equation we obtain

$$-\Im m_0(z) + \eta|m_0(z)|^2 + d_N^{-1}|m_0(z)|^2\int \frac{t^2\Im m_0(z)}{|tm_0(z)+1|^2}dH_N(t) = 0,$$

which implies

$$0 \le d_N^{-1}\int \frac{t^2|m_0(z)|^2}{|tm_0(z)+1|^2}dH_N(t) = 1 - \frac{|m_0(z)|^2}{\Im m_0(z)}\eta < 1 - \delta \tag{3.34}$$

for some positive constant δ. Here we have used the fact that $\eta \sim 1$ and (i) of Lemma 2.3. Similarly, by the facts

$$(m_N(z))^{-1} - d_N^{-1} \int \frac{t}{tm_0(z)+1} dH_N(t) + z = \delta_0$$

and $m_N(z) \sim 1$ with ζ-high probability when $\eta \sim 1$, we can obtain

$$0 \le d_N^{-1} \int \frac{t^2 |m_N(z)|^2}{|tm_N(z)+1|^2} dH_N(t) = 1 - \frac{|m_N(z)|^2}{\Im m_N(z)}(\eta - \Im \delta_0'(z)) < 1$$

with ζ-high probability, where $\delta_0'(z)$ is some function such that $|\delta_0'(z)| \le \varphi^{C_\zeta}/(N\eta)^{1/2}$. Here we have used the fact that $\eta \sim 1$ and the bound (3.33).

Therefore we have

$$\left| 1 - d_N^{-1} \int \frac{t^2 m_N(z) m_0(z)}{(tm_N(z)+1)(tm_0(z)+1)} dH_N(t) \right| \ge c_0 \tag{3.35}$$

for some positive constant c_0. Hence, by (3.32), (3.33) and (3.35) we have

$$\Lambda(z) \le \varphi^{C_\zeta} \frac{1}{(N\eta)^{1/2}}.$$

Thus we complete the proof. □

Now ,we consider to extend the results to the case of $\eta \ll 1$. To this end, we will provide the desired bounds ((3.16) and (3.17)) under the condition that the event $\Xi^c(z)$ happens at first, then we prove that the event $\Xi^c(z)$ holds with ζ-high probability. Such a strategy is also parallel to that of the null case in [27]. Similar to Lemma 3.5, we now need the following lemma for the first step.

LEMMA 3.8. *Under Condition 1.1 and (3.4), for any $\zeta > 0$, there exists a positive constant C_ζ such that*

$$\bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \Gamma(z, \varphi^{C_\zeta})$$

with ζ-high probability.

To prove Lemma 3.8, we will need the following lemma.

LEMMA 3.9. *For $z \in S_r(\tilde{c}, 5C_\zeta)$, when the event $\Xi^c(z)$ happens , for any $G_{ij}^{(\mathbb{T})}(z)$ with $|\mathbb{T}| = O(1)$ and $i, j \notin \mathbb{T}$ we have*

$$\max_{i \notin \mathbb{T}} |G_{ii}^{(\mathbb{T})} - G_{ii}| = O(\Lambda_o^2), \quad G_{ii}^{(\mathbb{T})} \sim 1, \quad \Lambda_o^{(\mathbb{T})} := \max_{i \ne j, i, j \notin \mathbb{T}} |G_{ij}^{(\mathbb{T})}(z)| \le C\Lambda_o, \tag{3.36}$$

holds with ζ-high probability.

PROOF. Note that when the event $\Xi^c(z)$ happens, we have

$$G_{ii} \sim 1, \quad \Lambda_o(z) \le (\log N)^{-1}.$$

Now by (iii) of Lemma 3.1 and the induction method, we can easily conclude the proof. □

Now we come to show Lemma 3.8

Proof of Lemma 3.8. At first, we use the discussion in the proof of Lemma 3.5 again to claim that it suffices to prove the result for any fixed $z \in S_r(\tilde{c}, 5C_\zeta)$.

By the definition in (3.14), it suffices to show that (3.16) and (3.17) hold with ζ-high probability in $\Xi^c(z)$. That means, we need to go back to the proof of Lemma 3.5. But this time we have the condition that $\Xi^c(z)$ happens instead of $\eta \sim 1$. Note that by definition, in $\Xi^c(z)$ we have

$$\Lambda(z) \le \Lambda_d(z) \le (\log N)^{-1}.$$

Using (i) of Lemma 2.3, we have

$$m_N(z) \sim 1, \quad G_{ii}(z) \sim 1, \quad \text{in } \Xi^c(z). \tag{3.37}$$

Note that in the proof of Lemma 3.5, every term except $v_2(z)$ and $v_4(z)$ can finally be bounded by some quantity in terms of $\max_{i \ne j} ||\mathcal{G}^{(ij)}||_{HS}$ or $\max_i ||\mathcal{G}^{(i)}||_{HS}$. Therefore, at first we will bound $||\mathcal{G}^{(ij)}||_{HS}$ and $||\mathcal{G}^{(i)}||_{HS}$. Note that

$$\begin{aligned} ||\mathcal{G}^{(ij)}||_{HS} &= \sqrt{\frac{\Im Tr\mathcal{G}^{(ij)}}{\eta}} = \sqrt{\frac{\Im Tr G^{(ij)}}{\eta} + O(N)} \\ &= \sqrt{\frac{\Im Tr G}{\eta} + O(\frac{N\Lambda_o^2}{\eta}) + O(N)}, \end{aligned} \tag{3.38}$$

where the last step follows from (3.36). Now similar to (3.24), with the aid of (3.37), we see in $\Xi^c(z)$, there exists

$$\begin{aligned} \Lambda_o(z) &\le C\frac{\varphi^{C_\zeta}}{N} \max_{i\ne j} ||\mathcal{G}^{(ij)}(z)||_{HS} \le C\varphi^{C_\zeta}\sqrt{\frac{\Im m_N(z)}{N\eta} + O(\frac{\Lambda_o^2}{N\eta}) + O(\frac{1}{N})} \\ &= C\varphi^{C_\zeta}\sqrt{\frac{\Im m_N(z)}{N\eta} + O(\frac{1}{N})} + o(\Lambda_o), \end{aligned}$$

where we have used the fact that $z \in S_r(\tilde{c}, 5C_\zeta)$. Thus by (iii) of Lemma 2.3 we have

$$\Lambda_o(z) \le \varphi^{C_\zeta}\Psi(z) \tag{3.39}$$

with ζ-high probability in $\Xi^c(z)$. Inserting (3.39) into (3.38), we have

$$\frac{1}{N}||\mathcal{G}^{(ij)}(z)||_{HS} \le \Psi(z) \tag{3.40}$$

with ζ-high probability in $\Xi^c(z)$. Analogously, we also have

$$\frac{1}{N}||\mathcal{G}^{(i)}(z)||_{HS} \le \Psi(z) \tag{3.41}$$

with ζ-high probability in $\Xi^c(z)$

With the aid of (3.37), (3.40) and (3.41) it is not difficult to see (by using Lemma 3.9) the estimates between (3.24) and (3.29) are still valid in $\Xi^c(z)$ for all $z \in S_r(\tilde{c}, 5C_\zeta)$.

Therefore, it remains to estimate R_i, v_{1i} and v_2, v_3, v_4. Actually, we need the following lemma.

Lemma 3.10. *Under Condition 1.1, for any $\zeta > 0$, there exists a constant $C_\zeta > 0$, when $z \in S_r(\tilde{c}, 5C_\zeta)$, we have with ζ-high probability,*

$$||(m_N^{(i)}(z)\Sigma + I)^{-1}||_{op}, ||(m_N(z)\Sigma + I)^{-1}||_{op}, ||(m_0(z)\Sigma + I)^{-1}||_{op} \le C_\zeta \tag{3.42}$$

in $\Xi^c(z)$. Moreover, we have with ζ-high probability,

$$\frac{1}{N} Tr|\mathcal{G}^{(i)}(z)| \le (\log N)^{O(1)}, \text{for all } z \in S_r(\tilde{c}, 5C_\zeta). \tag{3.43}$$

Note that by the estimates for R_i, v_{1i} and v_2, v_3, v_4 in the proof of Lemma 3.5, we can easily see that once Lemma 3.10 holds, Lemma 3.8 follows. Actually, with the aid of Lemma 3.10, we can easily check that

$$|v_2|, |v_3|, |v_4| \le O(\varphi^{C_\zeta}\Psi^2), \quad |R_i|, |v_{1i}| \le O(\varphi^{C_\zeta}\Psi), \quad k = 1, \ldots, N \tag{3.44}$$

with ζ-high probability in $\Xi^c(z)$.

Hence, it suffices to prove Lemma 3.10 below. To this end, we need the following crude bound on the counting function of eigenvalues. Hereafter, we will use $N_I(A)$ to denote the number of the eigenvalues of an Hermitian matrix A in the interval I.

Lemma 3.11. *Under Condition 1.1, for any $\zeta > 0$ there exists some constant $C_\zeta > 0$ such that for any interval $I \subset [C_l, \infty)$ with length $|I| \ge \varphi^{5C_\zeta}/N$, we have*

$$N_I(\mathcal{W}) \le C_\zeta N|I| \tag{3.45}$$

with ζ-high probability.

Proof. Let $\eta \ge \varphi^{5C_\zeta}/N$. Note that for any interval $I = [E - \eta/2, E + \eta/2] \subset [C_l, \infty]$ (thus $E - \eta/2 \ge C_l$) we have the elementary inequality

$$N_I(\mathcal{W}) \le CN\eta \Im m_N(z), \quad z = E + i\eta,$$

thus

$$N_I(\mathcal{W}) \le C\eta \sum_{i=1}^{N} \Im G_{ii}(z).$$

Now we assume $N_I(\mathcal{W}) > C_\zeta N\eta$ for any large constant C_ζ to get a contradiction. It suffices to show for any $\zeta > 0$, there exists a positive constant C'_ζ such that

$$|\Im G_{ii}(z)| \le C'_\zeta \tag{3.46}$$

with ζ-high probability if $N_I(\mathcal{W}) > C_\zeta N\eta$. To verify (3.46) under the assumption of $N_I(\mathcal{W}) > C_\zeta N\eta$, we rewrite (i) of Lemma 3.1 as

$$G_{ii} = -\frac{1}{z + z\sum_{k=1}^{M} \frac{1}{\lambda_k(\mathcal{W}^{(i)}) - z} |\langle \mathbf{u}_k(\mathcal{W}^{(i)}), \mathbf{r}_i\rangle|^2},$$

where $\lambda_1(\mathcal{W}^{(i)}) \ge \lambda_2(\mathcal{W}^{(i)}) \ge \ldots \ge \lambda_M(\mathcal{W}^{(i)})$ are the eigenvalues of $\mathcal{W}^{(i)}$ and $\mathbf{u}_k(\mathcal{W}^{(i)}), k = 1\ldots, M$ are their corresponding unit eigenvectors. Then we have

$$\begin{aligned} \Im G_{ii}(z) &\le \frac{1}{\eta + \eta\sum_{k=1}^{M} \frac{\lambda_k(\mathcal{W}^{(i)})}{(\lambda_k(\mathcal{W}^{(i)}) - E)^2 + \eta^2} |\langle \mathbf{u}_k(\mathcal{W}^{(i)}), \mathbf{r}_i\rangle|^2} \\ &\le \frac{C\eta}{C_l \sum_{k:|\lambda_k(\mathcal{W}^{(i)}) - E| \le \eta/2} |\langle \mathbf{u}_k(\mathcal{W}^{(i)}), \mathbf{r}_i\rangle|^2} \end{aligned} \tag{3.47}$$

with ζ-high probability. Now by Cauchy interlacing property again we know for any interval I there exists

$$|N_I(\mathcal{W}) - N_I(\mathcal{W}^{(i)})| \le 1. \tag{3.48}$$

Thus by assumption, we have

$$|N_I(\mathcal{W}_N^{(i)})| = \sharp\{k : |\lambda_k(\mathcal{W}^{(i)}) - E| \le \eta/2\} \ge C_\zeta N\eta$$

with some sufficiently large C_ζ. Now we set the projection matrix

$$P_z := \sum_{k:|\lambda_k(\mathcal{W}^{(i)})-E|\le\eta/2} \mathbf{u}_k(\mathcal{W}^{(i)})(\mathbf{u}_k(\mathcal{W}^{(i)}))^*.$$

By assumption we have

$$TrP_z = TrP_z^2 \ge C_\zeta N\eta.$$

Thus by Lemma 3.4, there exists

$$\sum_{k:|\lambda_k(\mathcal{W}^{(i)})-E|\le\eta/2} |\langle \mathbf{u}_k(\mathcal{W}^{(i)}), \mathbf{r}_i\rangle|^2 = X_i^*\Sigma^{1/2}P_z\Sigma^{1/2}X_i$$
$$= \frac{1}{N}Tr\Sigma^{1/2}P_z\Sigma^{1/2} + O\left(\frac{\varphi^{C\zeta}}{N}||\Sigma^{1/2}P_z\Sigma^{1/2}||_{HS}\right) \ge C''_\zeta\eta$$

with ζ-high probability for some positive constant C''_ζ. Thus by (3.47) we see that (3.46) holds with ζ-high probability. Therefore we conclude the proof. □

Now we come to prove Lemma 3.10.

