LOCAL MARCHENKO-PASTUR LAW AT THE HARD EDGE OF THE SAMPLE COVARIANCE ENSEMBLE

ANASTASIS KAFETZOPOULOS AND ANNA MALTSEV

School of Mathematical Sciences, Queen Mary University of London, Mile End Road, E1 4NS

Abstract. Consider an N by N matrix X of complex entries with iid real and imaginary parts. We show that the local density of eigenvalues of X^*X converges to the Marchenko-Pastur law on the optimal scale with probability 1. We also obtain rigidity of the eigenvalues in the bulk and near both hard and soft edges. Here we avoid logarithmic and polynomial corrections by working directly with high powers of expectation of the Stieltjes transforms. We work under the assumption that the entries have a finite 4th moment and are truncated at $N^{1/4}$, or alternatively with exploding moments. In this work we simplify and adapt the methods from prior papers of Götze-Tikhomirov and Cacciapuoti-Maltsev-Schlein to covariance matrices.

1. Introduction

In this paper we obtain optimal large deviation bounds on the Stieltjes transform for the sample covariance random matrix ensemble. Let X be a $M \times N$ matrix with components $x_{ij} = \text{Re } x_{ij} + i \,\text{Im } x_{ij}$. Assume that Re x_{ij} and Im x_{ij} are independent identically distributed (iid) real random variables with mean zero and variance $\frac{1}{2}$ so that

$$\mathbb{E}x_{ij} = 0 \quad \text{and} \quad \mathbb{E}|x_{ij}|^2 = 1 \qquad i = 1, ..., N,\, j = 1, \ldots, M\,. \tag{1.1}$$

and

$$d := M/N.$$

In what follows we shall denote by X_N the scaled matrix

$$X_N = X/\sqrt{N}. \tag{1.2}$$

We are interested in the analysis of the asymptotic empirical spectral measure of the matrix $X_N^* X_N$ for $N \to \infty$, when $M = N$. This is the case when the limiting measure has a one over square root singularity near 0 with typical distance between eigenvalues on the order of $\frac{1}{N^2}$. We are able to obtain results on the hard edge, the bulk, and the soft edge in a unified way.

Let s_α, $\alpha = 1, ..., N$, be the eigenvalues of $X_N^* X_N$. Since $X_N^* X_N$ is Hermitian and positive definite we can assume that $0 \le s_1 \le s_2 \le ... \le s_N$. We denote by n_N the empirical spectral distribution s_α,

$$n_N(E) = \frac{1}{N} \#\{\alpha \le N \mid s_\alpha \le E\} \tag{1.3}$$

and

$$\mathcal{N}(I) = \#\{\alpha \le N \mid s_\alpha \in I\} \tag{1.4}$$

For any $\theta \in \mathbb{C}$ with Im $\theta \neq 0$ we define the Stieltjes transform of n_N as

$$\Delta_N(\theta) = \int_{\mathbb{R}} \frac{1}{x - \theta} dn_N(x) = \frac{1}{N} \operatorname{Tr}(X_N^* X_N - \theta)^{-1} = \frac{1}{N} \sum_{\alpha=1}^{N} \frac{1}{s_\alpha - \theta}\,. \tag{1.5}$$

We denote by ν the probability distribution of Re x_{ij} and Im x_{ij}. In this paper we assume that

$$\sup_{N \ge 1} \sup_{1 \le j,k \le N} \mathbb{E}|x_{jk}|^4 =: \mu_4 < \infty, \tag{1.6}$$

E-mail address: `anastasis.kafetzopoulos@gmail.com, a.maltsev@qmul.ac.uk`.
Date: August 6, 2023.

and that there exists a constant $D > 0$ such that for all N:

$$\sup_{1\le j,k\le N} |x_{jk}| \le DN^{1/4}. \tag{1.7}$$

These assumptions are the same as in the papers of Götze-Tikhomirov [6, 5], and with easy modifications all the proofs and results hold as well for x_{ij} such that $\mathbb{E}|x_{ij}|^q \le (Cq)^{cq}$ for universal constants C, c. Alternatively, it is sufficient to assume that

$$\mathbb{E}|x_{jk}|^{4q} \le D^{4q-4}N^{q-1}\mu_4 \tag{1.8}$$

for some constants μ_4 and D, and all $q \in \mathbb{N}$.

The first results about universality of covariance matrices date back to 1967. Let

$$\lambda_\pm = (1 \pm \sqrt{d'})^2,$$

Marchenko Pastur in [11] show that $d\nu_N \to \rho$ weakly with probability 1, where ρ is the Marchenko-Pastur distribution, given by

$$\rho_{MP}(E) = \frac{1}{2\pi}\sqrt{\frac{(\lambda_+ - E)(E-\lambda_-)}{E^2}}, \tag{1.9}$$

whenever $E \in [\lambda_-, \lambda_+]$ and 0 otherwise. In the case of a square matrix X, the density of the Marchenko-Pastur distribution is

$$\rho(E) = \begin{cases} \frac{1}{2\pi}\sqrt{\frac{4}{E} - 1} & 0 < E \le 4 \\ 0 & \text{otherwise} \end{cases} \tag{1.10}$$

and for any θ such that Im $\theta \neq 0$ we denote by Δ the associated Stieltjes transform

$$\Delta(\theta) = \int_{\mathbb{R}} \frac{1}{x-\theta}\rho(x)dx \tag{1.11}$$

which satisfies the quadratic equation

$$\Delta = -\frac{1}{\theta(\Delta+1)}. \tag{1.12}$$

In [11], the convergence of the density of states is on intervals whose sizes are independent of N. In this case, the intervals that are away from the endpoints contain an order of N eigenvalues. A natural question to study is whether the convergence remains on intervals whose size (we call the interval size scale) goes to zero as N grows.

In [3], Erdös-Schlein-Yau-Yin establish convergence of the empirical spectral density for general covariance matrices to the Marchenko-Pastur law in the bulk for $d < 1$ on small intervals. They use a decomposition by minors for the diagonal elements of the resolvent to establish a self-consistent equation for the Stieltjes transform Δ_N of $d\nu_N$. Large deviation estimates and a continuity argument are then used show the convergence of the spectral measure on small intervals (involving polynomial corrections) in the bulk distribution. These methods have been extended to the hard edge and logarithmic rather than polynomial corrections by Cacciapuoti-Maltsev-Schlein in [1]. More precisely, the authors show that the fluctuation of the Stieltjes transform $\sqrt{E}\Delta_N$ away from $\sqrt{E}\Delta$ is on the order of $\sqrt{\frac{\sqrt{E}}{N\eta}}$ and they obtain convergence of the counting function of eigenvalues everywhere including close to the hard edge. Eigenvalue rigidity with polynomial corrections for the bulk and soft edges for entries with subexponential decay can be found in Pillai-Yin [12].

A related question is that of the universality of the correlation function of the eigenvalues. Results in the bulk using local laws and a local relaxation flow can be found in [3, 12]. A similar result in [16] by proving a version of the four moment theorem for random covariance matrices for any $0 < d \le 1$ in the bulk of the spectrum. Wang [17] extends these results to the soft edge (cf Remark 1.8 in [17]). For the hard edge, universality of the joint distribution of low-lying eigenvalues has been established by Tao-Vu in [15]. Another related question is about the rate of convergence of the density of states to the Marchenko-Pastur law. In [6], the authors establish that the Kolmogorov distance between the expected spectral measure and the Marchenko-Pastur law is $O(N^{-1})$. Additionally, there has been some remarkable progress on similar questions in the case of Wigner (matrices with i.i.d. entries up to Hermitian symmetry) and more general

Wigner-type matrices [10, 9]. The authors use homogenization theory, which relies on coupling two Dyson Brownian motions, to establish the Gaussianity of fluctuation of individual eigenvalues in the bulk of the spectrum.

In this paper we obtain optimal bounds on the expectations of high moments of the fluctuation $\Lambda = \Delta_N - \Delta$ on the optimal scale. Our methods and results apply to the bulk as well as the soft and hard edges. The main objective of this work is to extend the results and methodology of [2] to a hard edge setup. We were able to simplify the proof of Theorem 1 in [2] avoiding different cases for the bulk and edges. Unlike in the Wigner case, where both edges are soft, the presence of the hard edge at 0 allows us to extend the bounds on the real part of the Stieltjes transform to the negative real line, thus also yielding a fluctuation for the individual eigenvalue near the hard edge that is decreasing with the eigenvalue number. This paper also improves on [1] by removing the logarithmic corrections and improving the fluctuation bounds. We also extended the proofs in [5, 6] on fluctuations of quadratic forms to a soft edge setup by improving a factor of $|\Delta|$ to a factor of $\operatorname{Im}\Delta$.

To state our theorem we define the domain $S_{E,\eta}$ where we obtain our bounds:

$$S_{E,\eta} := \{4\eta > c(E^2 + \eta^2 - 4E)\} \tag{1.13}$$

for some $c > 0$. This domain is chosen so that $\operatorname{Im}(\Delta + 1/2)^2 \geq c \operatorname{Re}((\Delta + 1/2)^2)$ which we need for the proof of Proposition 3.2. While all the proofs work for all $c > 0$ not dependent on N, we will specifically work with $c = 1$ to allow us the opportunity to illustrate it the following picture, Figure 1.