Proof of Lemma 3.10. At first, we will show (3.43). By definition we have

$$\frac{1}{N}Tr|\mathcal{G}^{(i)}(z)| = \frac{1}{N}\sum_{k=1}^{M}\frac{1}{|\lambda_k(\mathcal{W}^{(i)})-z|}.$$

Note that when $z = E+i\eta \in S_r(\tilde{c}, 5C_\zeta)$ one has $\eta \ge \varphi^{5C_\zeta}/N$. Now we split $\mathbb{R}$ into

$$\mathbb{R} = (-\infty, \lambda_r - 2\tilde{c}) \cup \left(\bigcup_{k=1}^{K_n} I_k\right) \cup (C_r, \infty) := I_0 \cup \left(\bigcup_{k=1}^{K_n} I_k\right) \cup I_{K_n+1},$$

where $K_n = O(\eta^{-1})$ and I_k are non-intersecting intervals with $|I_k| = \eta, k = 1, \ldots, K_n$. Specifically,

$$I_1 := [\lambda_r - 2\tilde{c}, \lambda_r - 2\tilde{c} + \eta).$$

Now we write

$$\frac{1}{N}Tr|\mathcal{G}^{(i)}(z)| = \frac{1}{N}\sum_{k=0}^{K_n+1}\sum_{l:\lambda_l(\mathcal{W}^{(i)})\in I_k}\frac{1}{\sqrt{(\lambda_l(\mathcal{W}^{(i)})-E)^2+\eta^2}}.$$

Then for $z = E+i\eta \in S_r(\tilde{c}, 5C_\zeta)$, by invoking Lemma 3.11 and (3.48) we have

$$\frac{1}{N}Tr|\mathcal{G}^{(i)}(z)| \le \log^{O(1)} N$$

with ζ-high probability.

Now we start to show (3.42). Note that in $\Xi^c(z)$, we have

$$|m_N(z) - m_0(z)|, |m_N^{(i)}(z) - m_0(z)| \le (\log N)^{-1}.$$

Therefore, it suffices to show that

$$||(m_0(z)\Sigma + I)^{-1}||_{op} = O(1), \quad \text{for } z \in S_r(\tilde{c}, 5C_\zeta), \tag{3.49}$$

which follows from (iv) of Lemma 2.3 immediately. □

Therefore, we conclude the proof of Lemma 3.8. □

Moreover, by definitions we also have the following lemma.

LEMMA 3.12. *Let K be some positive number such that $1 \leq K \ll (N\eta)^{1/2}$. Then for $z \in S_r(\tilde{c}, 5C_\zeta)$, in $\Gamma(z,K)$ we have*

$$|\mathcal{D}(m_N)(z)| \leq |[Y]| + O(K^2\Psi^2(z)) + \infty \mathbf{1}_{\Xi(z)}. \tag{3.50}$$

PROOF. Note that by the assumption on K, we have

$$K\Psi \ll 1, \quad \text{in} \quad \Xi^c(z).$$

Hence, by the fact that $G_{ii}(z), m_N(z) \sim 1$ in $\Xi^c(z)$ it is not difficult to see

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}(G_{ii}(z))^{-1} &= (m_N(z))^{-1} + O(\max_i |G_{ii}(z) - m_N(z)|^2) \\ &= (m_N(z))^{-1} + O(K^2\Psi^2(z)), \quad \text{in } \Omega^c(z,K) \cap \Xi^c(z)\end{aligned}$$

from (3.31). Moreover, we recall the identity

$$(G_{ii}(z))^{-1} = -z + d_N^{-1}\int \frac{t}{tm_N(z)+1} dH_N(t) - Y_i,$$

which implies

$$(m_N(z))^{-1} = -z + d_N^{-1}\int \frac{t}{tm_N(z)+1} dH_N(t) - [Y] + O(K^2\Psi^2(z)), \quad \text{in } \Omega^c(z,K) \cap \Xi^c(z).$$

Then by the definition of $\mathcal{D}(u)(z)$ we obtain

$$|\mathcal{D}(m_N)(z)| \leq |[Y]| + O(K^2\Psi^2(z)), \quad \text{in } \Omega^c(z,K) \cap \Xi^c(z),$$

which implies (3.50) in $\Gamma(z,K)$. □

Now we are on the stage to extend the estimate on $\Lambda(z)$ to the case of $\eta \ll 1$. Actually, we have the following crucial Lemma which is an extension of Lemma 6.12 of [27] to our non-null case.

LEMMA 3.13. *Let $K = \varphi^{O(1)}$ and $L = O(1)$ be two positive numbers satisfying*

$$\varphi^L \geq K^2(\log N)^4.$$

And let $\mathbf{S}$ be an event satisfying

$$\mathbf{S} \subset \bigcap_{z\in S_r(\tilde{c},L)} \Gamma(z,K) \cap \bigcap_{z\in S_r(\tilde{c},L),\eta=1} \Xi^c(z).$$

Assume that in $\mathbf{S}$ one has

$$|\mathcal{D}(m_N)(z)| \leq \delta(z) + \infty\mathbf{1}_{\Xi(z)}, \quad \textit{for all } z \in S_r(\tilde{c},L),$$

where $\delta : \mathbb{C} \to \mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$ is a continuous function. Moreover, $\delta(z)$ is decreasing in η and $|\delta(z)| \leq (\log N)^{-8}$. Then there exists some positive constant C such that

$$\Lambda(z) \leq C(\log N)\frac{\delta(z)}{\sqrt{\kappa+\eta+\delta(z)}}, \quad \textit{for all } z \in S_r(\tilde{c},L) \tag{3.51}$$

holds in $\mathbf{S}$ and

$$\mathbf{S} \subset \bigcap_{z\in S_r(\tilde{c},L)} \Xi^c(z). \tag{3.52}$$

PROOF. Note that we need to prove (3.51) and (3.52). By definition of $\Xi(z)$ it suffices to prove that (3.51) holds when

$$\Lambda_o(z) + \Lambda_d(z) \le (\log N)^{-1} \tag{3.53}$$

and then prove (3.53) holds for all $z \in S_r(\tilde{c}, L)$ when the event $\mathbf{S}$ occurs. To this end, we define the set of η

$$I_E := \{\eta : \Lambda_o(E + i\hat{\eta}) + \Lambda_d(E + i\hat{\eta}) \le (\log N)^{-1}, \forall \hat{\eta} \ge \eta, E + i\hat{\eta} \in S_r(\tilde{c}, L)\}$$

At first, we come to show that (3.51) holds for $z = E + i\eta$ with $\eta \in I_E$. By assumption, we have

$$|\mathcal{D}(m_N)(E + i\eta)| \le \delta(E + i\eta), \quad \forall \eta \in I_E.$$

Note that when $\eta \in I_E$, we have

$$m_N(z) = m_0(z) + O((\log N)^{-1}) \sim 1. \tag{3.54}$$

Thus similar to (3.32), by the definition of $\mathcal{D}(u)(z)$ here we also have

$$(m_0(z) - m_N(z)) \left[1 - d_N^{-1} \int \frac{t^2 m_N(z) m_0(z)}{(tm_N(z) + 1)(tm_0(z) + 1)} dH_N(t)\right] = \delta_0(z), \tag{3.55}$$

with

$$|\delta_0(z)| \le O(\delta(z)), \quad \eta \in I_E.$$

Similarly, in order to obtain a bound for Λ, we need to derive an estimate for the quantity

$$\left|1 - d_N^{-1} \int \frac{t^2 m_N(z) m_0(z)}{(tm_N(z) + 1)(tm_0(z) + 1)} dH_N(t)\right|.$$

However, unlike (3.35) whose estimation is based on the fact of $\eta \sim 1$, we need to do it for general η this time. Thus more delicate calculation should be taken.

At first, by the discussion towards (3.32), we see (3.51) holds naturally when $\eta \sim 1$. Therefore, in the sequel, we will assume $\eta \le \hat{\eta}$ for some sufficiently small constant $\hat{\eta}$. Now we can write (3.55) as

$$\begin{aligned} &(m_0(z) - m_N(z)) \\ &\times \left[1 - d_N^{-1} \int \frac{t^2 m_0^2(z) dH_N(t)}{(tm_0(z) + 1)^2} + d_N^{-1}(m_0(z) - m_N(z)) \int \frac{t^2 m_0(z) dH_N(t)}{(tm_N(z) + 1)(tm_0(z) + 1)^2}\right] \\ &= \delta_0(z). \end{aligned} \tag{3.56}$$

By (i) and (iv) of Lemma 2.3 and (3.54), we have

$$\left|\int \frac{t^2 m_0(z)}{(tm_N(z) + 1)(tm_0(z) + 1)^2} dH_N(t)\right| \le C, \quad \text{if } z \in S_r(\tilde{c}, L). \tag{3.57}$$

for some positive constant C. Now we come to show that when $\eta \le \hat{\eta}$ and $\tilde{c}$ is chosen to be sufficiently small, we also have

$$\left|\int \frac{t^2 m_0(z)}{(tm_N(z) + 1)(tm_0(z) + 1)^2} dH_N(t)\right| \ge C^{-1}, \quad \text{if } z \in S_r(\tilde{c}, L) \tag{3.58}$$

for some sufficiently large positive constant C. Note that by using (i) and (iv) of Lemma 2.3 and (3.54) again, it is easy to see

$$\left|\int \frac{t^2}{(tm_N(z) + 1)(tm_0(z) + 1)^2} dH_N(t) - \int \frac{t^2}{(tm_0(z) + 1)^3} dH_N(t)\right| = O((\log N)^{-1})$$

when $\eta \in I_E$. Thus it suffices to show that when $\eta \leq \hat{\eta}$ and $\tilde{c}$ is chosen to be sufficiently small, there exists

$$|\int \frac{t^2}{(tm_0(z)+1)^3} dH_N(t)| \geq C^{-1}, \quad \text{if } z \in S_r(\tilde{c}, L).$$

For convenience we set

$$\mathcal{J}(t) := tm_0(z) + 1.$$

Note that by (iv) of Lemma 2.3, we know both $\Re\mathcal{J}(t)$ and $\Im\mathcal{J}(t)$ are positive when $z \in S_r(\tilde{c}, L)$. Let

$$\theta(t) := \arg(\mathcal{J}(t)) \in (0, \pi/2).$$

It is not difficult to see from the proof of (iv) of Lemma 2.3 that if $\tilde{c}$ is chosen to be sufficiently small, we can guarantee that the parameter c_0 therein is much larger than $\tilde{c}$. Then by (ii) of Lemma 2.3, when $\eta \geq \hat{\eta}$ we also have

$$\Im m_0(z) \leq C\sqrt{\tilde{c} + \hat{\eta}}.$$

Therefore, when $\hat{\eta}$ and $\tilde{c}$ are sufficiently small, we have that $\Re\mathcal{J}(t)$ is much larger than $\Im\mathcal{J}(t)$ such that $\theta(t) \leq \pi/18$(say). That means we have

$$0 < (\theta(t))^3 \leq \pi/6, \quad \text{thus} \quad \Re(\mathcal{J}(t))^3 \geq \frac{1}{2}|\mathcal{J}(t)|^3$$

when $z \in S_r(\tilde{c}, L)$. Then it is easy to see

$$\left| \int \frac{t^2}{(tm_0(z)+1)^3} dH_N(t) \right| \geq \Re \int \frac{t^2}{(tm_0(z)+1)^3} dH_N(t) \geq C^{-1}$$

for some positive constant C. Then (3.57) and (3.58) together imply that

$$\int \frac{t^2 m_0(z)}{(tm_N(z)+1)(tm_0(z)+1)^2} dH_N(t) \sim 1, \quad z \in S_r(\tilde{c}, L), \quad \eta \in I_E, \quad \eta \leq \hat{\eta}$$

when $\hat{\eta}$ and $\tilde{c}$ are sufficiently small. Therefore, by (3.56) we have

$$a(z)(m_0(z) - m_N(z))^2 + (1 - b(z))(m_0(z) - m_N(z)) = \delta_0(z), \tag{3.59}$$

where

$$a(z) \sim 1, \quad b(z) = d_N^{-1} \int \frac{t^2 m_0^2(z)}{(tm_0(z)+1)^2} dH_N(t).$$

Now we need to provide an upper bound for

$$|1 - b(z)|.$$

Recall the formula in (3.34) and set

$$c(z) := d_N^{-1} \int \frac{t^2 |m_0(z)|^2}{|tm_0(z)+1|^2} dH_N(t) = 1 - \frac{|m_0(z)|^2}{\Im m_0(z)} \eta < 1. \tag{3.60}$$

Note that since $1 > c(z) > |b(z)|$, it is obvious that

$$|1 - b(z)| \geq \max\{1 - c(z), |c(z) - b(z)|\}. \tag{3.61}$$

Thus it suffices to estimate the quantity

$$|c(z) - b(z)|.$$

To this end, we use an idea in [11] which is provided for studying the stability of the Stieltjes transform of $\mathcal{W}_N$ type matrices in the bulk case. Actually, we can extend the discussion in [11] to the edge case for our model as well. For ease of presentation, we set

$$\mathcal{F}(t) = \frac{tm_0(z)}{tm_0(z)+1}.$$

Then

$$\begin{aligned} |c(z) - b(z)| &= d_N^{-1}\left|\int (|\mathcal{F}(t)|^2 - \mathcal{F}^2(t))dH_N(t)\right| \\ &= d_N^{-1}\left|2\int (\Im\mathcal{F}(t))^2 dH_N(t) - 2i\int \Re\mathcal{F}(t)\Im\mathcal{F}(t)dH_N(t)\right| \\ &\geq 2d_N^{-1}\int (\Im\mathcal{F}(t))^2 dH_N(t) \\ &\geq C(\Im m_0(z))^2\int |\mathcal{F}(t)|^2 dH_N(t). \end{aligned}$$

Here we have used the relation

$$\frac{\Im\mathcal{F}(t)}{|\mathcal{F}(t)|} = \frac{\Im m_0(z)}{|tm_0(z)+1||m_0(z)|}, \tag{3.62}$$

and the fact that

$$m_0(z) \sim 1, \quad 1 + tm_0(z) \sim 1, \quad \text{for } \; z \in S_r(\tilde{c}, 0), \; t \in Spec(\Sigma). \tag{3.63}$$

Moreover, by definition and (3.63) we also have

$$\int |\mathcal{F}(t)|^2 dH_N(t) \sim 1.$$

Therefore, if $\Im m_0(z) \sim 1$, we have

$$|c(z) - b(z)| \sim \Im m_0(z). \tag{3.64}$$

If $\Im m_0(z) \leq \epsilon$ for some sufficiently small constant $\epsilon > 0$, we can estimate $|c(z) - b(z)|$ as follows. Set

$$\mathcal{F}(t) = e^{i\phi(t)}|\mathcal{F}(t)|.$$

From (3.62) we see $\Im\mathcal{F}(t) > 0$, thus we have $\phi \in (0, \pi)$. By the assumption of $\Im m_0(z) \leq \epsilon$, (3.62) and (3.63) we obtain

$$\sin\phi \leq C\epsilon,$$

thus

$$|\cos\phi| \geq 1/2$$

for sufficiently small ϵ. Moreover, by (iv) of Lemma 2.3 and continuity, we have either $\cos\phi \geq 1/2$ or $\cos\phi \leq -1/2$ holds uniformly on $t \in [\lambda_M(\Sigma), \lambda_1(\Sigma)]$. Therefore, we have

$$|c(z) - b(z)| = d_N^{-1}\left|\int |\mathcal{F}(t)|^2 - \mathcal{F}^2(t) dH_N(t)\right|$$

$$
\begin{aligned}
&= 2d_N^{-1}\left|\int |\mathcal{F}(t)|^2 e^{i\phi}\sin\phi dH_N(t)\right| \\
&\geq 2d_N^{-1}\left|\int |\mathcal{F}(t)|^2 \cos\phi\sin\phi dH_N(t)\right| - 2d_N^{-1}\int |\mathcal{F}(t)|^2\sin^2\phi dH_N(t) \\
&\geq d_N^{-1}\int |\mathcal{F}(t)|^2(\sin\phi - 2\sin^2\phi)dH_N(t) \\
&\geq (2d_N)^{-1}\int |\mathcal{F}(t)|^2\sin\phi dH_N(t) \\
&\geq c\Im m_0(z)\int |\mathcal{F}(t)|^2 dH_N(t), \qquad (3.65)
\end{aligned}
$$

where in the last step we have used (3.62) and the fact that $\sin\phi = \Im\mathcal{F}(t)/\mathcal{F}(t)$. Therefore, by (3.64) and (3.65) we always have

$$|1-b(z)| \geq |c(z)-b(z)| \geq \epsilon\Im m_0(z)$$

for some positive constant ϵ. Similar to the analysis in (3.65), when $\Im m_0(z)$ is sufficiently small, we also have

$$
\begin{aligned}
|c(z)-b(z)| &\leq 2d_N^{-1}\left|\int |\mathcal{F}(t)|^2\cos\phi\sin\phi dH_N(t)\right| + 2d_N^{-1}\int |\mathcal{F}(t)|^2\sin^2\phi dH_N(t) \\
&\leq C\Im m_0(z) \qquad (3.66)
\end{aligned}
$$

for some positive constant C. Moreover, by (3.60) and (3.61) we also have

$$|1-b(z)| \geq 1-c(z) = \frac{|m_0(z)|^2}{\Im m_0(z)}\eta \sim \frac{\eta}{\Im m_0(z)}. \qquad (3.67)$$

Thus we have for some positive constants ϵ, ϵ',

$$|1-b(z)| \geq \epsilon\max\{\Im m_0(z), \frac{\eta}{\Im m_0(z)}\} \geq \epsilon'\sqrt{\kappa+\eta}, \qquad (3.68)$$

which is implied by (ii) of Lemma 2.3. Moreover, when $\Im m_0(z)$ is sufficiently small, by (3.66) we also have

$$|1-b(z)| \leq |1-c(z)| + |c(z)-b(z)| = O(\frac{\eta}{\Im m_0(z)}) + O(\Im m_0(z)) = O(\sqrt{\kappa+\eta}). \qquad (3.69)$$

Note that the function $\delta(E+i\eta)$ is decreasing in η, thus we can set

$$\eta_1 = \sup_{I_E}\left\{\eta : \delta(E+i\eta) \geq (\log N)^{-1}(\kappa+\eta)\right\}.$$

Solving (3.59) we obtain two solutions of $m_N(z)$ as $m_{N1}(z)$ and $m_{N2}(z)$ such that

$$m_{N1}(z) - m_0(z) = \frac{-(1-b(z)) + \sqrt{(1-b(z))^2 + 4a(z)\delta_0(z)}}{2a(z)},$$

and

$$m_{N2}(z) - m_0(z) = \frac{-(1-b(z)) - \sqrt{(1-b(z))^2 + 4a(z)\delta_0(z)}}{2a(z)}.$$

Here the square root is chosen to guarantee that

$$\sqrt{(1-b(z))^2 + 4a(z)\delta_0(z)} = (1-b(z)) + O(\delta_0(z))$$

when $\eta \sim 1$.

When $\Im m_0(z)$ is small enough, by (3.68) and (3.69) we see

$$1 - b(z) \sim \sqrt{\kappa + \eta}.$$

Therefore, when $\eta \le \eta_1$, by definition we have

$$\kappa + \eta \le (\log N)^{-7}$$

which implies

$$\Im m_0(z) \ll 1$$

by Lemma 2.3. Thus for $\eta \le \eta_1$, we have

$$|1 - b(z)| \sim \sqrt{\kappa + \eta} \le C(\log N)\sqrt{\delta(z)}. \tag{3.70}$$

For $\eta > \eta_1$, by the fact that $\delta(z)$ is decreasing in η we know

$$\delta(E + i\eta) \le (\log N)^{-1}(\kappa + \eta_1) \le (\log N)^{-1}(\kappa + \eta) \le C(\log N)^{-1}|1 - b(z)|^2. \tag{3.71}$$

Therefore, (3.70) and (3.71) imply that

$$\begin{aligned} |m_{N1}(z) - m_{N2}(z)| \quad &\ge \quad C|1 - b(z)| \ge C\sqrt{\kappa + \eta}, \qquad \eta > \eta_1 \\ &\le \quad C(\log N)\sqrt{\delta(z)}, \qquad \eta \le \eta_1. \end{aligned} \tag{3.72}$$

Observe that in $\Xi^c(z)$ we have

$$|m_0(z) - m_N(z)| \le (\log N)^{-1}.$$

Thus, when $\eta \sim 1$, it is obvious that

$$m_N(z) = m_{N1}(z) = m_0(z) + \frac{2\delta_0(z)}{\Big((1 - b(z)) + \sqrt{(1 - b(z))^2 + 4a(z)\delta_0(z)}\Big)}$$

and

$$|m_0(z) - m_N(z)| \le C\frac{\delta(z)}{|1 - b(z)|} \le C\frac{\delta(z)}{\sqrt{\kappa + \eta}}.$$

Now by (3.72) we know that $m_{N1}(z) \ne m_{N2}(z)$ when $\eta \ge \eta_1$. Therefore, for $\eta \ge \eta_1$, by continuity we have

$$m_N(z) = m_{N1}(z),$$

and

$$|m_0(z) - m_N(z)| \le C\frac{\delta(z)}{|1 - b(z)|} \le C\frac{\delta(z)}{|1 - b(z)| + \sqrt{\delta(z)}} \le C\frac{\delta(z)}{\sqrt{\kappa + \eta + \delta(z)}}.$$

When $\eta \le \eta_1$, we have

$$\begin{aligned} |m_0(z) - m_N(z)| \quad &\le \quad |m_0(z) - m_{N1}(z)| + |m_{N1}(z) - m_{N2}(z)| \\ &\le C(\log N)\sqrt{\delta(z)} \le C(\log N)\frac{\delta(z)}{\sqrt{\kappa + \eta + \delta(z)}}. \end{aligned}$$

Therefore, we proved (3.51) when $\eta \in I_E$.

It remains to show that when $\mathbf{S}$ occurs, I_E is just exactly $[\varphi^L N^{-1}, 1]$. The proof is nearly the same as the counterpart in [27]. However, for the convenience of the reader, we reproduce it here. We assume $I_E \neq [\varphi^L N^{-1}, 1]$ to get a contradiction. Note that if $I_E \neq [\varphi^L N^{-1}, 1]$, we can set $\eta_0 = \inf I_E$ which satisfies

$$\Lambda_o(E + i\eta_0) + \Lambda_d(E + i\eta_0) = (\log N)^{-1}. \tag{3.73}$$

Then by the definition of $\Gamma(z, K)$, we see in $\mathbf{S}$

$$\begin{aligned} &\Lambda_o(E + i\eta_0) + \max_i |G_{ii}(E + i\eta_0) - m_N(E + i\eta_0)| \\ &\leq K\Psi(E + i\eta_0) \leq K\varphi^{-L/2} \leq (\log N)^{-2} \end{aligned} \tag{3.74}$$

by the assumption of $\varphi^L \geq K^2(\log N)^4$. Moreover, since $\eta_0 \in I_E$, we also have

$$\Lambda(E + i\eta_0) \leq C(\log N)\sqrt{\delta(E + i\eta_0)} = O((\log N)^{-3}). \tag{3.75}$$

Thus (3.74) and (3.75) together imply that

$$\Lambda_o(E + i\eta_0) + \Lambda_d(E + i\eta_0) \leq O((\log N)^{-2}),$$

which contradicts with (3.73) when N is sufficiently large. Thus we complete the proof. □

Now we can start to prove Theorem 3.3.

Proof of Theorem 3.3. Note that by the definitions of $\Omega(z, K)$ and $\Gamma(z, K)$, we have for any $\zeta > 0$, there exists a positive constants C_ζ such that

$$\begin{aligned} |\mathcal{D}(m_N)(z)| &\leq |[Y]| + O(\max_i |G_{ii}(z) - m_N(z)|^2) + \infty \mathbf{1}_{\Xi(z)} \\ &\leq \varphi^{C_\zeta}\Psi(z) + \infty \mathbf{1}_{\Xi(z)}, \quad \forall z \in S_r(\tilde{c}, 5C_\zeta) \end{aligned}$$

holds on the event

$$\bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \Gamma(z, \varphi^{C_\zeta}).$$

Now set

$$\mathbf{S} = \bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \Gamma(z, \varphi^{C_\zeta}) \cap \bigcap_{z \in S_r(\tilde{c}, 5C_\zeta), \eta = 1} \Xi^c(z).$$

and

$$\delta(z) = \varphi^{C_\zeta}(N\eta)^{-1/2}$$

as in Lemma 3.13. We can get that

$$\Lambda(z) \leq \varphi^{C_\zeta}(N\eta)^{-1/4}, \quad \forall z \in S_r(\tilde{c}, 5C_\zeta)$$

holds in $\mathbf{S}$. Moreover, by Lemma 3.13 we know

$$\mathbf{S} \subset \bigcap_{z \in S_r(\tilde{c}, 5C_\zeta)} \Xi^c(z). \tag{3.76}$$

Thus we have

$$\Lambda \leq \Lambda_d \leq (\log N)^{-1}, \quad \Psi(z) \leq C(N\eta)^{-1/2}, \quad \text{in } \mathbf{S}.$$

Moreover, by the definition of $\mathbf{S}$, Lemma 3.8 and the fact that

$$\bigcap_{z\in S_r(\tilde{c},5C_\zeta),\eta=1} \Xi^c(z)$$

holds with ζ-high probability which is implied by Lemma 3.5 and 3.6, we see that $\mathbf{S}$ holds with ζ-high probability. Then by (3.76) and Lemma 3.8 we also know

$$\bigcap_{z\in S_r(\tilde{c},5C_\zeta)} \Omega^c(z,\varphi^{C_\zeta})$$

holds with ζ-high probability. Therefore, by the definition of $\Omega(z,\varphi^{C_\zeta})$ we see that Theorem 3.3 follows. □

3.2. *Improved bound for* $[Y]$. In order to prove Theorem 3.2, the main thing is to provide a stronger bound for $[Y]$. More precisely, we need stronger bounds for $[T]$ and $v_1(z)$. Here

$$[T]=\frac{1}{N}\sum_{i=1}^N T_i(z).$$

Similarly, we will denote

$$[U]=\frac{1}{N}\sum_{i=1}^N U_i(z).$$

Recall the decomposition

$$Y_i=T_i+U_i+V=T_i+U_i+\sum_{k=1}^4 v_k,$$

which implies that

$$[Y]=[T]+[U]+\sum_{k=1}^4 v_k.$$

Noting that by definitions, in $\Gamma(z,K)\cap\Xi^c(z)$, we have

$$|U_i|,|v_2|,|v_3|,|v_4|=O(K^2\Psi^2).$$

Such a strong bound is enough for our purpose. Thus our main task in this subsection is to improve the bounds for $[T]$ and v_1. Actually, we will prove the following lemma.