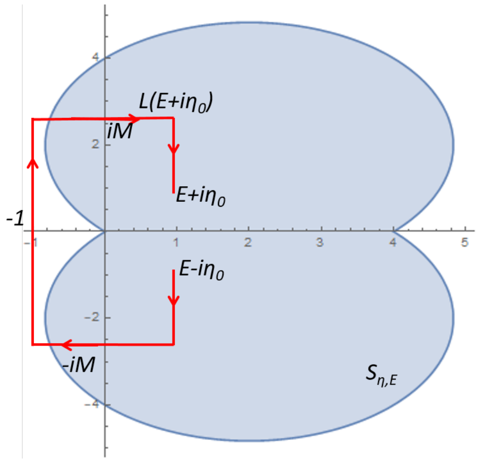

FIGURE 1. The set $S_{\eta,E}$ shaded in blue and the integration contour $L(z_0)$ from (7.1) in red

Theorem 1. *Let X_N be a $N \times N$ matrix as described in equation (1.2), and assume (1.6) and (1.7). Let Δ_N and Δ be the Stieltjes transforms defined in equations (1.5) and (1.11). Moreover set $\theta = E + i\eta$, with $\frac{N\eta}{|\sqrt{\theta}|} \geq M$ for some suitably large M. Then there exist positive constants c_0, C such that for each $K > 0$ and $1 \leq q \leq c_0 \left(\frac{N\eta}{|\sqrt{\theta}|}\right)^{1/8}$ and $\theta \in S_{E,\eta}$ or $E < 0$*

$$\mathbb{P}\left(|\Delta_N(\theta) - \Delta(\theta)| \geq \frac{K}{N\eta}\right) \leq \frac{(Cq)^{cq^2}}{K^q} \tag{1.14}$$

Furthermore, for any $E \in \mathbb{R}$ and $\eta > 0$ such that $\frac{N\eta}{|\sqrt{\theta}|} \geq M$ we have that

$$\mathbb{P}\left(|\operatorname{Im}(\Delta_N(\theta) - \Delta(\theta))| \geq \frac{K}{N\eta}\right) \leq \frac{(Cq)^{cq^2}}{K^q}. \tag{1.15}$$

Remark 1.1. We notice that the particular an upper bound on q given in Theorems 1, 2, and 3 is used for the proof of Lemma 5.1, specifically in equation (5.11), then it gets halved in Proposition 6.4. The power $1/4$ in Lemma 5.1 could be relaxed further if desired to any finite power. However, we notice that if q gets large with N, the $(Cq)^{cq^2}$ in the numerator of (1.14) and (1.15) renders the bound trivial for even modestly large values of $N\eta$ such as a power of a $\log N$. Our results are best used for short scales and very small intervals, smaller than $\eta \leq (\log N)^c$ for some c, or q should remain reasonably small such as $\log \log N$. Other papers that use a bootstrapping argument in probability rather than in expectation such as [1] cover the regimes with logarithmic corrections, obtaining better results.

We then use our Theorem 1 to obtain fluctuation estimates on the counting function as stated in the next theorem. Letting

$$n_{MP}(E) = \int_0^E \rho(x)dx, \tag{1.16}$$

we compare it to n_N.

Theorem 2. *With assumptions and M as in Theorem 1, there exist constants $M_0, N_0, C, c > 0$ such that for any $K > 0$ and $E \geq \frac{M_0}{N^2}$*

$$\mathbb{P}\left(|n_N(E) - n_{MP}(E)| \geq K \min\left\{\sqrt{E}, \frac{\log N}{N}\right\}\right) \leq \frac{(Cq)^{cq^2}}{K^q} \tag{1.17}$$

for all $E \in \mathbb{R}$, $K > 0, N > N_0$, $q \leq M$.

We use the above estimate to obtain rigidity estimates that is how far each eigenvalue can fluctuate away from its classical location. We define the classical locations of the eigenvalues, predicted by the Marchenko-Pastur distribution, as the points γ_a, $(a = 1, ..., N)$ such that

$$\int_0^{\gamma_a} \rho(E)dE = \frac{a}{N}.$$

In particular, we obtain the fluctuation of eigenvalues near the hard edge to be of the order of $\frac{\log N}{N^2}$. The fluctuations of eigenvalues in the bulk and soft edges of both the Gaussian Unitary Ensemble and the Wishart Ensemble are known to be respectively of the order $\frac{\sqrt{\log N}}{N}$ in the bulk and $\frac{\sqrt{\log k}}{k^{1/3}N^{2/3}}$ for the kth eigenvalue from the edge, $k \to \infty$ (see [8, 14]). To our knowledge similar results are not yet available for the hard edge.

Theorem 3. *With assumptions and M as in Theorem 1, there exist constants $C, c, N_0, \epsilon > 0$ such that*

$$\mathbb{P}\left(|\lambda_a - \gamma_a| \geq K\frac{\log N}{N}\left(\frac{a}{N}\right)\right) \leq \frac{(Cq)^{cq^2}}{K^q} \tag{1.18}$$

for $a = 1, ..., \lceil N/2 \rceil$, $N > N_0$, $K > 0$, and any $q \leq M$. Furthermore, for $a \leq \log N$ we have that

$$\mathbb{P}\left(|\lambda_a - \gamma_a| \geq K\frac{a^2}{N^2}\right) \leq \frac{(Cq)^{cq^2}}{K^{q/2}}. \tag{1.19}$$

In this theorem the factor $\frac{a}{N}$ accounts for the higher density at the hard edge. Here we focus on hard-edge rigidity, since proofs of soft-edge rigidity require control of the largest eigenvalue which, to our knowledge, is not currently available in the case of truncated entries with four moments, in either Wigner or Sample Covariance case.

2. Useful Identities

In this section we collect some useful known identities. Let $\mathbb{J}, \mathbb{J}_1, \mathbb{J}_2 \subset \{1, ..., N\}$. We will denote by $X^{(\mathbb{J})}$ the submatrix of X_N with columns of indices $\mathbb{J}$ removed, and $X_{(\mathbb{J})}$ with rows of indices $\mathbb{J}$ removed.
We define the resolvent matrices

$$G^{(\mathbb{J}_1)}_{(\mathbb{J}_2)} := \left((X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)})^* X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)} - \theta\right)^{-1} \quad \text{and} \quad \mathcal{G}^{(\mathbb{J}_1)}_{(\mathbb{J}_2)} := \left(X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)} (X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)})^* - \theta\right)^{-1}. \tag{2.1}$$

When our arguments work for any $\mathbb{J}_1, \mathbb{J}_2$ we will mention this and then suppress them for ease of notation, and we will write $G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),ij}$ for the ijth element. We notice here that $G^{(\mathbb{J})}$ is the minor of $G := G^{(\emptyset)}$ with $\mathbb{J}$-th rows and $\mathbb{J}$-th columns removed. Lastly we notice that

$$\operatorname{Tr} \mathcal{G}^{(\mathbb{J}_1)}_{(\mathbb{J}_2)} = \frac{|\mathbb{J}_1| - |\mathbb{J}_2|}{\theta} + \operatorname{Tr} G^{(\mathbb{J}_1)}_{(\mathbb{J}_2)} \tag{2.2}$$

Similarly we introduce

$$\Delta_{N,(\mathbb{J}_2)}^{(\mathbb{J}_1)} := \operatorname{Tr} G_{(\mathbb{J}_2)}^{(\mathbb{J}_1)} \quad \text{and} \quad \Lambda_{(\mathbb{J}_2)}^{(\mathbb{J}_1)} := \Delta_{N,(\mathbb{J}_2)}^{(\mathbb{J}_1)} - \Delta \tag{2.3}$$

and we use Δ_N and Λ when $\mathbb{J}_1, \mathbb{J}_2 = \emptyset$. We will use $\mathbf{x}_k$ and $\mathbf{x}^k$ for rows and columns of $\sqrt{N} X_N$ respectively. We state some well-known identities for resolvent entries (Lemma 2.3 of [12]).

Lemma 2.1. *With $G_{(\mathbb{J}_2)}^{(\mathbb{J}_1)}$ as before for $i, j \neq k$, we have*

$$G_{(\mathbb{J}_2),ij}^{(\mathbb{J}_1)} = G_{(\mathbb{J}_2),ij}^{(\mathbb{J}_1 \cup \{k\})} + \frac{G_{(\mathbb{J}_2),ik}^{(\mathbb{J}_1)} G_{(\mathbb{J}_2),ki}^{(\mathbb{J}_1)}}{G_{(\mathbb{J}_2),kk}^{(\mathbb{J}_1)}}. \tag{2.4}$$

Furthermore, as seen for example in (3.2) of [2], we have the following relationship between the (k,k) element of G^2 and Im G_{kk}, and the same holds for $G_{(\mathbb{J}_2)}^{(\mathbb{J}_1)}, \mathcal{G}_{(\mathbb{J}_2)}^{(\mathbb{J}_1)}$:

$$|G(z)|_{kk}^2 = \frac{(\text{Im } G(z))_{kk}}{\eta} \tag{2.5}$$

yielding that

$$|(G^2)_{kk}| \leq \frac{\text{Im } G_{kk}}{\eta}. \tag{2.6}$$

Next we observe that using the proof of (3.10) in [2] we can also obtain the following for the resolvent of the sample covariance ensemble, and the proof works for $G_{(\mathbb{J}_2)}^{(\mathbb{J}_1)}, \mathcal{G}_{(\mathbb{J}_2)}^{(\mathbb{J}_1)}$ for any $\mathbb{J}_1, \mathbb{J}_2$:

Lemma 2.2. *With G and $\mathcal{G}$ as before, we have that*

$$G_{11}(E + i\eta/s) \leq s G_{11}(E + i\eta). \tag{2.7}$$

Furthermore we have the following bounds on the Stieltjes transform of the Marchenko-Pastur law. For $E > 0$ we set $\kappa := |E - 4|$. For any fixed $E_0', E_0 > 0$ and $\eta_0 > 0$ there exist constants $C > 0$ such that

$$\left| \Delta + \frac{1}{2} \right| \geq C(\kappa^2 + \eta^2)^{\frac{1}{4}} \geq C\sqrt{\kappa + \eta}, \tag{2.8}$$

and

$$c \frac{\eta}{\sqrt{\kappa + \eta}} \leq \text{Im}\Delta \leq C \frac{\eta}{\sqrt{\kappa + \eta}}, \tag{2.9}$$