Lemma 3.14. *Let $K=\varphi^{O(1)}$ and $0<L=O(1)$ satisfying $\varphi^L\geq K^2(\log N)^4$. Suppose that for some event*

$$\Theta\subset\bigcap_{z\in S_r(\tilde{c},L)}\left(\Gamma(z,K)\cap\Xi^c(z)\right),$$

we have

$$\Lambda(z)\leq\tilde{\Lambda}(z)\leq 1,\quad \forall z\in S_r(\tilde{c},L),$$

where $\tilde{\Lambda}(z)$ is some deterministic number. And we also have

$$\mathbb{P}(\Theta^c)\leq\exp(-p(\log N)^2)$$

for some

$$1 \ll p \ll \min\{\varphi^{L/2} K^{-1} (\log N)^{-1}, \frac{1}{2} K^{\tau^{-1}} (\log N)^{-2}\}, \tag{3.77}$$

where τ is the parameter in Lemma 3.4. Then there exists an event $\Theta' \subset \Theta$ such that

$$\mathbb{P}((\Theta')^c) \leq \frac{1}{2} \exp(-p) \tag{3.78}$$

and for any $z \in S_r(\tilde{c}, L)$, we have

$$|[T]| + |v_1| \leq C p^5 K^2 \left(\tilde{\Psi}^2 + \tilde{\Lambda} \tilde{\Psi} \right), \quad \tilde{\Psi} := \sqrt{\frac{\Im m_0 + \tilde{\Lambda}}{N\eta}}, \quad \text{in } \Theta', \tag{3.79}$$

which implies

$$|[Y]| \leq C p^5 K^2 \left(\tilde{\Psi}^2 + \tilde{\Lambda} \tilde{\Psi} \right), \quad \text{in } \Theta'.$$

Remark 3.15. *This lemma can be regarded as an extension of Lemma 7.1 of [27] to the non-null case. Moreover, it is also analogous to Lemma 5.2 of [19], Corollary 4.2 of [21] and Lemma 4.1 of [12] for Wigner matrices or the adjacency matrices of Erdös-Rényi graphs.*

By definitions, we have

$$[T] = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{r}_i^* \mathcal{G}^{(i)} \mathbf{r}_i - \frac{1}{N} Tr \mathcal{G}^{(i)} \Sigma \right),$$

and

$$v_1 = \hat{v}_1 + \check{v}_1$$

with

$$\hat{v}_1 := -m_N(z) \frac{1}{N} \sum_{i=1}^{N} v_{1i}(z)$$

$$\check{v}_1 := -\frac{1}{N} \sum_{i=1}^{N} (G_{ii}(z) - m_N(z)) v_{1i}(z).$$

Here $v_{1i}(z)$ is defined in (3.10).

Observe that in $\bigcap_{z \in S_r(\tilde{c}, L)} (\Gamma(z, K) \cap \Xi^c(z))$, there exist the following bounds

$$|v_{1i}|, |G_{ii}(z) - m_N(z)| \leq K \Psi(z).$$

Thus in Θ there exists

$$|\check{v}_1| \leq C K^2 \tilde{\Psi}^2.$$

Noting that in $\Xi^c(z)$, we also have

$$m_N(z) \sim 1.$$

Hence, it suffices to bound $[T]$ and

$$\omega := \omega(z) = \frac{1}{N} \sum_{i=1}^{N} v_{1i}(z).$$

Note that both $[T]$ and ω are in the form of

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i^* A^{(i)}\mathbf{x}_i - \frac{1}{N}TrA^{(i)}) \tag{3.80}$$

with some matrix $A^{(i)}$ independent of $\mathbf{x}_i$ but constructed from all the other $\mathbf{x}_j, j \neq i$. If we use the notation $\mathbb{E}_i$ to denote the expectation with respect to $\mathbf{x}_i$, we can also write

$$[T] = \frac{1}{N}\sum_{i=1}^{N}(1-\mathbb{E}_i)\mathbf{r}_i^*\mathcal{G}^{(i)}\mathbf{r}_i$$

and

$$\omega = \frac{1}{N}\sum_{i=1}^{N}(1-\mathbb{E}_i)\mathbf{r}_i^*\mathcal{G}^{(i)}\Sigma(m_N^{(i)}(z)\Sigma + I)^{-1}\mathbf{r}_i.$$

In the sequel, we will use the notation

$$Q_{\mathbb{A}} := \prod_{k\in\mathbb{A}}(1-\mathbb{E}_k),$$

and $Q_{\{i\}} = 1-\mathbb{E}_i$ will be simply denoted by Q_i. To bound summation in the form of (3.80) , we will adopt the abstract decoupling lemma of [27] (see Lemma 7.3 therein), which is a large deviation lemma for weakly hierarchically coupled random variables. We cite it as the following lemma.

LEMMA 3.16 (Lemma 7.3, [27]). *Let $Z_1,\dots,Z_N$ be random variables which are functions of $x_{ij}, 1\le i\le M, 1\le j\le N$. Let Θ be an event and p an even integer, which may depend on N. Suppose the following assumptions hold with some constants $C_1, c_1 > 0$.*

(i) There exist some deterministic positive numbers $\mathcal{X} < 1$ and $\mathcal{Y}$ such that for any set $\mathbb{A}\subset\{1,2,\dots,N\}$ with $i\in\mathbb{A}$ and $|\mathbb{A}|\le p$, $\mathbf{1}(\Theta)(Q_{\mathbb{A}}Z_i)$ can be written as the sum of two new random variables

$$\mathbf{1}(\Theta)(Q_{\mathbb{A}}Z_i) = Z_{i,\mathbb{A}} + \mathbf{1}(\Theta)Q_{\mathbb{A}}\mathbf{1}(\Theta^c)\tilde{Z}_{i,\mathbb{A}}$$

and

$$|Z_{i,\mathbb{A}}| \le \mathcal{Y}(C_1\mathcal{X}|\mathbb{A}|)^{|\mathbb{A}|}, \quad |\tilde{Z}_{i,\mathbb{A}}| \le \mathcal{Y}N^{C_1|\mathbb{A}|}.$$

Here $\mathbf{1}(\Theta)$ represents the indicator function of Θ.

(ii) For Z_i, we have the rough definite bound

$$\max_i |Z_i| \le \mathcal{Y}N^{C_1}.$$

(iii) For Θ we have

$$\mathbb{P}(\Theta^c) \le \exp(-c_1(\log N)^{3/2}p).$$

Then, under the assumptions (i), (ii) and (iii) we have

$$\mathbb{E}\left(N^{-1}\sum_{i=1}^{N}Q_iZ_i\right)^p \le (Cp)^{4p}[\mathcal{X}^2 + N^{-1}]^p\mathcal{Y}^p \tag{3.81}$$

for some $C > 0$ and any sufficiently large N.

REMARK 3.17. *We remind here similar results also appeared in some previous work, such as Theorem 5.6 of [12] and Lemma 4.1 of [21].*

REMARK 3.18. *Noting that once we have (3.81), we can use the Markov inequality to get*

$$\mathbb{P}\left(|N^{-1}\sum_{i=1}^{N} Q_i Z_i| \geq C\mathcal{Y}p^5[\mathcal{X}^2 + N^{-1}]\right) \ll \frac{1}{2}\exp(-p).$$

With the aid of Lemma 3.16, we now come to prove Lemma 3.14.

PROOF OF LEMMA 3.14. Note that it suffices to find an event $\Theta' \in \Theta$ satisfying (3.78) such that

$$|[T]| \leq Cp^5 K^2 \tilde{\Psi}^2 \tag{3.82}$$

and

$$|\omega| \leq Cp^5 K^2 \left(\tilde{\Psi}^2 + \tilde{\Lambda}\tilde{\Psi}\right)$$

in Θ'. To simplify the discussion on ω we introduce the quantities

$$\begin{aligned} \tilde{\nu}_i : &= \mathbf{r}_i^* \mathcal{G}^{(i)} \Sigma (m_0(z)\Sigma + I)^{-1} \mathbf{r}_i \\ \tilde{v}_{1i} : &= Q_i \tilde{\nu}_i = \mathbf{r}_i^* \mathcal{G}^{(i)} \Sigma (m_0(z)\Sigma + I)^{-1} \mathbf{r}_i - \frac{1}{N} Tr(m_0(z)\Sigma + I)^{-1} \Sigma \mathcal{G}^{(i)} \Sigma \end{aligned}$$

and

$$\tilde{\omega} = \frac{1}{N}\sum_{i=1}^{N} \tilde{v}_{1i}.$$

It suffices to show that in Θ', there exist (3.82) and

$$|\omega - \tilde{\omega}| \leq C\left(K\tilde{\Lambda}\tilde{\Psi} + K^2\tilde{\Psi}^2\right), \tag{3.83}$$

$$|\tilde{\omega}| \leq Cp^5 K^2 \tilde{\Psi}^2. \tag{3.84}$$

At first, we come to show (3.83). Noting that by definition, we have

$$\begin{aligned} |v_{1i} - \tilde{v}_{1i}| &= \left| Q_i \left(\mathbf{r}_i^* \mathcal{G}^{(i)} \Sigma \left[(m_N^{(i)}(z)\Sigma + I)^{-1} - (m_0(z)\Sigma + I)^{-1} \right] \mathbf{r}_i \right) \right| \\ &= |m_N^{(i)}(z) - m_0(z)| \left| Q_i \left(\mathbf{r}_i^* \mathcal{G}^{(i)} \Sigma (m_N^{(i)}(z)\Sigma + I)^{-1} \Sigma (m_0(z)\Sigma + I)^{-1} \mathbf{r}_i \right) \right| \\ &\leq CK(\Lambda + \Lambda_o^2)\Psi \leq C(K\Lambda\Psi + K^3\Psi^3) \leq C(K\Lambda\Psi + K^2\Psi^2). \end{aligned}$$

Here in the first inequality above we have used the fact that $|m_N^{(i)}(z) - m_0(z)| \leq \Lambda + C\Lambda_o^2$ and in the last step we have used the fact that $K\Psi \ll 1$ which is implied by the assumption on K and L immediately. Note here we have used the fact that in $\Gamma(z, K) \cap \Xi^c(z)$,

$$|R_i| = |Q_i\left(\mathbf{r}_i^* \mathcal{G}^{(i)} \Sigma (m_N^{(i)}(z)\Sigma + I)^{-1} \Sigma (m_0(z)\Sigma + I)^{-1} \mathbf{r}_i\right)| \leq K\Psi(z).$$

Thus it remains to show (3.82) and (3.84). We will adopt Lemma 3.16 to prove both of these two inequlities.

We remind here that we learn the proof strategy which will be used below from [12].