$\forall E_0 \leq E \leq E_0',\ 0 < \eta \leq \eta_0,\ \kappa \geq \eta.$

3. Equations for Λ

Lemma 3.1. *Take $\theta = E + i\eta$. For any $N \geq N_0$ one has*

$$\Delta_{N,(\mathbb{J}_2)}^{(\mathbb{J}_1)} = -\frac{1}{N} \sum_{k=1}^{N} \frac{1}{\theta \left(1 + \Delta_{N,(\mathbb{J}_2)}^{(\mathbb{J}_1)} + T_k + \Upsilon_{(\mathbb{J}_2)}^{(\mathbb{J}_1 \cup \{k\})}\right)} = -\frac{1}{N} \sum_{k=1}^{N} \frac{1}{\theta \left(1 + \Delta_{N,(\mathbb{J}_2)}^{(\mathbb{J}_1)} + \mathcal{T}_k + Y_{(\mathbb{J}_2 \cup \{k\})}^{(\mathbb{J}_1)}\right)} \tag{3.1}$$

with

$$|T_k|, |\mathcal{T}_k| \leq \frac{\left|\, |\mathbb{J}_1| - |\mathbb{J}_2| \,\right| + 1}{N\eta} \tag{3.2}$$

and

$$\Upsilon_{(\mathbb{J}_2)}^{(\mathbb{J}_1 \cup \{k\})} := (\mathbb{I} - \mathbb{E}_{\boldsymbol{x}^k})(\boldsymbol{x}^k/\sqrt{N})^* \mathcal{G}_{(\mathbb{J}_2)}^{(\mathbb{J}_1 \cup \{k\})} \boldsymbol{x}^k/\sqrt{N} \quad \text{and} \quad Y_{(\mathbb{J}_2 \cup \{k\})}^{(\mathbb{J}_1)} := (\mathbb{I} - \mathbb{E}_{\boldsymbol{x}_k}) \frac{\boldsymbol{x}_k}{\sqrt{N}} G_{(\mathbb{J}_2 \cup \{k\})}^{(\mathbb{J}_1)} \boldsymbol{x}_k^*/\sqrt{N}. \tag{3.3}$$

$$T_k := \frac{1}{N} \operatorname{Tr} \mathcal{G}_{(\mathbb{J}_2)}^{(\mathbb{J}_1) \cup \{k\}} - \frac{1}{N} \operatorname{Tr} G_{(\mathbb{J}_2)}^{(\mathbb{J}_1)} \quad \text{and} \quad \mathcal{T}_k := \frac{1}{N} \operatorname{Tr} G_{(\mathbb{J}_2)}^{(\mathbb{J}_1) \cup \{k\}} - \frac{1}{N} \operatorname{Tr} \mathcal{G}_{(\mathbb{J}_2)}^{(\mathbb{J}_1)}, \tag{3.4}$$

where $\boldsymbol{x}^k$ is the $k-th$ column of the matrix X and $\boldsymbol{x}_k$ is the $k-th$ row of X.

Proof. First we recall that for any matrix X and θ with non-zero imaginary part we have that

$$X(X^*X-\theta\mathbf{I})^{-1}X^* = XX^*(XX^*-\theta)^{-1}. \tag{3.5}$$

This can be proved by expanding the inverse in a Taylor series, followed by the use of matrix associativity. Furthermore using (3.5) and $\mathbf{I}=(XX^*-\theta\mathbf{I})(XX^*-\theta\mathbf{I})^{-1}$ we obtain that

$$X(X^*X-\theta\mathbf{I})^{-1}X^*-\mathbf{I} = \theta(XX^*-\theta)^{-1}. \tag{3.6}$$

By the definition of Δ_N and using (3.6) we get

$$\begin{aligned} G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kk} &= \frac{1}{\frac{|\mathbf{x}^k|^2}{N}-\theta-(\mathbf{x}^k)^*X^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}\left((X^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)})^*X^{(\mathbb{J}_1)\cup\{k\}}_{(\mathbb{J}_2)}-\theta\right)^{-1}(X^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)})^*\mathbf{x}^k} \\ &= -\frac{1}{\theta\left(1+(\mathbf{x}^k/\sqrt{N})^*\mathcal{G}^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}\mathbf{x}^k/\sqrt{N}\right)} = -\frac{1}{\theta\left(1+\Delta^{(\mathbb{J}_1)}_{N,(\mathbb{J}_2)}+T_k+\Upsilon^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}\right)}, \end{aligned} \tag{3.7}$$

which yields

$$\Delta^{(\mathbb{J}_1)}_{N,(\mathbb{J}_2)} = -\frac{1}{N}\sum_{k=1}^{N}\frac{1}{\theta\left(1+\Delta^{(\mathbb{J}_1)}_{N,(\mathbb{J}_2)}+T_k+\Upsilon^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}\right)} \tag{3.8}$$

where $\Upsilon^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}$ is as in (3.3) and

$$\begin{aligned} T_k &= \frac{1}{N}\operatorname{Tr}(X^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}(X^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)})^*-\theta)^{-1}-\frac{1}{N}\operatorname{Tr}((X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)})^*(X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)})-\theta)^{-1} \\ &= -\frac{|\mathbb{J}_1|-|\mathbb{J}_2|}{N\theta}+\frac{1}{N}\operatorname{Tr}((X^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)})^*X^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}-\theta)^{-1}-\frac{1}{N}\operatorname{Tr}((X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)})^*X^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}-\theta)^{-1}. \end{aligned} \tag{3.9}$$

Rewriting, we obtain

$$\begin{aligned} T_k &= -\frac{|\mathbb{J}_1|-|\mathbb{J}_2|}{N\theta}+\frac{1}{N}\left(\sum_{i\neq k}G^{(k)}_{ii}-\sum_{i=1}^{N}G_{ii}\right) = -\frac{|\mathbb{J}_1|-|\mathbb{J}_2|}{N\theta}+\frac{1}{N}\left(\sum_{i\neq k}\left(G_{ii}-\frac{G_{ik}G_{ki}}{G_{kk}}\right)-\sum_{i=1}^{N}G_{ii}\right) \\ &= -\frac{|\mathbb{J}_1|-|\mathbb{J}_2|}{N\theta}-\frac{1}{N}\left(\sum_{i\neq k}\frac{G_{ik}G_{ki}}{G_{kk}}-G_{kk}\right) = -\frac{|\mathbb{J}_1|-|\mathbb{J}_2|}{N\theta}-\frac{1}{N}\frac{1}{G_{kk}}\sum_{i=1}^{N}G_{ik}G_{ki} = -\frac{|\mathbb{J}_1|-|\mathbb{J}_2|}{N\theta}-\frac{(G^2)_{kk}}{NG_{kk}}. \end{aligned} \tag{3.10}$$

We now use (2.5) to obtain

$$|T_k| \leq \frac{|\,|\mathbb{J}_1|-|\mathbb{J}_2|\,|}{N|\theta|}+\frac{\operatorname{Im} G_{kk}}{|G_{kk}|N\eta} \tag{3.11}$$

yielding that

$$|\sqrt{\theta}|\,|T_k| \leq \frac{(|\,|\mathbb{J}_1|-|\mathbb{J}_2|\,|+1)|\sqrt{\theta}|}{N\eta}.$$

Note also

$$\mathcal{G}^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kk} = -\frac{1}{\theta\left(1+(\mathbf{x}_k/\sqrt{N})\left((X^{(\mathbb{J}_1)}_{(\mathbb{J}_2\cup\{k\})})^*X^{(\mathbb{J}_1)}_{(\mathbb{J}_2\cup\{k\})}-\theta\right)^{-1}\mathbf{x}^*_k/\sqrt{N}\right)} \tag{3.12}$$

which similarly yields the second part of (3.1), recalling (2.2). □

Rewriting (3.1) using $\frac{1}{A+\epsilon} = \frac{1}{A} - \frac{\epsilon}{A(A+\epsilon)}$ we obtain (also for any $\mathbb{J}_1, \mathbb{J}_2$, thus we suppress them here)

$$\begin{aligned}\Delta_N &= -\frac{1}{N}\sum_{k=1}^{N} \frac{1}{\theta(1+\Delta)+\theta\Lambda+\theta(T_k+\Upsilon^{(\{k\})})} \\ &= -\frac{1}{\theta(1+\Delta)} - \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\theta(1+\Delta)} \frac{\theta\Lambda+\theta(T_k+\Upsilon^{(\{k\})})}{\theta\left(1+\Delta_N+(T_k+\Upsilon^{(\{k\})})\right)} \\ &= \Delta - \frac{\Delta}{N}\sum_{k=1}^{N}\theta\Lambda G_{kk} + -\frac{\Delta}{N}\sum_{k=1}^{N} G_{kk}\theta(T_k+\Upsilon^{(\{k\})}) \\ &= \Delta - \Delta\theta\Lambda\Delta_N + -\frac{\Delta}{N}\sum_{k=1}^{N} G_{kk}\theta(T_k+\Upsilon^{(\{k\})})\end{aligned} \tag{3.13}$$

This yields that

$$\Lambda = \theta\Delta\Lambda(\Delta+\Lambda) + \frac{\Delta}{N}\sum_{k=1}^{N} G_{kk}\theta(T_k+\Upsilon^{(\{k\})}) \tag{3.14}$$

Let

$$R := N^{-1}\sum_{k=1}^{N} G_{kk}(T_k+\Upsilon^{(\{k\})}) \tag{3.15}$$

and (similarly can define $R^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}$) which yields the following quadratic for Λ

$$\theta\Delta\Lambda^2 + (\theta\Delta^2-1)\Lambda + \Delta\theta R = 0. \tag{3.16}$$

Dividing by $\theta\Delta$, using that $\theta\Delta = -\frac{1}{1+\Delta}$ and the quadratic formula, yields

$$-(\Delta+1/2) \pm \sqrt{(\Delta+1/2)^2 - R} \tag{3.17}$$

as two solutions. From definition of Λ in (2.3) it follows that $\operatorname{Im}\Lambda > \operatorname{Im}\Delta$, thus if we take the branch cut of the square root to be on the positive reals so that the imaginary part of the square root is always positive, we obtain that

$$\Lambda = -(\Delta+1/2) + \sqrt{(\Delta+1/2)^2 - R} \tag{3.18}$$

We also notice that the second solution, call it $\tilde{\Lambda}$, to (3.16) is given by

$$\tilde{\Lambda} = -\Lambda - 2\Delta - 1. \tag{3.19}$$

The following proposition is analogous to Proposition 2.2 of [2].

Proposition 3.2. *Let $\theta = E + i\eta$. There exists a constant $C > 0$, such that:*

$$|\Lambda| \le C\min\left\{\frac{|R|}{|\Delta+\frac{1}{2}|}, \sqrt{|R|}\right\}, \tag{3.20}$$

for all $(E,\eta) \in S_{\eta,E}$ as well as for any $E < 0$. Furthermore, for any $E \in \mathbb{R}$ and $\eta > 0$ we have that

$$|\operatorname{Im}\Lambda| < C\min\left\{\frac{|R|}{|\Delta+\frac{1}{2}|}, \sqrt{|R|}\right\} \tag{3.21}$$

and

$$\min\{|\Lambda|, |\tilde{\Lambda}|\} \le C\sqrt{|R|}. \tag{3.22}$$

Analogous statements hold for $\Lambda^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}$ with $R^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}$.

Proof. To show (3.20), we apply (2.17) of [2] with $a = (\Delta + \frac{1}{2})^2$ and $\mathrm{b} = -R$. Since $\operatorname{Im} \Delta > 0$, with our choice of branch cut we have that $\sqrt{(\Delta + 1/2)^2} = \Delta + 1/2$, and we recall that we defined $S_{\eta,E}$ in (1.13) to be exactly the set where $\operatorname{Im}(\Delta + 1/2)^2 \geq c \operatorname{Re}((\Delta + 1/2)^2)$ for some $c > 0$. Note that (3.21) follows directly from (2.18) of [2], while the proof of (3.22) is identical to the proof of (2.16) in [2].

Recalling that $\Delta_N(z) = \frac{1}{N} \sum_\alpha \frac{1}{s_\alpha - E - i\eta}$ and noting that for $E < 0$ the real part of each summand is positive we conclude that $\operatorname{Re} \Delta_N > 0$ for $E < 0$, and similar to our argument about the imaginary part of Λ, we see from (2.3) that $\operatorname{Im} \Lambda > -\operatorname{Im} \Delta$ while from (3.19) we see that $\operatorname{Re} \tilde{\Lambda} < -\operatorname{Re} \Delta - 1$. Since we have that

$$\operatorname{Re} \Lambda = -\operatorname{Re}(\Delta + 1/2) + \operatorname{Re}\left(\sqrt{(\Delta + 1/2)^2 - R}\right)$$
$$\operatorname{Re} \tilde{\Lambda} = -\operatorname{Re}(\Delta + 1/2) - \operatorname{Re}\left(\sqrt{(\Delta + 1/2)^2 - R}\right)$$

we see that $\operatorname{Re}\left(\sqrt{(\Delta + 1/2)^2 - R}\right) > 0$ and thus $|\operatorname{Re} \Lambda| < |\operatorname{Re} \tilde{\Lambda}|$ and thus one part of (3.20) follows from (3.22). For the other part of (3.20), we estimate that

$$|\Lambda| = \left| \frac{R}{\sqrt{(\Delta + 1/2)^2 - R} + (\Delta + 1/2)} \right| \leq \left| \frac{R}{\Delta + 1/2} \right| \tag{3.23}$$

where the last inequality follows since both real and imaginary parts of both summands in the denominator are positive. The fact that $\operatorname{Re}(\Delta + \frac{1}{2}) \geq 0$ comes from the definition of our spectral domain, namely by the constraint $E^2 + \eta^2 - 4E \leq 4\eta$.

□

4. Bounds on quadratic forms

Here we obtain the necessary bounds on quadratic forms.

Lemma 4.1. *Let* $\mathcal{G} = \mathcal{G}^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}$ *or* $G^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}$ *for some* $\mathbb{J}_1, \mathbb{J}_2$. *Let* $\Upsilon := \frac{1}{N}(\mathbb{I} - \mathbb{E}_{\mathbf{x}})\mathbf{x}^* \mathcal{G} \mathbf{x}$, *assuming* (1.6), (1.7) *for elements of* $\mathbf{x}$. *Then we have that*

$$\mathbb{E}|\Upsilon|^{2q} \leq (Cq)^{cq} \left(\frac{\mathbb{E}(\operatorname{Im} \operatorname{Tr} \mathcal{G})^q}{N^q (N\eta)^q} + \frac{\mathbb{E}|\mathcal{G}_{11}|^{2q}}{N^q} + \frac{\mathbb{E}|\mathcal{G}_{11}|^q}{(N\eta)^q} \right). \tag{4.1}$$

Moreover, we have a more precise inequality

$$\mathbb{E}|\Upsilon|^{2q} \leq \left(\frac{Cq}{N\eta}\right)^{cq} \left(\mathbb{E}\left(\frac{\operatorname{Im} \operatorname{Tr} \mathcal{G}}{N}\right)^q + \mathbb{E}\left|\frac{\mathcal{G}_{11}}{\sqrt{N}}\right|^q \right) + \frac{(Cq)^{cq} \mathbb{E}(|\mathcal{G}_{12}|^{2q} + |\mathcal{G}_{11}|^{2q})}{N^q}. \tag{4.2}$$

Proof. We start by the decomposition:

$$\Upsilon = \frac{1}{N} \sum_{j \neq l} \overline{x_j} x_l \mathcal{G}_{jl} + \frac{1}{N} \sum_j (|x_{jk}|^2 - 1) \mathcal{G}_{jj} = \epsilon_2 + \epsilon_1,$$

where

$$\epsilon_2 := \frac{1}{N} \sum_{j \neq l} \overline{x_j} x_l \mathcal{G}_{jl} \text{ and } \epsilon_1 := \frac{1}{N} \sum_j (|x_j|^2 - 1) \mathcal{G}_{jj}. \tag{4.3}$$

We use Rosenthal's inequality (see e.g. Lemma 1 in [7]) to obtain:

$$\mathbb{E}|\epsilon_1|^{2q} \leq (Cq)^{2q} N^{-2q} \left[\sum_j \mathbb{E}|x_j|^{4q} \mathbb{E}|\mathcal{G}_{jj}|^{2q} + \left(\mu_4 \sum_j \mathbb{E}|\mathcal{G}_{jj}|^2 \right)^q \right]. \tag{4.4}$$

We notice that

$$|x_l| \leq DN^{1/4} \Rightarrow \mu_{4q} := \mathbb{E}|x_l|^{4q} \leq D^{4q-4} N^{q-1} \mu_4, \tag{4.5}$$

which yields that

$$\mathbb{E}|\epsilon_1|^{2q} \leq (Cq)^{2q} N^{-q} \mathbb{E}|\mathcal{G}_{jj}|^{2q}.$$

For ϵ_2 we use Corollary 1 from [7]. We notice that in our notation $a_{kl} = \frac{\mathcal{G}_{kl}}{N}$, $\sigma^2 = \mathbb{E}|x_l|^2 = 1$ and μ_{2p} is given in (4.5). We also notice that while Corollary 1 in [7] is formulated for real variables, it works in an identical way up to constants by separating real and imaginary parts of the random variables. With this translation of notation in place and absorbing the constants in (4.5) into the global constant, we obtain

$$\mathbb{E}|\epsilon_2|^{2q} \leq (C_1 q)^{4q}\mathbb{E}\left(\left(\sum_{k,l=1,k\neq l}^{N}\left|\frac{\mathcal{G}_{kl}}{N}\right|^2\right)^q + N^{q/2-1}\sum_{k=1}^{N}\left(\sum_{l=1,l\neq k}^{N}\left|\frac{\mathcal{G}_{kl}}{N}\right|^2\right)^q + N^{q-2}\sum_{k,l=1,k\neq l}^{N}\left|\frac{\mathcal{G}_{kl}}{N}\right|^{2q}\right) \tag{4.6}$$

We obtain the bounds proportional to $\mathbb{E}\left|\frac{\mathrm{Tr}\,\mathrm{Im}\,\mathcal{G}}{N^2\eta}\right|^q$ and $N^{-q/2}\mathbb{E}\left|\frac{\mathrm{Im}\;\mathcal{G}_{11}}{N\eta}\right|^q$ for the first two terms on the RHS of (4.6) respectively using that $\sum_{l=1}^{N}|\mathcal{G}_{jl}|^2 \leq \eta^{-1}\mathrm{Im}\;\mathcal{G}_{jj}$, which yields

$$\mathbb{E}|\epsilon_2|^{2q} \leq \frac{(Cq)^{4q}}{(N\eta)^q}\left(\mathbb{E}\left|\frac{\mathrm{Im}\;\mathcal{G}_{11}}{\sqrt{N}}\right|^q + \mathbb{E}\left|\frac{\mathrm{Tr}\,\mathrm{Im}\;\mathcal{G}}{N}\right|^q\right) + (Cq)^{4q}\mathbb{E}\left|\frac{\mathcal{G}_{12}}{\sqrt{N}}\right|^{2q}. \tag{4.7}$$

and putting all the bounds together (4.2) follows.
To obtain (4.1) we need to bound the off-diagonal $\mathbb{E}|\mathcal{G}_{12}|$ in terms of the diagonal. We observe that

$$|\mathcal{G}_{lj}| \leq \frac{1}{2}\sqrt{\frac{\mathrm{Im}\;\mathcal{G}_{ll}}{\eta}} + \frac{1}{2}\sqrt{\frac{\mathrm{Im}\;\mathcal{G}_{jj}}{\eta}} \tag{4.8}$$

which can be obtained as follows. Let u_j be the jth normalized eigenvector of $\mathcal{G}$ and $\lambda_0 = 0$. Then

$$|\mathcal{G}_{lj}| = \left|\sum_{q=0}^{N}\frac{u_{lq}u_{qj}}{\lambda_q - z}\right| \leq \sum_{q=0}^{N}\frac{|u_{lq}u_{qj}|}{|\lambda_q - z|} \leq \frac{1}{2}\sum_{q=0}^{N}\frac{|u_{lq}|^2+|u_{qj}|^2}{|\lambda_q - z|} \leq \frac{1}{2}\sqrt{\sum_{q=0}^{N}\frac{|u_{lq}|^2}{|\lambda_q - z|^2}} + \frac{1}{2}\sqrt{\sum_{q=0}^{N}\frac{|u_{qj}|^2}{|\lambda_q - z|^2}} \tag{4.9}$$

where in the last step we recall that the eigenvectors are normalized and use Jensen's ineqality. □