At first, we come to deal with the simpler one (3.82). Note that

$$z + z\mathbf{r}_i^* \mathcal{G}^{(i)} \mathbf{r}_i = -(G_{ii}(z))^{-1}.$$

Thus

$$T_i = Q_i (zG_{ii}(z))^{-1}.$$

Since $z \in S_r(\tilde{c}, L)$, it suffices to bound

$$\frac{1}{N} \sum_{i=1}^{N} Q_i (G_{ii}(z))^{-1}.$$

Now we can set $Z_i = (G_{ii}(z))^{-1}$ in Lemma 3.16. Then by Lemma 3.16 and the Markov inequality it suffices to show that for $i \in \mathbb{A} \subset \{1, 2, \dots, N\}$ and $|\mathbb{A}| \le p$ there exists $Z_{i,\mathbb{A}}$ and $\tilde{Z}_{i,\mathbb{A}}$ satisfying

$$\mathbf{1}(\Theta)(Q_{\mathbb{A}} Z_i) = Z_{i,\mathbb{A}} + \mathbf{1}(\Theta) Q_{\mathbb{A}} \mathbf{1}(\Theta^c) \tilde{Z}_{i,\mathbb{A}}, \quad Z_{i,\mathbb{A}} \le \mathcal{Y}(C\mathcal{X}|\mathbb{A}|)^{|\mathbb{A}|}, \quad \tilde{Z}_{i,\mathbb{A}} \le \mathcal{Y} N^{C|\mathbb{A}|} \tag{3.85}$$

with

$$\mathcal{X} = K\tilde{\Psi}, \quad \mathcal{Y} = C \tag{3.86}$$

for some positive constant C. The proof of (3.85) is totally the same as the counterpart in the null case, see the proof of Lemma 7.4 and Lemma 7.1 of [27]. The proof only depends on the equations listed in Lemma 3.1, thus is also valid for our non-null case. So here we omit the details. Our main target is to prove (3.84) in the sequel.

By Lemma 3.16, it suffices to show that there exists an event Θ with probability

$$\mathbb{P}(\Theta) \ge 1 - \exp(-(\log N)^{3/2} p)$$

such that for $i \in \mathbb{A} \subset \{1, 2, \dots, N\}$ and $|\mathbb{A}| \le p$ there exists $\tilde{\nu}_{i,\mathbb{A}}$ and $\hat{\nu}_{i,\mathbb{A}}$ satisfying

$$\mathbf{1}(\Theta)(Q_{\mathbb{A}} \tilde{\nu}_i) = \tilde{\nu}_{i,\mathbb{A}} + \mathbf{1}(\Theta) Q_{\mathbb{A}} \mathbf{1}(\Theta^c) \hat{\nu}_{i,\mathbb{A}}, \quad \tilde{\nu}_{i,\mathbb{A}} \le \mathcal{Y}(C\mathcal{X}|\mathbb{A}|)^{|\mathbb{A}|}, \quad \hat{\nu}_{i,\mathbb{A}} \le \mathcal{Y} N^{C|\mathbb{A}|} \tag{3.87}$$

for some constant $C > 0$, where $\mathcal{X}$ and $\mathcal{Y}$ are specified to be those in (3.86).

Noting that by the fact $\eta \gg N^{-1}$ in $S_r(\tilde{c}, L)$ and the truncation on x_{ij}, we see (3.87) holds naturally when $\mathbb{A} = \{i\}$. Thus we can always assume $|\mathbb{A}| \ge 2$ in the sequel. Now let $\mathcal{A} = \mathcal{A}(\mathcal{W})$ be a function of $\mathcal{W}$ and denote $\mathcal{A}^{(\mathbb{T})} = \mathcal{A}(\mathcal{W}^{(\mathbb{T})})$. Pursuing the idea in [12], we define

$$\mathcal{A}^{\mathbb{S},\mathbb{U}} := \sum_{\mathbb{S} \setminus \mathbb{U} \subset \mathbb{V} \subset \mathbb{S}} (-1)^{|\mathbb{V}|} \mathcal{A}^{(\mathbb{V})},$$

for any $\mathbb{S}, \mathbb{U} \subset \{1, 2, \dots, N\}$. Then by definition it is not difficult to see for any $\mathbb{S} \subset \{1, \dots, N\}$, there exists

$$\mathcal{A} = \sum_{\emptyset \subset \mathbb{U} \subset \mathbb{S}} \mathcal{A}^{\mathbb{S},\mathbb{U}}.$$

Moreover, it is not difficult to see $\mathcal{A}^{\mathbb{S},\mathbb{U}}$ is independent of the k-th column of X if $k \in \mathbb{S} \setminus \mathbb{U}$. Thus we have

$$Q_{\mathbb{S}} \mathcal{A} = Q_{\mathbb{S}} \sum_{\emptyset \subset \mathbb{U} \subset \mathbb{S}} \mathcal{A}^{\mathbb{S},\mathbb{U}} = Q_{\mathbb{S}} \mathcal{A}^{\mathbb{S},\mathbb{S}},$$

which implies that

$$Q_{\mathbb{A}}\tilde{\nu}_i = Q_i Q_{\mathbb{A}\setminus\{i\}}\tilde{\nu}_i = Q_{\mathbb{A}}(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}}.$$

Now we set

$$\tilde{\nu}_{i,\mathbb{A}} := \mathbf{1}(\Theta)Q_{\mathbb{A}}\mathbf{1}(\Theta)(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}}, \quad \hat{\nu}_{i,\mathbb{A}} := (\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}}.$$

It is easy to get the bound of $\hat{\nu}_{i,\mathbb{A}}$ in (3.87), thus we only need to handle $\tilde{\nu}_{i,\mathbb{A}}$ below. That is to say, we shall show that for $2 \le |\mathbb{A}| \le p$, there exist

$$\left|\mathbf{1}(\Theta)(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}}\right| \le \mathcal{Y}(C\mathcal{X}|\mathbb{A}|)^{|\mathbb{A}|}, \quad \mathcal{X} = K\tilde{\Psi}, \quad \mathcal{Y} = C. \tag{3.88}$$

Now we need the following two lemmas.

Lemma 3.19. *Let p be an even number satisfying*

$$1 \ll p \ll \varphi^{L/2}K^{-1}(\log N)^{-1},$$

then in $\bigcap_{z\in S_r(\tilde{c},L)}(\Gamma(z,K)\cap\Xi^c(z))$, for any $\mathbb{T}$ with $|\mathbb{T}|\le p$, there exist

$$\max_{i,j\notin T}|G_{ij}^{(\mathbb{T})}| \le C\max_{i,j}|G_{ij}|, \quad \min_{i\notin\mathbb{T}}|G_{ii}^{(\mathbb{T})}| \ge c\min_i|G_{ii}|,$$
$$\frac{1}{N}\Im Tr\mathcal{G}^{(\mathbb{T})} \le \frac{1}{N}\Im Tr\mathcal{G} + Cp(K^2\Psi^2 + \frac{1}{N})$$

for some positive constants C, c.

For ease of presentation, we denote

$$M_0 := M_0(z) = (m_0(z)\Sigma + I)^{-1}.$$

And now we set the event

$$\Upsilon(z) := \bigcap_{i\neq j}\left\{\mathbf{r}_i^*\mathcal{G}^{(\mathbb{T}\cup\{i,j\})}(z)\Sigma M_0(z)\mathbf{r}_j^* \le K\Psi(z) \text{ for all } \mathbb{T}\in\{1,\ldots,N\}, i,j\notin\mathbb{T}, |\mathbb{T}|\le p\right\}.$$

Then we have the following lemma

Lemma 3.20. *Let p be an even number satisfying*

$$1 \ll p \ll \min\{\frac{1}{2}K^{\tau^{-1}}(\log N)^{-2}, \varphi^{L/2}K^{-1}(\log N)^{-1}\}.$$

Then in $\bigcap_{z\in S_r(\tilde{c},L)}(\Gamma(z,K)\cap\Xi^c(z))$, we have $\bigcap_{z\in S_r(\tilde{c},L)}\Upsilon(z)$ holds with probability larger than

$$1 - \exp(-CK^{\tau^{-1}})$$

with some positive constant C.

In the sequel, we will assume

$$\Theta \subset \bigcap_{z\in S_r(\tilde{c},L)}(\Gamma(z,K)\cap\Xi^c(z)\cap\Upsilon(z)).$$

Now we prove (3.88) assuming Lemma 3.19 and 3.20 at first and postpone the proofs of these two lemmas after that. As a warm-up, we start with the case of $|\mathbb{A}| = 2$ and $|\mathbb{A}| = 3$.

Let $\mathbb{A} = \{i, j\}$ with some $j \neq i$. Note

$$(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}} = \tilde{\nu}_i^{\{j\},\{j\}} = \tilde{\nu}_i - \tilde{\nu}_i^{(j)}.$$

Note that for any set $\mathbb{T}$ we have the relation

$$\mathbf{r}_i^* \mathcal{G}^{(\mathbb{T}\cup\{i,j\})} \mathbf{r}_j = z^{-1} \frac{G_{ij}^{(\mathbb{T})}}{G_{ii}^{(\mathbb{T})} G_{jj}^{(\mathbb{T}\cup\{i\})}}, \quad \text{for} \quad i, j \notin \mathbb{T}.$$

Thus we have

$$\begin{aligned}
(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}} &= \mathbf{r}_i^* \mathcal{G}^{(i)} \Sigma M_0 \mathbf{r}_i - \mathbf{r}_i^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i \\
&= -\mathbf{r}_i^* \mathcal{G}^{(i)} \mathbf{r}_j \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i \\
&= -\frac{1}{1 + \mathbf{r}_j^* \mathcal{G}^{(ij)} \mathbf{r}_j} \mathbf{r}_i^* \mathcal{G}^{(ij)} \mathbf{r}_j \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i \\
&= z G_{jj}^{(i)} \mathbf{r}_i^* \mathcal{G}^{(ij)} \mathbf{r}_j \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i \\
&= G_{ij} (G_{ii})^{-1} \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i \\
&\leq O(\mathcal{X}^2) \qquad \text{in} \quad \Theta.
\end{aligned} \tag{3.89}$$

If $|\mathbb{A}| = 3$, without loss of generality, we assume that $\mathbb{A} = \{i, j, k\}$, it is easy to see

$$\begin{aligned}
(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}} &= G_{ij} (G_{ii})^{-1} \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i - G_{ij}^{(k)} (G_{ii}^{(k)})^{-1} \mathbf{r}_j^* \mathcal{G}^{(ijk)} \Sigma M_0 \mathbf{r}_i \\
&= (G_{ij} - G_{ij}^{(k)}) (G_{ii})^{-1} \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i \\
&\quad + G_{ij}^{(k)} ((G_{ii})^{-1} - (G_{ii}^{(k)})^{-1}) \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i \\
&\quad + G_{ij}^{(k)} (G_{ii}^{(k)})^{-1} (\mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i - \mathbf{r}_j^* \mathcal{G}^{(ijk)} \Sigma M_0 \mathbf{r}_i).
\end{aligned} \tag{3.90}$$

Note that for any set $\mathbb{T}$ such that $i, j, k \notin \mathbb{T}$, there exist

$$\begin{aligned}
&G_{ij}^{(\mathbb{T})} - G_{ij}^{(\mathbb{T}\cup\{k\})} = G_{ik}^{(\mathbb{T})} (G_{kk}^{(\mathbb{T})})^{-1} G_{kj}^{(\mathbb{T})}, \\
&(G_{ii}^{(\mathbb{T})})^{-1} - (G_{ii}^{(\mathbb{T}\cup\{k\})})^{-1} = -(G_{ii}^{(\mathbb{T})})^{-1} G_{ik}^{(\mathbb{T})} (G_{kk}^{(\mathbb{T})})^{-1} G_{ki}^{(\mathbb{T})} (G_{ii}^{(\mathbb{T}\cup\{k\})})^{-1}
\end{aligned} \tag{3.91}$$

and

$$\mathbf{r}_j^* \mathcal{G}^{(\mathbb{T}\cup\{i,j\})} \Sigma M_0 \mathbf{r}_i - \mathbf{r}_j^* \mathcal{G}^{(\mathbb{T}\cup\{i,j,k\})} \Sigma M_0 \mathbf{r}_i = G_{jk}^{(\mathbb{T}\cup\{i\})} (G_{jj}^{(\mathbb{T}\cup\{i\})})^{-1} \mathbf{r}_k^* \mathcal{G}^{(\mathbb{T}\cup\{i,j,k\})} \Sigma M_0 \mathbf{r}_i. \tag{3.92}$$

Thus we have

$$\begin{aligned}
(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\},\mathbb{A}\setminus\{i\}} &= \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i (G_{ii})^{-1} G_{ik} (G_{kk})^{-1} G_{kj} \\
&\quad - \mathbf{r}_j^* \mathcal{G}^{(ij)} \Sigma M_0 \mathbf{r}_i (G_{ii})^{-1} G_{ik} (G_{kk})^{-1} G_{ki} (G_{ii}^{(k)})^{-1} G_{ij}^{(k)} \\
&\quad + \mathbf{r}_k^* \mathcal{G}^{(ijk)} \Sigma M_0 \mathbf{r}_i (G_{ii}^{(k)})^{-1} G_{ij}^{(k)} (G_{jj}^{(i)})^{-1} G_{jk}^{(i)} \\
&\leq O(\mathcal{X}^3), \qquad \text{in} \quad \Theta.
\end{aligned}$$

Now we use the idea in [12] (see Section 5.2 therein) to introduce a class of rational functions of resolvent matrix elements. For any fixed positive integer n we define

(i) a sequence of integers $(i_r)_{r=1}^{n+1}$ satisfying $i_k \neq i_{k+1}$ for $1 \leq k \leq n-1$ while $i_1 = i_{n+1}$;

(ii) a collection of sets $(\mathbb{U}_\alpha)_{\alpha=1}^n$ satisfying $i_1, i_2 \in \mathbb{U}_1$ and $i_\alpha, i_{\alpha+1} \notin \mathbb{U}_\alpha$, $\alpha \geq 2$, and $\emptyset \subset \mathbb{U}_\alpha \subset \mathbb{A}$ for all $1 \leq \alpha \leq n$;

(iii) a collection of sets $(\mathbb{T}_\beta)_{\beta=2}^n$ satisfying $i_\beta \notin \mathbb{T}_\beta$, and $\emptyset \subset \mathbb{T}_\beta \subset \mathbb{A}$ for all $2 \le \beta \le n$.
Then we define the random rational function

$$D\left((i_r)_{r=1}^{n+1}, (\mathbb{U}_\alpha)_{\alpha=1}^n, (\mathbb{T}_\beta)_{\beta=2}^n\right) := \frac{P}{Q}, \tag{3.93}$$

where

$$P = \mathbf{r}_{i_1}^* \mathcal{G}^{(\mathbb{U}_1)} \Sigma M_0 \mathbf{r}_{i_2} \prod_{\alpha=2}^n G_{i_\alpha, i_{\alpha+1}}^{(\mathbb{U}_\alpha)}, \quad Q = \prod_{\beta=2}^n G_{i_\beta, i_\beta}^{(\mathbb{T}_\beta)}.$$

Noting that in the cases of $|A| = 2$ and $|A| = 3$, we can write $(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\}, \mathbb{A}\setminus\{i\}}$ as a summation of rational functions in the form of

$$\pm D\left((i_r)_{r=1}^{n+1}, (\mathbb{U}_\alpha)_{\alpha=1}^n, (\mathbb{T}_\beta)_{\beta=2}^n\right). \tag{3.94}$$

Actually, for general $\mathbb{S}$ we have the following lemma.

Lemma 3.21. *Let $\mathbb{S} \subset \{1, \ldots, N\}$ and $i \notin \mathbb{S}$. Then we have*

$$(\tilde{\nu}_i)^{\mathbb{S},\mathbb{S}} = \sum_{n=|\mathbb{S}|+1}^{2|\mathbb{S}|} D_n, \quad D_n = \sum_{k=1}^{K_n} D_{n,k},$$

where

$$\sum_{n=|\mathbb{S}|+1}^{2\mathbb{S}} K_n \le 4^{|\mathbb{S}|} |\mathbb{S}|!,$$

and each $D_{n,k}$ is in the form of (3.94), with appropriate chosen sets $(\mathbb{U}_\alpha)_{\alpha=1}^n$ and $(\mathbb{T}_\beta)_{\beta=2}^n$ which may be different for each $F_{n,k}$.

Proof. Set

$$\mathbb{A} = \mathbb{S} \cup \{i\}.$$

We prove it by induction. Note that we already proved the cases of $|\mathbb{A}| = 2, 3$. At first, it is not difficult to check the following relation,

$$(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\}, \mathbb{A}\setminus\{i\}} = (\tilde{\nu}_i)^{\mathbb{A}\setminus\{i,j\}, \mathbb{A}\setminus\{i,j\}} - \left((\tilde{\nu}_i)^{\mathbb{A}\setminus\{i,j\}, \mathbb{A}\setminus\{i,j\}}\right)^{(j)}$$

for any $j \ne i$ while $j \in A$. Now we assume that $(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i,j\}, \mathbb{A}\setminus\{i,j\}}$ can be written in the form of

$$(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i,j\}, \mathbb{A}\setminus\{i,j\}} = \sum_{n=|\mathbb{A}\setminus\{i,j\}|+1}^{2|\mathbb{A}\setminus\{i,j\}|} D_n, \quad D_n = \sum_{k=1}^{K_n} D_{n,k}$$

with some $D_{n,k}$ in the form of (3.94) and K_n satisfying

$$\sum_{n=|\mathbb{A}\setminus\{i,j\}|+1}^{2|\mathbb{A}\setminus\{i,j\}|} K_n \le 4^{|\mathbb{A}\setminus\{i,j\}|} |\mathbb{A} \setminus \{i,j\}|!.$$

Then

$$(\tilde{\nu}_i)^{\mathbb{A}\setminus\{i\}, \mathbb{A}\setminus\{i\}} = \sum_{n=|\mathbb{A}\setminus\{i,j\}|+1}^{2|\mathbb{A}\setminus\{i,j\}|} (D_n - D_n^{(j)}) = \sum_{n=|\mathbb{A}\setminus\{i,j\}|+1}^{2|\mathbb{A}\setminus\{i,j\}|} \sum_{k=1}^{K_n} (D_{n,k} - D_{n,k}^{(j)}).$$

Now we assume that there are $(i_r)_{r=1}^{n+1}$, $(\mathbb{U}_\alpha)_{\alpha=1}^n$, $(\mathbb{T}_\beta)_{\beta=2}^n$ such that

$$D_{n,k} = \mathbf{r}_{i_1}^* \mathcal{G}^{(\mathbb{U}_1)} \Sigma M_0 \mathbf{r}_{i_2} \prod_{\alpha=2}^n G_{i_\alpha, i_{\alpha+1}}^{(\mathbb{U}_\alpha)} \prod_{\beta=2}^n \left(G_{i_\beta, i_\beta}^{(\mathbb{T}_\beta)} \right)^{-1}.$$

Then

$$\begin{aligned} D_{n,k} - D_{n,k}^{(j)} &= \mathbf{r}_{i_1}^* \mathcal{G}^{(\mathbb{U}_1)} \Sigma M_0 \mathbf{r}_{i_2} \prod_{\alpha=2}^n G_{i_\alpha, i_{\alpha+1}}^{(\mathbb{U}_\alpha)} \prod_{\beta=2}^n \left(G_{i_\beta, i_\beta}^{(\mathbb{T}_\beta)} \right)^{-1} \\ &\quad - \mathbf{r}_{i_1}^* \mathcal{G}^{(\mathbb{U}_1 \cup \{j\})} \Sigma M_0 \mathbf{r}_{i_2} \prod_{\alpha=2}^n G_{i_\alpha, i_{\alpha+1}}^{(\mathbb{U}_\alpha \cup \{j\})} \prod_{\beta=2}^n \left(G_{i_\beta, i_\beta}^{(\mathbb{T}_\beta \cup \{j\})} \right)^{-1} \end{aligned}$$

Now obviously $D_{n,k} - D_{n,k}^{(j)}$ can be written as a sum of $(2n-1)$ terms, and every term contains one and only one factor in one of the following forms

$$\begin{aligned} &\mathbf{r}_{i_1}^* \mathcal{G}^{(\mathbb{U}_1)} \Sigma M_0 \mathbf{r}_{i_2} - \mathbf{r}_{i_1}^* \mathcal{G}^{(\mathbb{U}_1 \cup \{j\})} \Sigma M_0 \mathbf{r}_{i_2}, \\ &G_{i_\alpha, i_{\alpha+1}}^{(\mathbb{U}_\alpha)} - G_{i_\alpha, i_{\alpha+1}}^{(\mathbb{U}_\alpha \cup \{j\})}, \quad \left(G_{i_\beta, i_\beta}^{(\mathbb{T}_\beta)} \right)^{-1} - \left(G_{i_\beta, i_\beta}^{(\mathbb{T}_\beta \cup \{j\})} \right)^{-1}. \end{aligned}$$

Now using (3.90) and (3.91) we see that $D_{n,k} - D_{n,k}^{(j)}$ is a sum of terms in the form of (3.93) with n replaced by $n+1$ or $n+2$. Moreover, the total number of these terms in $D_{n,k} - D_{n,k}^{(j)}$ is not more than $2n-1$. These facts together imply that we can write

$$(\tilde{\nu}_i)^{\mathbb{A} \setminus \{i\}, \mathbb{A} \setminus \{i\}} = \sum_{n=|\mathbb{A} \setminus \{i\}|+1}^{2|\mathbb{A} \setminus \{i\}|} \sum_{k=1}^{K'_n} D'_{n,k}$$

and

$$\begin{aligned} \left| \sum_{n=|\mathbb{A} \setminus \{i\}|+1}^{2|\mathbb{A} \setminus \{i\}|} \sum_{k=1}^{K'_n} \right| &\leq \left| \sum_{n=|\mathbb{A} \setminus \{i,j\}|+1}^{2|\mathbb{A} \setminus \{i,j\}|} K_n (2n-1) \right| \\ &\leq 4^{|\mathbb{A} \setminus \{i,j\}|+1} |\mathbb{A} \setminus \{i,j\}|! |\mathbb{A} \setminus \{i,j\}| \\ &\leq 4^{|\mathbb{A} \setminus \{i\}|} |\mathbb{A} \setminus \{i\}|! \end{aligned}$$

Thus we can complete the proof by induction. □

Now by the definition of D_n, Lemma 3.19 and 3.20 we can find an event

$$\Theta \subset \bigcap_{z \in S_r(\tilde{c}, L)} (\Gamma(z, K) \cap \Xi^c(z) \cap \Upsilon(z))$$

with probability

$$\mathbb{P}(\Theta) \geq 1 - \exp(-p(\log N)^{3/2})$$

such that (3.88) holds. Thus we conclude the proof of Lemma 3.14. □

Now we come to prove Lemma 3.19 and Lemma 3.20.

PROOF OF LEMMA 3.19. Noting that by (iii) of Lemma 3.1, we have

$$(G_{ii})^{-1} = (G_{ii}^{(j)})^{-1} - \frac{G_{ij}G_{ji}}{G_{ii}G_{jj}G_{ii}^{(j)}} = (1 + O(\mathcal{X}^2))(G_{ii}^{(j)})^{-1}$$

and

$$|G_{ij}^{(k)}| \leq \left| G_{ij} - \frac{G_{ik}G_{kj}}{G_{kk}} \right| \leq \Lambda_o(1 + O(\mathcal{X})), \quad i \neq j,$$

which imply

$$\min_{i \neq k} |G_{ii}^{(k)}| \geq (1 - O(\mathcal{X})) \min |G_{ii}|, \quad \max_{i,j \neq k} |G_{ij}^{(k)}| \leq (1 + O(\mathcal{X})) \max_{i,j} |G_{ij}|. \tag{3.95}$$

Thus we can get the first two inequalities of Lemma 3.19 by induction and the assumption that $p\mathcal{X} \ll 1$.

Now we come to show the third inequality of Lemma 3.19. Note that for $z \in S_r(\tilde{c}, 0)$

$$\begin{aligned} \left| \frac{1}{N} \Im Tr \mathcal{G}^{(\mathbb{T})} - \frac{1}{N} \Im Tr \mathcal{G} \right| &\leq \left| \frac{1}{N} \Im Tr G^{(\mathbb{T})} - \frac{1}{N} \Im Tr G \right| + O(\frac{p\eta}{N}) \\ &\leq \frac{1}{N} \sum_{i \notin \mathbb{T}} |G_{ii}^{(\mathbb{T})} - G_{ii}| + O(\frac{p}{N}). \end{aligned}$$

Now by (3.95) and (iii) of Lemma 3.1, it is not difficult to see that

$$|G_{ii}^{(\mathbb{T})} - G_{ii}| \leq pK^2\Psi^2.$$

Therefore, we conclude the proof. □

PROOF OF LEMMA 3.20. Similarly, it suffices to prove the result for any fixed $z \in S_r(\tilde{c}, L)$. By using Lemma 3.4 and (iv) of Lemma 2.3 we have that

$$\mathbf{r}_i^* \mathcal{G}^{(\mathbb{T} \cup \{i,j\})}(z) \Sigma M_0(z) \mathbf{r}_j^* \leq K \frac{1}{N} ||\mathcal{G}^{(\mathbb{T} \cup \{i,j\})}(z)||_{HS} = K \sqrt{\frac{\Im Tr \mathcal{G}^{(\mathbb{T} \cup \{i,j\})}(z)}{N^2 \eta}}$$

holds with probability larger than

$$1 - N^C \exp(-C' K^{\tau^{-1}})$$

for some positive constants C, C'. Now by Lemma 3.19 and the assumption on p one has

$$\mathbf{r}_i^* \mathcal{G}^{(\mathbb{T} \cup \{i,j\})}(z) \Sigma M_0(z) \mathbf{r}_j^* \leq K \sqrt{\frac{\Im Tr \mathcal{G}(z)}{N^2 \eta} + \Psi^2(z)} \leq CK\Psi(z), \quad \text{in} \quad \Gamma(z, K) \cap \Xi^c(z).$$

Now note

$$\sharp\{\mathbb{T} \subset \{1, \ldots, N\} : |\mathbb{T}| \leq p\} = O(N^{p+1}).$$

Thus in $\Gamma(z, K) \cap \Xi^c(z)$, Υ holds with probability larger than

$$1 - O(N^{p+C} \exp(-C' K^{\tau^{-1}})) \geq 1 - \exp(-\frac{C'}{2} K^{\tau^{-1}})$$

by the assumption on p. Thus we conclude the proof. □

3.3. *Strong local MP type law around λ_r.* Now we start to prove Theorem 3.2. The proof relies on a bootstrap strategy we have mentioned in Introduction and is similar to that of the null case in [27]. The main difference is that the iterate rate in the bootstrap process for our non-null case is slower than that of the null case in [27]. Roughly speaking, we will show that if for some exponent τ, $\Lambda(z) \le (N\eta)^{-\tau}$ holds up to some logarithmic factor with high probability, then $\Lambda(z) \le (N\eta)^{\frac{-(1+3\tau)}{4}}$ up to some logarithmic factor with high probability.