5. Non-optimal bound and bootstrap argument in the bulk

Let

$$\lambda^{(\mathbb{J}_1)}_{(\mathbb{J}_2)} := \max\{|\Lambda^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}|\chi_{S_{E,\eta}}, \min\{|\Lambda^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}|, |\tilde{\Lambda}^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}|\}, |\;\mathrm{Im}\;\Lambda^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}|\}, \tag{5.1}$$

where $\chi_{S_{E,\eta}}$ is the indicator function of $S_{E,\eta}$ By Proposition 3.2, $|\theta|^q\mathbb{E}|\lambda^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}|^{2q} \le C^{2q}|\theta|^q\mathbb{E}|R^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}|^q$. Taking expectation of a power of θR we obtain (as in [2])

$$\begin{aligned}
\mathbb{E}|\theta R^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}|^q &\le \frac{|\theta|^q}{N}\sum_{k=1}^N \mathbb{E}\left|\left(T_k + \Upsilon^{(\mathbb{J}_1\cup\{k\})}_{(\mathbb{J}_2)}\right) G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kk}\right|^q \\
&\le |\theta|^q\,\mathbb{E}\left|\left(T_1 + \Upsilon^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2)}\right) G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}\right|^q \\
&\le \frac{\mathbb{E}\left|C\theta(|\,|\mathbb{J}_1|-|\mathbb{J}_2|\,|+1)G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}\right|^q}{(N\eta)^q} + |C\theta|^q\sqrt{\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}\mathbb{E}\left|\Upsilon^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2)}\right|^{2q}} \\
&\le \frac{\mathbb{E}\left|C\theta(|\,|\mathbb{J}_1|-|\mathbb{J}_2|\,|+1)G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}\right|^q}{(N\eta)^q} + |C\theta q|^{cq}\sqrt{\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}\frac{\mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^{2q}}{N^q}} \\
&\quad + |C\theta q|^{cq}\sqrt{\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}\left(\mathbb{E}\frac{(\text{Im Tr}\,\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2)})^q}{(N\eta)^qN^q} + \frac{\mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^q}{(N\eta)^q}\right)} \\
&\le \frac{\mathbb{E}\left|C\theta(|\,|\mathbb{J}_1|-|\mathbb{J}_2|\,|+1)G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}\right|^q}{(N\eta)^q} + |C\theta q|^{cq}\sqrt{\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}\frac{\mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^{2q}}{N^q}} \\
&\quad + |C\theta|^q\sqrt{\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}\frac{(Cq)^{cq}}{(N\eta)^q}\left(\left(\frac{|\,|\mathbb{J}_1|+1-|\mathbb{J}_2|\,|}{N\eta}\right)^q + \mathbb{E}\left(\text{Im }\Delta + \text{Im }\Lambda^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2)}\right)^q + \mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^q\right)} \\
&\le |C\theta q|^{cq}\sqrt{\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}\frac{\mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^{2q}}{N^q}} \\
&\quad + |Cq|^{cq}\,\frac{|\theta|^{\frac{q}{4}}\sqrt{\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}}}{(N\eta)^{\frac{q}{2}}}\left(|\,|\mathbb{J}_1|+1-|\mathbb{J}_2|\,|^q + \sqrt{\mathbb{E}|\sqrt{\theta}\lambda|^q} + \sqrt{\mathbb{E}|\sqrt{\theta}\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^q}\right) \quad (5.2)
\end{aligned}$$

In the second to last line, the term $\frac{|\mathbb{J}_1|+1-|\mathbb{J}_2|}{N\eta}$ arises from equation (2.2), and Im $\Lambda^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2)}$ is close to Im $\Lambda^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}$ similar to (3.10). Since for any $x,\delta>0$, $x^{1/4} < \delta x + \delta^{-1/3}$, setting $\delta = \left(2(\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q})^{1/2}\frac{(Cq(|\,|\mathbb{J}_1|-|\mathbb{J}_2|\,|+1))^{cq}|\theta|^{q/4}}{(N\eta)^{q/2}}\right)^{-1}$ and using Cauchy-Schwarz inequality on $\sqrt{\mathbb{E}|\sqrt{\theta}\lambda|^q}$, we get

$$\begin{aligned}
|\theta|^q\mathbb{E}\lambda^{2q} \le \frac{(Cq(|\,|\mathbb{J}_1|-|\mathbb{J}_2|\,|+1))^{cq}}{(N\eta)^{q/6}}|\theta|^{q/12}\left((\mathbb{E}|\sqrt{\theta}\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^{2q})^{1/2} + |\,|\mathbb{J}_1|-|\mathbb{J}_2|+1|^{\frac{q}{2}}\right)\left((\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q})^{1/2}+1\right) \\
+ |C\theta q|^{cq}\sqrt{\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}|^{2q}\frac{\mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}|^{2q}}{N^q}}. \quad (5.3)
\end{aligned}$$

Lemma 5.1. *Let* $q < \left(\frac{N\eta}{|\sqrt{\theta}|}\right)^{1/4}$, $N\eta > |\sqrt{\theta}|M$ *for some constant* $M>0$*, fixed* E*. Assume that* $\mathbb{J}_1, \mathbb{J}_2$ *are such that* $0 \le |\mathbb{J}_1|-|\mathbb{J}_2| \le Cq$ *for a uniform constant* C*. Then with definitions as before,* $\mathbb{E}|G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),11}\sqrt{\theta}|^q \le C^q$ *and* $\mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}\sqrt{\theta}|^q \le C^q$*, for some constant* C*.*

Proof. We will implement an induction argument similar to [2, 5]. The induction hypothesis will be that for $\eta_j = \eta_0/16^j$ for some constant η_0, any $\mathbb{J}_{1,j}, \mathbb{J}_{2,j}$ with $|\mathbb{J}_{1,j}| = |\mathbb{J}_{2,j}| \le |\log_{16}\eta| + 1 - j =: L_j$ and $k \notin \mathbb{J}_{1,j}$

$$\mathbb{E}|G^{(\mathbb{J}_1\cup\mathbb{J}_{1,j})}_{(\mathbb{J}_2\cup\mathbb{J}_{2,j}),11}(\eta_j)\sqrt{\theta}|^q < C_0^q \quad \text{and} \quad \mathbb{E}|\mathcal{G}^{(\mathbb{J}_1\cup\mathbb{J}_{1,j}\cup\{1\})}_{(\mathbb{J}_2\cup\mathbb{J}_{2,j}),22}(\eta_j)\sqrt{\theta}|^q < C_0^q \quad (5.4)$$

for $q < \left(\frac{N\eta_j}{|\sqrt{E}|}\right)^{1/4}$ for a universal constant C_0. We notice that this holds to initiate our induction for η_0 constant. Letting $\eta_{j+1} = \eta_j/16$ and $L_{j+1} = L_j - 1$ we will show that inequality (5.4) taken at η_j implies the

same inequality with the same constant C_0 for η_{j+1}. For easier notation, we will suppress the dependence on $\mathbb{J}_1, \mathbb{J}_2$ mentioning only the step where they come up (which is equation (5.12)).
From the induction hypothesis and Lemma 2.2 we see that

$$\mathbb{E}|G^{(\mathbb{J}_{1,j})}_{(\mathbb{J}_{2,j}),11}(\eta_{j+1})\sqrt{\theta}|^q < (16C_0)^q \quad \text{and} \quad \mathbb{E}|\mathcal{G}^{(\mathbb{J}_{1,j}\cup\{1\})}_{(\mathbb{J}_{2,j}),22}(\eta_{j+1})\sqrt{\theta}|^q < (16C_0)^q \tag{5.5}$$

for any $\mathbb{J}_{1,j}, \mathbb{J}_{2,j}$ with $|\mathbb{J}_{1,j}| = |\mathbb{J}_{2,j}| \leq L - j$. This will need to be improved to the bound C_0^q for any $\mathbb{J}_{1,j+1}, \mathbb{J}_{2,j+1}$ of size up to $L-j-1$.
From (3.7) and (3.19) we obtain that

$$G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11} = \Delta - \sqrt{\theta}\Delta(\sqrt{\theta}\Lambda^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})} - \sqrt{\theta}T_1 - \sqrt{\theta}\Upsilon^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})})G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11} \tag{5.6}$$

$$G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11} = \Delta - \sqrt{\theta}\Delta(-\sqrt{\theta}\tilde{\Lambda}^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}) - \sqrt{\theta}T_1 - \sqrt{\theta}\Upsilon^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})})G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11} + \theta\Delta(2\Delta+1)G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}. \tag{5.7}$$