Proof of Theorem 3.2. We assume $\zeta \ge 1$. By Theorem 3.3, Lemma 3.8 and Lemma 3.20, we know that for any $\zeta > 0$, there exists a positive constant C_ζ such that

$$\Theta_1 \subset \bigcap_{z \in S_r(\tilde{c}, 10C_\zeta)} \left(\Gamma(z, \varphi^{C_\zeta}) \cap \Xi^c(z) \cap \Upsilon(z)\right)$$

holds with $(\zeta + 2(\log 1.3)^{-1})$-high probability. In the sequel, we assume

$$C_\zeta \ge \tau(10\zeta + 20(\log 1.3)^{-1}), \tag{3.96}$$

where $\tau \ge 1$ is the parameter in Lemma 3.4. By Lemma 3.12 we know

$$|\mathcal{D}(m_N)(z)| \le C\varphi^{2C_\zeta}\Psi^2 + |[Y]|, \quad \text{in } \Theta_1.$$

Now let $\Lambda_1 = 1$, thus $\Lambda \le \Lambda_1$ in Θ_1. Set

$$p = p_1 = -\log(1 - \mathbb{P}(\Theta_1))/(\log N)^2$$

in Lemma 3.14. Then we have

$$p_1 = C\varphi^{\zeta + 2(\log 1.3)^{-1}}/(\log N)^2.$$

Recall the parameters K, L used in Lemma 3.14, $L = 10C_\zeta$, $K = \varphi^{C_\zeta}$. Thus it is easy to check (3.77) by (3.96).

By Lemma 3.14, we know for $z \in S_r(\tilde{c}, 10C_\zeta)$ there exists an event Θ_2 such that

$$\Theta_2 \subset \Theta_1, \quad \mathbb{P}(\Theta_2) = 1 - \exp(-p_1)$$

and

$$|[Y]| \le \varphi^{2C_\zeta + 5\zeta + 10(\log 1.3)^{-1}}(\Psi_1^2 + \Lambda_1\Psi_1), \quad \Psi_1 := \sqrt{\frac{\Im m_0 + \Lambda_1}{N\eta}}, \quad \text{in } \Theta_2.$$

Then in Θ_2, we have

$$|\mathcal{D}(m_N)(z)| \le \varphi^{2C_\zeta + 5\zeta + 10(\log 1.3)^{-1}}(\Psi_1^2 + \Lambda_1\Psi_1). \tag{3.97}$$

Note that the r.h.s. of (3.97) is decreasing in η and by (3.96) it is less than $(\log N)^{-8}$. Then by using Lemma 3.13 we have

$$\Lambda(z) \le \Lambda_2(z) := \varphi^{2C_\zeta + 6\zeta + 10(\log 1.3)^{-1}}\left(\frac{\Psi_1^2}{\sqrt{\kappa + \eta + \Psi_1^2}} + \frac{\Lambda_1\Psi_1}{\sqrt{\kappa + \eta + \Lambda_1\Psi_1}}\right). \tag{3.98}$$

Now by using the fact that

$$\Im m_0(z) \le O(\sqrt{\kappa + \eta}), \quad \text{in } S_r(\tilde{c}, 10C_\zeta)$$

one can easily get that

$$\frac{\Psi_1^2}{\sqrt{\kappa+\eta+\Psi_1^2}} \le C\Lambda_1^{1/2}(N\eta)^{-1/2}. \tag{3.99}$$

Moreover, in Θ_2 we have

$$\frac{\Lambda_1\Psi_1}{\sqrt{\kappa+\eta+\Lambda_1\Psi_1}} \quad \le \quad \Lambda_1 \cdot \frac{\Psi_1}{(\kappa+\eta+\Lambda_1\Psi_1)^{1/4}} \cdot \frac{1}{(\Lambda_1\Psi_1)^{1/4}}. \tag{3.100}$$

Note that we have

$$\frac{1}{(\Lambda_1\Psi_1)^{1/4}} \le \frac{1}{(\Lambda_1\sqrt{\frac{\Lambda_1}{N\eta}})^{1/4}}, \tag{3.101}$$

and

$$\frac{\Psi_1^2}{(\kappa+\eta+\Lambda_1\Psi_1)^{1/2}} \le \frac{\Im m_0(z)/(N\eta)}{\sqrt{\kappa+\eta}} + \frac{\Lambda_1/(N\eta)}{(\Lambda_1\sqrt{\frac{\Lambda_1}{N\eta}})^{1/2}} \le C\left((N\eta)^{-1} + \Lambda_1^{1/4}(N\eta)^{-3/4}\right). \tag{3.102}$$

Inserting (3.101) and (3.102) into (3.100) we have

$$\frac{\Lambda_1\Psi_1}{\sqrt{\kappa+\eta+\Lambda_1\Psi_1}} \le C\left(\Lambda_1^{5/8}(N\eta)^{-3/8} + \Lambda_1^{3/4}(N\eta)^{-1/4}\right). \tag{3.103}$$

Combining (3.99) and (3.103) we have for $z \in S_r(\tilde{c}, 10C_\zeta)$, there exists

$$\Lambda \le \Lambda_2 \le \varphi^{2C_\zeta+6\zeta+10(\log 1.3)^{-1}}\left(\Lambda_1^{3/4}(N\eta)^{-1/4}\right) \ll 1.$$

Then iterating this process for $K_0 := \log\log N/\log 1.3$ times, we have

$$\Theta_{K_0} \subset \Theta_{K_0-1}, \quad \mathbb{P}(\Theta_{K_0}) = 1 - \exp(-p_{K_0})$$

with

$$p_{K_0} = -\log[1-\mathbb{P}(\Theta_{K_0-1})]/(\log N)^2 = C\varphi^{\zeta+2(\log 1.3)^{-1}}(\log N)^{-2K_0} \ge \varphi^\zeta$$

and in Θ_{K_0} one has

$$\begin{aligned} \Lambda(z) \le \Lambda_{K_0}(z) \quad &:= \quad \varphi^{2C_\zeta+6\zeta+10(\log 1.3)^{-1}}\Lambda_{K_0-1}^{3/4}(N\eta)^{-1/4} \\ &\le \quad \varphi^{2C_\zeta+6\zeta+10(\log 1.3)^{-1}}(N\eta)^{-1+(3/4)^{K_0}} \\ &\le \quad \varphi^{2C_\zeta+6\zeta+10(\log 1.3)^{-1}+1}(N\eta)^{-1}, \end{aligned}$$

where in the last step we have used the fact that

$$N^{(3/4)^{K_0}} \le \varphi$$

which can be easily verified by the definition of K_0. Thus we complete the proof of (3.11) by adjusting the constant C_ζ in Theorem 3.2. Moreover, by $\Theta_{K_0} \subset \Theta_1$ and the definition of Θ_1 we obtain (3.12). □

4. Convergence rate on the right edge. In this section, we will prove Theorem 1.5. To this end, we need to translate the information on $m_N(z)$ to the eigenvalues around λ_r. Actually, we will turn to prove the result for the truncated matrices satisfying condition 1.5. In other words, we will verify the following slight modification of Theorem 1.5.

THEOREM 4.1 (Convergence rate around λ_r for truncated matrix). *Under Condition 1.1 and (3.4), for any $\zeta > 0$ there exists a constant C_ζ such that the following events hold with ζ-high probability.*

(i): For the largest eigenvalue $\lambda_1(\mathcal{W})$, there exists

$$|\lambda_1(\mathcal{W}) - \lambda_r| \le N^{-2/3}\varphi^{C_\zeta}.$$

(ii): For any

$$E_1, E_2 \in [\lambda_r - \tilde{c}, C_r],$$

there exists

$$|(F_N(E_1) - F_N(E_2)) - (F_0(E_1) - F_0(E_2))| \le \frac{C(\log N)\varphi^{C_\zeta}}{N}. \tag{4.1}$$

PROOF OF THEOREM 1.5 ASSUMING THEOREM 4.1. Note that by (2.1), (3.2) and (3.3) we can easily recover the results from the truncated matrix to the original one with overwhelming probability. □

Therefore, it suffices to prove Theorem 4.1 in the sequel.

PROOF OF THEOREM 4.1. Relying on the strong local MP type law, the proof is analogous to the counterparts in [21] and [27]. Note that by Lemma 2.1, it is easy to see

$$1 - F_0(\lambda_r - \varphi^{C_\zeta}N^{-2/3}) \sim \frac{\varphi^{\frac{3}{2}C_\zeta}}{N}.$$

Then if (ii) of Theorem 4.1 holds,

$$\lambda_1(W) \ge \lambda_r - N^{-2/3}\varphi^{C_\zeta}$$

holds with ζ-high probability. Thus it suffices to prove that (ii) and

$$\lambda_1(W) \le \lambda_r + N^{-2/3}\varphi^{C_\zeta} \tag{4.2}$$

hold with ζ-high probability.

At first we come to verify (4.2). We recall the fact that

$$C_l \le \lambda_1(W) \le C_r$$

holds with ζ-high probability.

Therefore, to show (4.2), it suffices to prove for any $\zeta > 0$, there exists some $C_\zeta > 0$ such that

$$\max\{\lambda_j(\mathcal{W}) : \lambda_j(\mathcal{W}) \le C_r\} \le \lambda_r + N^{-2/3}\varphi^{C_\zeta} \tag{4.3}$$

with ζ-high probability. To this end, we recall the iteration process in (3.97) and (3.98). In the sequel, we set $C_\zeta = 5D_\zeta$. By (i) of Theorem 3.2, we know that with ζ-high probability,

$$|\mathcal{D}(m)(z)| \le \varphi^{D_\zeta}\left(\frac{\Im m_0 + \frac{1}{N\eta}}{N\eta} + \frac{1}{N\eta}\sqrt{\frac{\Im m_0 + \frac{1}{N\eta}}{N\eta}}\right) \le C\varphi^{D_\zeta}\frac{\Im m_0 + \frac{1}{N\eta}}{N\eta}$$

for some constants $D_\zeta, C > 0$. Thus we have

$$\Lambda(z) \le C(\log N)\varphi^{D_\zeta} \frac{\delta}{\sqrt{\kappa+\eta+\delta}}, \quad \delta := \frac{\Im m_0 + \frac{1}{N\eta}}{N\eta}.$$

Now for any $E \ge \lambda_r + N^{-2/3}\varphi^{5D_\zeta}$, we choose

$$\eta := \varphi^{-D_\zeta} N^{-1/2} \kappa^{1/4}.$$

Note here

$$\kappa = E - \lambda_r \ge N^{-2/3}\varphi^{5D_\zeta}.$$

It is easy to check the following relations,

$$\kappa \gg \varphi^{D_\zeta}\eta, \quad \frac{\sqrt{\kappa}}{N\eta^2} \gg 1. \tag{4.4}$$

By using (ii) of Lemma 2.3, we have

$$\Im m_0(z) = C\frac{\eta}{\sqrt{\kappa}}.$$

Thus by the last inequality of (4.4) we have

$$\Im m_0(z) \ll \frac{1}{N\eta} \tag{4.5}$$

and thus

$$\delta \le \frac{C}{N\sqrt{\kappa}} + (N\eta)^{-2}.$$

Then we can get

$$\Lambda(z) \le C\varphi^{D_\zeta}\left(\frac{\eta}{\kappa} + \frac{1}{N\eta\sqrt{\kappa}}\right)\frac{1}{N\eta} \ll \frac{1}{N\eta}, \tag{4.6}$$

where we have used the fact that

$$N\eta\sqrt{\kappa} \ge C\varphi^{5D_\zeta/2}.$$

Then (4.5) and (4.6) together imply

$$\Im m_N(z) \ll \frac{1}{N\eta}.$$

Now by the basic relation

$$N(E-\eta, E+\eta) \le CN\eta\Im m_N(z) \ll 1$$

we know there is no eigenvalue in $[E-\eta, E+\eta]$. Thus (4.3) is verified, so (4.2) follows.