The analogous statements for $\mathcal{G}^{(\mathbb{J}_{1,j}\cup\{1\})}_{(\mathbb{J}_{2,j}),22}$ follow similarly from (3.12) and (3.19):

$$\mathcal{G}^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1}),22} = \Delta - \sqrt{\theta}\Delta\left[\sqrt{\theta}\Lambda^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1}),11} - \sqrt{\theta}\mathcal{T}_1 - \sqrt{\theta}Y^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1}\cup\{2\})}\right]\mathcal{G}^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1}),22}. \tag{5.8}$$

This yields that

$$|G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}| \leq |\Delta| + |G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|(|\sqrt{\theta}\Lambda^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}| + |\sqrt{\theta}T^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})}| + |\sqrt{\theta}\Upsilon^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})}|)|\sqrt{\theta}\Delta|$$

$$|G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}| \leq \left|\frac{\Delta}{1-\theta\Delta(2\Delta+1)}\right| + |G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|(|\sqrt{\theta}\tilde{\Lambda}^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}| + |\sqrt{\theta}T^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})}| + |\sqrt{\theta}\Upsilon^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})}|)|\sqrt{\theta}\Delta|$$

and using (1.11) we see that $1-\theta\Delta(2\Delta+1) = -\theta\Delta^2$ and thus $\frac{\Delta}{1-\theta\Delta(2\Delta+1)} = \frac{1}{\theta\Delta}$.
We will use the bounds $C_1 \leq |\sqrt{\theta}\Delta| \leq C_2$, valid in our domain, and let $C = \max\{C_1, C_2\}$. So, we have that:

$$|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}| \leq C\left[1 + |\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|(|\sqrt{\theta}|\min\{|\Lambda^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}|, |\tilde{\Lambda}^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}|\} + |\sqrt{\theta}T_1| + |\sqrt{\theta}\Upsilon^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})}|)\right]$$

and, taking power q, expectation, and using Cauchy-Schwarz we get at η_{j+1}

$$\begin{aligned}\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^q \leq C^q\big[1 &+ \sqrt{\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^{2q}}\sqrt{\mathbb{E}(|\sqrt{\theta}\lambda^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}|)^{2q}} \\ &+ \frac{|\sqrt{\theta}|^q}{(N\eta_{j+1})^q}\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^q + \sqrt{\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^{2q}}\sqrt{\mathbb{E}|\sqrt{\theta}\Upsilon^{(\mathbb{J}_{1,j+1}\cup\{1\})}_{(\mathbb{J}_{2,j+1})}|^{2q}}\big]\end{aligned} \tag{5.9}$$

Using the above, Lemma 4.1, and a calculation similar to (5.2) we obtain again at η_{j+1}

$$\begin{aligned}\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^q \leq (Cq)^{cq}\Bigg[1 &+ \sqrt{\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^{2q}}\sqrt{\mathbb{E}|\sqrt{\theta}\lambda^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}|^{2q}} + \frac{|\sqrt{\theta}|^q}{(N\eta_{j+1})^q}\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^q \\ &+ \sqrt{\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^{2q}}\frac{|\sqrt{\theta}|^{q/4}}{(N\eta_{j+1})^{q/2}}\sqrt{1 + \mathbb{E}|\sqrt{\theta}\lambda^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}|^q + \mathbb{E}|\sqrt{\theta}\mathcal{G}^{(\mathbb{J}_{1,j+1})\cup\{1\}}_{(\mathbb{J}_{2,j+1}),22}|^q} \\ &+ \sqrt{\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^{2q}\frac{\mathbb{E}|\sqrt{\theta}\mathcal{G}^{(\mathbb{J}_{1,j+1})\cup\{1\}}_{(\mathbb{J}_{2,j+1}),22}|^{2q}}{N^q}}\Bigg]\end{aligned} \tag{5.10}$$

We use (5.5) to bound the terms $\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^{2q}$ and $\mathbb{E}|\sqrt{\theta}\mathcal{G}^{(\mathbb{J}_{1,j+1})\cup\{1\}}_{(\mathbb{J}_{2,j+1}),22}|^{2q}$ in the above inequality, noting that $|\mathbb{J}_{1,j+1}\cup\{1\}| \leq L_j$. To use (5.5) we need $2q \leq \left(\frac{N\eta_j}{|\sqrt{E}|}\right)^{1/4}$, which gives us exactly the requirement that

$$q \leq \left(\frac{N\eta_j}{16|\sqrt{E}|}\right)^{1/4} = \left(\frac{N\eta_{j+1}}{|\sqrt{E}|}\right)^{1/4} \tag{5.11}$$

found in the in the induction hypothesis. Then using (5.5) on equation (5.3) at η_{j+1} and recalling that $||\mathbb{J}_1| - |\mathbb{J}_2|| \le Cq$ we obtain

$$\mathbb{E}|\sqrt{\theta}\lambda^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1})}|^q \le \frac{(Cq)^{cq}}{(N\eta_{j+1})^{q/6}}|\theta|^{q/12}\left((16C_0)^q + (Cq)^{q/2}\right)^2 + \frac{(Cq)^{cq}(16C_0)^{2q}}{N^{q/2}} \tag{5.12}$$

Substituting this into (5.10), we obtain that at η_{j+1}:

$$\begin{aligned}\mathbb{E}|\sqrt{\theta}G^{(\mathbb{J}_{1,j+1})}_{(\mathbb{J}_{2,j+1}),11}|^q \le (Cq)^{cq}\Bigg[&1 + (16C_0)^q\frac{|\theta|^{q/12}}{(N\eta_{j+1})^{q/6}}\left((16C_0)^q + (Cq)^{cq}\right) + \frac{(16C_0)^{3q}}{N^{q/2}}\\ &+ (16C_0)^q\frac{q^{cq}|\sqrt{\theta}|^{q/4}}{(N\eta_{j+1})^{q/2}}\sqrt{2 + \frac{(Cq)^{cq}}{(N\eta_{j+1})^{q/6}}|\theta|^{q/12}\left((16C_0)^q + (Cq)^{q/2}\right)^2 + (16C_0)^q}\Bigg]\\ &\le (Cq)^{cq}\left[2 + K^q\left(\frac{|\sqrt{\theta}|}{N\eta}\right)^{q/6}\right]\end{aligned} \tag{5.13}$$

for a constant $K > 0$ depending on C_0 and C. We can choose $C_0 > 2C$ and $\frac{N\eta}{|\sqrt{\theta}|} > M > K^6$, so that $K^q\left(\frac{|\sqrt{\theta}|}{N\eta}\right)^{q/6} < 1$ and therefore $\mathbb{E}[\sqrt{\theta}G_{11}(\eta_{j+1})]^q < C_0^q$ as required. We notice that all the steps are identical for $\mathcal{G}^{(\mathbb{J}_1\cup\{1\})}_{(\mathbb{J}_2),22}$ using (5.8), and exactly one row gets stripped as well as exactly one column so $|\mathbb{J}_{1,j+1}| = |\mathbb{J}_{2,j+1}|$. □

6. Optimal Bound for the Stieltjes transform

In this section we prove Theorem 1. We will use the matrix expansion algorithm from [2], which carries over directly as it is based entirely on linear algebra of resolvents. We will make a note of the important modifications. We note, importantly, that as we expand resolvent entries, we will be removing columns of X_N and we never need to remove rows. The expansion algorithm yields results in terms of high moments of the following quantities:

$$|\sqrt{\theta}G^{(\mathbb{J})}_{kk}|, \left|\frac{1}{\sqrt{\theta}G^{(\mathbb{J})}_{kk}}\right|, \left|(\mathbb{I} - \mathbb{E}_k)\frac{1}{\sqrt{\theta}G^{(\mathbb{J})}_{kk}}\right|, |\sqrt{\theta}G^{(\mathbb{J})}_{kl}|, \tag{6.1}$$

and we begin this section by estimating these moments.

To obtain optimal bounds on Λ near the soft edge the fluctuation bound on relevant quadratic forms (4.1) needs to be improved. For that purpose we will use (4.1) to obtain bounds on $|\sqrt{\theta}G^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}|$ as well as $|\sqrt{\theta}\mathcal{G}^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}|$ then use these in (4.2) to improve on the RHS of (4.2). For convenience of notation we introduce the control parameter

$$\mathcal{E}_q := \frac{1}{N^q|\theta|^{q/2}} + \max\left\{\frac{[\mathrm{Im}\,(|\theta|\Delta)]^q + \mathbb{E}|\theta\Lambda|^q}{(N\eta)^q}, \frac{|\theta|^q}{(N\eta)^{2q}}\right\}. \tag{6.2}$$

We now show how to estimate the last quantity in (6.1), using the formulas (see e.g. (2.20) of [12]) (valid also for any $\mathbb{J}_1, \mathbb{J}_2$, with $k, l \notin \mathbb{J}_1 \cup \mathbb{J}_2$)

$$\begin{aligned}\sqrt{\theta}G_{kl} &= \sqrt{\theta}G_{ll}\sqrt{\theta}G^{(\{l\})}_{kk}(\sqrt{\theta}(\mathbf{x}^k/\sqrt{N})^*\mathcal{G}^{(\{k,l\})}(\mathbf{x}^l/\sqrt{N})) =: \sqrt{\theta}G_{ll}\sqrt{\theta}G^{(\{l\})}_{kk}K_{kl}\\ \sqrt{\theta}\mathcal{G}_{kl} &= \sqrt{\theta}\mathcal{G}_{ll}\sqrt{\theta}\mathcal{G}_{(\{l\}),kk}(\sqrt{\theta}(\mathbf{x}_k/\sqrt{N})G_{(\{k,l\})}(\mathbf{x}_l/\sqrt{N}))^* =: \sqrt{\theta}\mathcal{G}_{ll}\sqrt{\theta}\mathcal{G}_{(\{l\}),kk}\mathcal{K}_{kl}.\end{aligned} \tag{6.3}$$

We can define $K^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}, \mathcal{K}^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}$ analogously. The following lemma provides the necessary bound on $\mathbb{E}|K_{kl}|^{2q}$ and an improved bound on $\Upsilon^{(\mathbb{J}_1)}_{(\mathbb{J}_2)}$.

Lemma 6.1. *Assume* (1.6) *and* (1.7) *for the entries of the matrix X_N as before and let $\theta = E + i\eta$. Then there exist constants $c, c_0, C, M_1, M_2 > 0$ such that*

$$\max\{\mathbb{E}|K_{kl}|^{2q}, \mathbb{E}|\mathcal{K}_{kl}|^{2q}\} \leq (Cq)^{cq}\mathcal{E}_q \tag{6.4}$$

for $E, \eta \in S_{E,\eta}$, $N > M_1$, $\frac{N\eta}{|\sqrt{\theta}|} > M_2$, $k \neq l \in \{1, ..., N\}$, $q \in \mathbb{N}$ with $q \leq c_0 N$. Assuming $|\,|\mathbb{J}_1| - |\mathbb{J}_2|\,| < Cq$ for some constant C, same inequality holds for $K^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}, \mathcal{K}^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}$.

Proof. The following argument is identical for $K^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}, \mathcal{K}^{(\mathbb{J}_1)}_{(\mathbb{J}_2),kl}$, so we work with K_{kl}. By the definition of K_{kl} and using the notation $\epsilon_{k1}, \epsilon_{k_2}$ for ϵ_1 and ϵ_2 as in (4.3) we get that:

$$\mathbb{E}|K_{kl}|^{2q} \leq \frac{(C|\sqrt{\theta}|)^q}{N^{2q}} \left(\mathbb{E}|\epsilon_{k2}|^{2q} + \mathbb{E}\sum_j |\mathcal{G}^{(kl)}_{jj} x_{kj} x_{lj}|^{2q} \right) \leq \frac{(Cq|\sqrt{\theta}|)^{cq}}{(N\eta)^q} \tag{6.5}$$

where $\mathbb{E}|\epsilon_{k2}|$ is bounded using (4.7), (4.6) and $\mathbb{E}\sum_j |\mathcal{G}^{(kl)}_{jj} x_{kj} x_{lj}|^{2q}$ is bounded by Rosenthal's inequality like $\mathbb{E}|\epsilon_{k1}|^{2q}$ in (4.4). We also use Lemma 5.1 to bound $\mathbb{E}|\mathcal{G}_{kk}|^{2q}$. Now using (6.5), (6.3), and Lemma 5.1 we obtain that

$$\mathbb{E}|G_{kl}|^{2q} \leq \frac{(Cq)^{cq}}{(N\eta)^q}. \tag{6.6}$$

To improve the bound (4.1), we see that using equation (4.2) and (6.6) as well as Lemma 5.1 (also using that $\frac{2}{(N\eta)^q N^{q/2}} \leq \frac{1}{(N\eta)^{2q}} + \frac{1}{N^q}$), we obtain

$$\mathbb{E}|\Upsilon|^{2q} \leq \left(\frac{Cq}{N\eta}\right)^{cq} \mathbb{E}|\operatorname{Im}\operatorname{Tr}\mathcal{G}|^q + (Cq)^{cq}\left(\frac{1}{N^q} + \frac{1}{(N\eta)^{2q}}\right) \tag{6.7}$$

and using (6.6) we can improve the bound on $\mathbb{E}|\epsilon_{2k}|^{2q}$ in (4.7), which yields (6.4).

□

Lemma 6.2. *Assume* (1.6) *and* (1.7) *for the entries of X_N as before and let $\theta = E + i\eta \in S_{E,\eta}$. There exist constants $c, C, M > 0$ such that*

$$\mathbb{E}\frac{1}{|\sqrt{\theta}G^{(\mathbb{J})}_{11}|^{2q}} \leq C^q,$$

for $\theta \in S_{E,\eta}$, $N\eta > |\sqrt{\theta}|M$, $q \leq c(N\eta)^{1/4}$ and $\mathbb{J} \subset \{1, ..., N\}$, with $|\mathbb{J}| \leq 2q$.

Proof. We can take $\mathbb{J} = \emptyset$ as the argument is similar in the general case. We have that:

$$\begin{aligned}\mathbb{E}\frac{1}{|\sqrt{\theta}G_{11}|^{2q}} &= \mathbb{E}|\sqrt{\theta}(1 + (\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1/N)|^{2q} \leq C^q + (C|\theta|)^q\mathbb{E}|(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1/N|^{2q} \\ &\leq C^q(1 + |\theta|^q\mathbb{E}|(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1/N - \mathbb{E}_{\mathbf{x}^1}(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1/N|^{2q} + \mathbb{E}|\mathbb{E}_{\mathbf{x}^1}\sqrt{\theta}(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1/N|^{2q}).\end{aligned}$$

The second term on the RHS is small by Lemma 4.1. For the third term, we find that:

$$\mathbb{E}|\mathbb{E}_{\mathbf{x}^1}\sqrt{\theta}(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1/N|^{2q} = \mathbb{E}\left|\frac{1}{N}\sqrt{\theta}\operatorname{Tr}(\mathcal{G}^{(1)})\right|^{2q} = \mathbb{E}\left|\frac{1}{N}\sqrt{\theta}\left(\frac{1}{\theta} + \operatorname{Tr}(G^{(1)})\right)\right|^{2q} \leq C^q, \tag{6.8}$$

where we used Lemma 5.1 and that $|\Delta^{(1)}_N - \Delta_N| \leq \frac{1}{N\eta}$ as in (3.11).

□

To estimate the third quantity in (6.1), we find by (6.7) that:

$$\left|(\mathbb{I}-\mathbb{E}_{\mathbf{x}^k})\frac{1}{\sqrt{\theta}G_{kk}}\right| = \left|-\sqrt{\theta}\Upsilon^{(\{k\})}\right| \leq (Cq)^{cq}\mathcal{E}_q. \tag{6.9}$$

Lastly, we also need a bound on $\mathbb{E}\left|\frac{1}{\mathbb{E}_{\mathbf{x}^1}\frac{1}{\sqrt{\theta}G_{11}}}\right|^q$ which we obtain in the following lemma.

Lemma 6.3. *Let $E,\eta \in S_{E,\eta}$, where $\theta = E+i\eta$. There exist constants $c, C, M > 0$ such that:*

$$\mathbb{E}\left|\frac{1}{\mathbb{E}_{\mathbf{x}^1}\frac{1}{\sqrt{\theta}G_{11}}}\right|^q \leq C^q,$$

for $N\eta \geq |\sqrt{\theta}|M$ and for $q\in\mathbb{N}$ with $q \leq c\left(\frac{N\eta}{|\sqrt{\theta}|}\right)^{1/4}$.

Proof. The proof is similar to Lemma 5.1 in [2]. We define

$$\widetilde{G_{11}} = \frac{1}{\mathbb{E}_{\mathbf{x}^1}\frac{1}{G_{11}}} = -\frac{1}{\theta(1+\operatorname{Tr}\mathcal{G}^{(\{1\})})}.$$

We calculate that

$$\left|\frac{d}{d\eta}\log\widetilde{G_{11}}(E+i\eta)\right| = \left|\frac{d}{d\eta}\log\left(\frac{1}{\theta}\right)+\frac{d}{d\eta}\log\left(\frac{1}{1+\operatorname{Tr}\mathcal{G}^{(\{1\})}}\right)\right| = \left|-\frac{i}{\theta}-\frac{\frac{d}{d\eta}\operatorname{Tr}\mathcal{G}^{(\{1\})}}{1+\operatorname{Tr}\mathcal{G}^{(\{1\})}}\right|.$$

We show that $\left|\frac{d}{d\eta}\operatorname{Tr}\mathcal{G}^{(\{1\})}\right| \leq \frac{\operatorname{Im}\operatorname{Tr}\mathcal{G}^{(\{1\})}}{\eta}$ as follows:

$$\frac{d}{d\eta}\operatorname{Tr}\mathcal{G}^{(1)} = \sum_{k=1}^N \frac{d}{d\eta}\mathcal{G}^{(\{1\})}_{kk}(\theta) = \sum_{k=1}^N i((\mathcal{G}^{(\{1\})})^2)_{kk} = \sum_{k=1}^N i\langle e_k, (\mathcal{G}^{(\{1\})})^2 e_k\rangle$$

$$\Rightarrow \left|\frac{d}{d\eta}\mathcal{G}^{(\{1\})}\right| \leq \sum_{k=1}^N ((\mathcal{G}^{(\{1\})})^*\mathcal{G}^{(\{1\})})_{kk} = \sum_{k=1}^N \frac{\operatorname{Im}(\mathcal{G}^{(\{1\})})_{kk}}{\eta} = \frac{\operatorname{Im}\operatorname{Tr}\mathcal{G}^{(\{1\})}}{\eta}. \tag{6.10}$$

We conclude that

$$\left|\frac{d}{d\eta}\log\widetilde{G_{11}}\right| \leq \frac{1}{|\theta|}+\frac{\operatorname{Im}\operatorname{Tr}\mathcal{G}^{(\{1\})}}{\eta|1+\operatorname{Tr}\mathcal{G}^{(\{1\})}|} \leq \frac{2}{\eta}, \tag{6.11}$$

yielding that

$$\left|\log\widetilde{G_{11}}(E+i\eta)-\log\widetilde{G_{11}}(E+i\eta/s)\right| = \left|\int_{\eta/s}^{\eta}\frac{d}{d\nu}\log\widetilde{G_{11}}(E+i\nu)d\nu\right| \leq \int_{\eta/s}^{\eta}\frac{2}{\nu}d\nu = \log s^2 \tag{6.12}$$

and thus $|\widetilde{G_{11}}(E+i\eta)| \leq s^2|\widetilde{G_{11}}(E+i\eta/s)|$. The proof now proceeds by induction on η just like in the proof of Lemma 5.1 using the identity

$$\sqrt{\theta}\widetilde{G_{11}} = \sqrt{\theta}G_{11}+\sqrt{\theta}G_{11}\sqrt{\theta}\widetilde{G_{11}}(\mathbb{I}-\mathbb{E}_{\mathbf{x}^1})(\sqrt{\theta}G_{11})^{-1} \tag{6.13}$$

as well as (6.9) and the results of Lemma 5.1. □

Lastly, we use the matrix expansion algorithm to take advantage of the fluctuations. Hence the following proposition, analogous to Lemma 4.1 of [2]:

Proposition 6.4. *Let $\mathcal{E}_q$ be the control parameter as in (6.2). There exist constants $C, M, c_0 > 0$ such that*

$$\mathbb{E}\left|\frac{1}{N}\sum_k \sqrt{\theta}\Upsilon^{(\{k\})}\sqrt{\theta}G_{kk}\right|^{2q} \leq (Cq)^{cq^2}\mathcal{E}_{4q}^{1/2}, \tag{6.14}$$

for $1 \leq q \leq c_0\left(\frac{N\eta}{|\sqrt{\theta}|}\right)^{1/8}$, $\frac{N\eta}{\sqrt{|\theta|}} \geq M$, $K>0$, $\theta = E+i\eta \in S_{E,\eta}$.