For (ii), with the aid of Theorem 3.2, the proof of (4.1) is just the same as the proof of (8.6) in [27]. The main strategy is to translate the closeness between m_N and m_0 to that of the spectral distributions (F_N and F_0). Such a strategy is independent of the matrix model. Actually, we only need to set the interval $[A_1, A_2] = [\lambda_r - \tilde{c}, C_r]$ in Lemma 8.1 of [27] and the remaining part of proof is totally the same as the counterpart in [27]. Thus here we do not reproduce the details.

Therefore, we complete the proof. □

5. A corollary of Theorems 3.2 and 4.1. Below we will provide a corollary of Theorems 3.2 and 4.1 which will be used in our subsequent work [7]. Such a corollary can establish an approximate equivalence between the distribution function of the largest eigenvalue and the expectation of a functional of the Stieltjes transform. This equivalence will be crucial in the so called *Green function comparison procedure* used in [7].

By using the square root behavior of $\rho_0(x)$ in Lemma 2.1 again, it is easy to see for any positive constant C_ζ,

$$1 - F_0(\lambda_r - 2\varphi^{C_\zeta} N^{-2/3}) \sim \frac{\varphi^{\frac{3}{2}C_\zeta}}{N}$$

when N is sufficiently large. Together with (4.1) we immediately get that

$$F_N(\lambda_r + 2\varphi^{C_\zeta} N^{-2/3}) - F_N(\lambda_r - 2\varphi^{C_\zeta} N^{-2/3}) \leq \frac{\varphi^{2C_\zeta}}{N} \tag{5.1}$$

with ζ-high probability. Moreover by Theorem 4.1, we also have

$$|\lambda_1(\mathcal{W}) - \lambda_r| \leq \varphi^{C_\zeta} N^{-2/3} \tag{5.2}$$

with ζ-high probability. Thus to show Theorem 1.4, it suffices to assume that

$$-\varphi^{C_\zeta} \leq s \leq \varphi^{C_\zeta}.$$

Now set

$$E_\zeta := \lambda_r + 2\varphi^{C_\zeta} N^{-2/3}.$$

For any $E \leq E_\zeta$ we denote

$$\chi_E := \mathbf{1}_{[E, E_\zeta]}.$$

For ease of presentation, we denote

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

and the number of eigenvalues of W_N in an interval $[E_1, E_2]$ by

$$\mathcal{N}(E_1, E_2) := N(F_N(E_2) - F_N(E_1)).$$

By definition we have

$$\mathcal{N}(E, E_\zeta) = Tr\chi_E(W_N).$$

Observe that

$$Tr\chi_{E-l} * \theta_\eta(W_N) = N\frac{1}{\pi}\int_{E-l}^{E_\zeta} \Im m_N(y + i\eta)dy,$$

which can be viewed as a smoothed version of the counting function $Tr\chi_E(W_N)$ on scale η. An obvious advantage of $Tr\chi_{E-l} * \theta_\eta(W_N)$ is that it can be represented in terms of the Stieltjes transform. The following lemma claim that we can replace $Tr\chi_E(W_N)$ by its smoothed approximation $Tr\chi_{E-l} * \theta_\eta(W_N)$.

Corollary 5.1. *Let* $\eta_1 = N^{-2/3-9\varepsilon}, l = \frac{1}{2}N^{-2/3-\varepsilon}$ *and* $h(x)$ *be a smooth cut-off function satisfying*

$$h(x) = 1 \quad \text{if} \quad |x| \le 1/9, \quad h(x) = 0 \quad \text{if} \quad |x| \ge 2/9, \quad h'(x) \le 0 \quad \text{if} \quad x \ge 0.$$

Then we have

$$\mathbb{E}h\left(Tr\chi_{E-l} * \theta_{\eta_1}(W_N)\right) \le \mathbb{P}\left(\lambda_1(W_N) \le E\right) \le \mathbb{E}h\left(Tr\chi_{E+l} * \theta_{\eta_1}(W_N)\right) + O(\exp(-\varphi^{C_\zeta}))$$

when N *is sufficiently large.*

At first, we need to prove the following lemma.

Lemma 5.2. *Let* $\eta_1 = N^{-2/3-9\varepsilon}$ *and* $l_1 = N^{-2/3-3\varepsilon}$ *for any* $\varepsilon > 0$. *If* E *satisfies*

$$|E - \lambda_r| \le \frac{3}{2}\varphi^{C_\zeta}N^{-2/3},$$

we have

$$|Tr\chi_E(W_N) - Tr\chi_E * \theta_{\eta_1}(W_N)| \le C\left(N^{-2\varepsilon} + \mathcal{N}(E - l_1, E + l_1)\right)$$

holds with ζ*-high probability.*

Proof. The proof is similar to that of Lemma 4.1 of [27]. Thus we will just sketch it below. By the assumption on E, (5.1) and (5.2), we can use (4.9) of [27], that is

$$\begin{aligned} |Tr\chi_E(W_N) - \chi_E * \theta_{\eta_1}(W_N)| &\le C(\mathcal{N}(E - l_1, E + l_1) + N^{-5\varepsilon}) \\ &+ CN\eta_1(E_\zeta - E)\int_{\mathbb{R}} \frac{1}{y^2 + l_1^2}\Im m_N(E - y + il_1)dy. \end{aligned}$$

By the bounds for $\lambda_1(\mathcal{W})$ in (1.12), we know

$$\int_{E-y\ge C_r} \frac{1}{y^2 + l_1^2}\Im m_N(E - y + il_1)dy = O(1)$$

with ζ-high probability. Now we further split the interval $[-\infty, C_r]$ into $(-\infty, \lambda_r - \tilde{c})$ and $(\lambda_r - \tilde{c}, C_r]$. When $E - y \in [\lambda_r - \tilde{c}, C_r]$, we have

$$|\Im m_N(E - y + il_1) - \Im m_0(E - y + il_1)| \le \frac{\varphi^{C_\zeta}}{Nl_1}$$

with ζ-high probability. When $E - y \in (-\infty, \lambda_r - \tilde{c})$, by assumption on E we have $y \ge \tilde{c}/2$. Thus we have

$$\begin{aligned} &\int_{-\infty<E-y\le C_r} \frac{1}{y^2 + l_1^2}\Im m_N(E - y + il_1)dy \\ = &\int_{\lambda_r-\tilde{c}\le E-y\le C_r} \frac{1}{y^2 + l_1^2}\Im m_N(E - y + il_1)dy + O(1) \end{aligned}$$

Moreover, it is not difficult to get

$$\begin{aligned} &\int_{\lambda_r-\tilde{c}\le E-y\le C_r} \frac{1}{y^2 + l_1^2}\Im m_N(E - y + il_1)dy \\ \le &\int_{\lambda_r-\tilde{c}\le E-y\le C_r} \frac{1}{y^2 + l_1^2}\Im m_0(E - y + il_1)dy + \frac{\varphi^{C_\zeta}}{Nl_1}\int_{\lambda_r-\tilde{c}\le E-y\le C_r} \frac{1}{y^2 + l_1^2}dy \end{aligned}$$

$$\leq C\int_{\lambda_r-\tilde{c}\leq E-y\leq C_r}\frac{1}{y^2+l_1^2}\sqrt{|E-y-\lambda_r|+l_1}dy+C\frac{\varphi^{C_\zeta}}{Nl_1^2}$$

with ζ-high probability. Then by the assumption on E and the definitions of η_1 and l_1, it is not difficult to obtain

$$N\eta_1(E_\zeta-E)\int_{\mathbb{R}}\frac{1}{y^2+l_1^2}\Im m_N(E-y+il_1)dy\leq N^{-2\varepsilon}$$

through elementary calculations. Thus we conclude the proof. □

Proof of Corollary 5.1. With the aid of Lemma 5.2, the proof of Corollary 5.1 is nearly the same to that of Corollary 4.2 of [27] for the null case. Hence we do not reproduce the details here. □

Acknowledgements. The authors are very grateful to Professors Laszlo Erdös, Horng-Tzer Yau and Jun Yin for useful comments on the former version of this paper which helped to improve the organization significantly.

References.

[1] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. Ann. Probab. **21**, 625648. (1993)
[2] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. Statist. Sinica **9**, 611-677.
[3] Z. D. Bai, J. Hu, and W. Zhou. Convergence rates to the Marcenko-Pastur type distribution. Stochastic Processes and their Applications. **122**, 68-92. (2012)
[4] Z. D. Bai, B. Miao and J. Tsay. Convergence rates of the spectral distributions of large Wigner matrices. Int. Math. J. **1**, 6590. (2002)
[5] Z. D. Bai, J. Silverstein. Spectral analysis of large dimensional random matrices. Science Press, Beijing. (2006)
[6] Z. D. Bai, Y.Q. Yin, and P.R. Krishnaiah. On the limiting empirical distribution of product of two random matrices when the underlying distribution is isotropic. J. Multiv. Ana., **19**, 189-200. (1986)
[7] Z. G. Bao, G. M. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. Preprint.
[8] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. J. Multiv. Anal., **12**: 1-38. (1982)
[9] N. El Karoui, Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. Ann. Probab. **35**, No. 2, 663-714. (2007)
[10] L. Erdös. Universality of Wigner random matrices: a survey of recent results. Russ. Math. Surv. **66**(3), 507. (2011)
[11] L. Erdös, B. Farrell, Local eigenvalue density for general MANOVA matries. arXiv:1207.0031v2.
[12] L. Erdös, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdös-Renyi graphs I: Local Semicircle law, arXiv: 1103. 1919. (2011)
[13] L. Erdös, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdös-Renyi graphs II: eigenvalue spacing and the extreme eigenvalues. Communications in Mathematical Physics, **314**(3), 587-640. (2012)
[14] L. Erdös, B. Schlein, H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. Ann. Probab. **37**(3), 815-852. (2009)
[15] L. Erdös, B. Schlein, H.-T. Yau. Local Semicircle Law and Complete Delocalization for Wigner Random Matrices. Communications in Mathematical Physics, **287**(2), 641-655. (2009)
[16] L. Erdös, B. Schlein, H.-T. Yau. Wegner Estimate and Level Repulsion for Wigner Random Matrices. Int. Math. Res. Notices, 2010(3): 436-479. (2010)
[17] L. Erdös, B. Schlein, H.-T. Yau. Universality of random matrices and local relaxation flow. Inventiones mathematicae, **185**(1), 75-119. (2011)
[18] L. Erdös, B. Schlein, H.-T. Yau, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. Ann. Inst. H. Poincaré Probab. Statist. 48(1), 1-46. (2012)
[19] L. Erdös, H.-T. Yau, and J. Yin. Universality for generalized Wigner matrices with Bernoulli distribution. arXiv: 1003.3813v7. (2011)
[20] L. Erdös, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. Probability Theory and Related Fields, **154**, Issue 1-2, 341-407. (2012)
[21] L. Erdös, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. Advances in Mathematics, **229**, Issue 3, 14351515. (2012)

[22] F. Götze and A. N. Tikhomirov. Rate of convergence to the semi-circular law. Probab. Theory Relat. Fields **127**, 228276. (2003)
[23] F. Götze and A. N. Tikhomirov. Rate of convergence in probability to the Marchenko-Pastur law. Bernoulli **10**(3), 503548. (2004)
[24] P. L. Hsu. On the distribution of roots of certain determinantal equations. Ann. Eugenics, **9**, 250-258. (1939)
[25] B. Y. Jing, G. M. Pan, Q. M. Shao, and W. Zhou. Nonparametric estimate of spectral density functions of sample covariance matrices: A first step. Ann. Statist. **38**(6) ,3724-3750. (2010)
[26] V. A. Marčenko, L.A. Pastur. Distribution for some sets of random matrices. Math. USSR-Sb., **1**, 457-483. (1967)
[27] N. S. Pillai, J. Yin, Universality of covariance matrices. arXiv:1110.2501v6.
[28] J. W. Silverstein, S. I. Choi, Analysis of the limiting spectral distribution of large dimensional random matrices. J. Multivariate Anal. **54**, 175-192. (1995)
[29] T. Tao, V. Vu. Random Matrices: Universality of Local Eigenvalue Statistics up to the Edge. Communications in Mathematical Physics, **298**(2), 549-572. (2010)
[30] T. Tao, V. Vu. Random matrices: Universality of local eigenvalue statistics. Acta Mathematica, **206**(1), 127-204. (2011)
[31] T. Tao, V. Vu. Random covariance matrices: Universality of local statistics of eigenvalues. Ann. Probab. **40**(3), 1285-1315. (2012)
[32] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. J. Multiv. Anal., **20**, 50-68. (1986)
[33] K. Wang. Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. Random matrices: Theory and Applications, **1**(1), 1150005. (2012)

Department of Mathematics,
Zhejiang University, P. R. China
E-mail: zhigangbao@zju.edu.cn

Division of Mathematical Sciences,
School of Physical and Mathematical Sciences,
Nanyang Technological University,
Singapore 637371
E-mail: gmpan@ntu.edu.sg

Department of Statistics and Applied Probability,
National University of Singapore,
Singapore 117546
E-mail: stazw@nus.edu.sg
URL: http://www.sta.nus.edu.sg/ stazw/