Proof. To match notation in [2], we introduce $W_k = \sqrt{\theta}\Upsilon_k\sqrt{\theta}G_{kk}$ and we split:

$$\frac{1}{N}\sum_k W_k = \frac{1}{N}\sum_k (\mathbb{I}-\mathbb{E}_k)W_k + \frac{1}{N}\sum_k \mathbb{E}_k W_k.$$

By Hölder's inequality,

$$\mathbb{E}\left|\frac{1}{N}\sum_k W_k\right|^{2q} \le C^q\mathbb{E}\left|\frac{1}{N}\sum_k(\mathbb{I}-\mathbb{E}_k)W_k\right|^{2q} + C^q\mathbb{E}|\mathbb{E}_1W_1|^{2q}. \tag{6.15}$$

To bound the second term in (6) above, using that $\sqrt{\theta}\Upsilon^{(\{k\})} = -(\mathbb{I}-\mathbb{E}_k)\frac{1}{\sqrt{\theta}G_{kk}}$, we obtain

$$\mathbb{E}_kW_k = \frac{\mathbb{E}_k[\sqrt{\theta}G_{kk}(\sqrt{\theta}\Upsilon^{(\{k\})})^2]}{\left(\mathbb{E}_k\frac{1}{\sqrt{\theta}G_{kk}}\right)}. \tag{6.16}$$

and applying Lemma 6.3 to (6.16), we get that:

$$\begin{aligned}\mathbb{E}|\mathbb{E}_1W_1|^{2q} &\le (\mathbb{E}|\sqrt{\theta}G_{11}|^{8q})^{\frac14}\left(\mathbb{E}\left|\frac{1}{\mathbb{E}_1\frac{1}{\sqrt{\theta}G_{11}}}\right|^{8q}\right)^{\frac14}(\mathbb{E}|\sqrt{\theta}\Upsilon^{(\{1\})}|^{8q})^{\frac12}\\ &\le (Cq)^{cq}\left(\frac{|\theta|^{4q}}{(N\eta)^{8q}}+\frac{(\mathrm{Im}\ |\theta|\Delta))^{4q}+\mathbb{E}|\theta\Lambda|^{4q}}{(N\eta)^{4q}}\right)^{\frac12},\end{aligned}$$

which is what we want.

In order to handle the first term of (6), we use the matrix expansion algorithm as in Section 5.2 of [2]. We notice that equations (5.7), (5.8), and (5.9) are the basis of the expansion algorithm, and they are equivalent to the following (see e.g. (2.18) in [12]):

$$\begin{aligned}\sqrt{\theta}G_{ij}^{(\mathbb{T})} &= \sqrt{\theta}G_{ij}^{(\mathbb{T}k)} + \frac{\sqrt{\theta}G_{ik}^{(\mathbb{T})}\sqrt{\theta}G_{kj}^{(\mathbb{T})}}{\sqrt{\theta}G_{kk}^{(\mathbb{T})}} \text{ for } i,j,k\notin\mathbb{T} \text{ and } i,j\ne k,\\ \frac{1}{\sqrt{\theta}G_{ii}^{(\mathbb{T})}} &= \frac{1}{\sqrt{\theta}G_{ii}^{(\mathbb{T}k)}} - \frac{\sqrt{\theta}G_{ik}^{(\mathbb{T})}\sqrt{\theta}G_{ki}^{(\mathbb{T})}}{\sqrt{\theta}G_{ii}^{(\mathbb{T})}\sqrt{\theta}G_{ii}^{(\mathbb{T}k)}\sqrt{\theta}G_{kk}^{(\mathbb{T})}} \text{ for } i,k\notin\mathbb{T} \text{ and } i\ne k\end{aligned} \tag{6.17}$$

Using the above equation (6.17), we see that in our case the steps of the expansion algorithm (5.13), (5.14), (5.15) in [2] are the same except that each resolvent entry is multiplied by a factor of $\sqrt{\theta}$. Using our definition of W, equation (5.6) in [2] becomes analogous to

$$(\mathbb{I}-\mathbb{E}_{k_s})W_{k_s} = (\mathbb{I}-\mathbb{E}_{k_s})\left[(\mathbb{I}-\mathbb{E}_{k_s})\frac{1}{\sqrt{\theta}G_{k_sk_s}}\right]\sqrt{\theta}G_{k_sk_s}, s=1,...,2q, \tag{6.18}$$

so the initial terms of the algorithm are $A^r := \sqrt{\theta}G_{k_rk_r}$ and $B^r := \frac{1}{\sqrt{\theta}G_{k_rk_r}}$ are the same as (5.16), (5.17) of [2] except that each resolvent entry is multiplied by a $\sqrt{\theta}$. Then (5.18), (5.19), and (5.20) of [2] carry over directly as well as properties (1) through (5) of relevant strings. We then obtain the desired result

$$\mathbb{E}\left|\frac{1}{N}\sum_k(\mathbb{I}-\mathbb{E}_k)W_k\right|^{2q} \le (Cq)^{cq^2}\mathcal{E}_{4q}^{1/2}$$

using the proof of (5.32) of [2]. It relies on counting the types of terms that result from the expansion algorithm. Since our algorithm yields the same type and number of terms in each step, the proof in our case will be identical. In [2], we notice the use of bounds (3.9) and Lemma 5.2 in (5.44) as well as in Case 2, bounds (5.26) and (3.4) in (5.43) and (5.49). We can replace (3.9), Lemma 5.2, (5.26), and (3.4) of [2] by our bounds on the relevant quantities in (6.1) as well as our (5.4).

□

Proof of Theorem 1. By Proposition 3.2, in order to control Λ, we need to control high moments of $R = N^{-1}\sum_{k=1}^{N} G_{kk}(T_k + \Upsilon^{(\{k\})})$. Taking expectation of $2q$ power we obtain

$$\mathbb{E}|\theta R|^{2q} \leq C^q \left(\mathbb{E}\left| \frac{1}{N}\sum_k \sqrt{\theta}T_k\sqrt{\theta}G_{kk}\right|^{2q} + \mathbb{E}\left| \frac{1}{N}\sum_k \sqrt{\theta}\Upsilon^{(\{k\})}\sqrt{\theta}G_{kk}\right|^{2q}\right). \tag{6.19}$$

For the first term by (3.11), we obtain

$$\mathbb{E}\left| \frac{1}{N}\sum_k \sqrt{\theta}T_k\sqrt{\theta}G_{kk}\right|^{2q} \leq C^q \frac{1}{N^{2q}|\theta|^q}. \tag{6.20}$$

while the second term is handled in Proposition 6.4, yielding that

$$\mathbb{E}|\theta R|^{2q} \leq (Cq)^{cq^2}\mathcal{E}_{4q}^{1/2}.$$

Here we are able simplify the analysis in [2] by only using the bounds proportional to R from Proposition 3.2 to control $E|\Lambda|^{2q}$ on $S_{E,\eta}$ and $E|\operatorname{Im}\Lambda|^{2q}$. Our simplifications carry over also to the Wigner case. We can assume that

$$[\operatorname{Im}(|\theta|\Delta)]^{2q} + \mathbb{E}|\theta\Lambda|^{2q} \geq \frac{|\theta|^{2q}}{(N\eta)^{2q}},$$

(otherwise $\mathbb{E}|\Lambda|^{2q} \leq \frac{1}{(N\eta)^{2q}}$, as we want) and in this case:

$$\mathcal{E}_{2q} = \frac{1}{N^{2q}|\theta|^q} + \frac{\operatorname{Im}(|\theta|\Delta)]^{2q} + \mathbb{E}|\theta\Lambda|^{2q}}{(N\eta)^{2q}} \leq \frac{\eta^q + \operatorname{Im}(|\theta|\Delta)]^{2q} + \mathbb{E}|\theta\Lambda|^{2q}}{(N\eta)^{2q}},$$

Using the bound proportional to $|R|$ from Proposition 3.2, we obtain

$$\begin{aligned}\mathbb{E}|\theta\Lambda|^q &\leq \frac{C^q\mathbb{E}|\theta R|^q}{|\Delta+\frac{1}{2}|^q} \leq \frac{(Cq)^{cq^2}}{|\Delta+\frac{1}{2}|^q}\left(\frac{\eta^q + [\operatorname{Im}(|\theta|\Delta)]^{2q}}{(N\eta)^{2q}}\right)^{1/2} = \frac{(Cq)^{cq^2}}{|\Delta+\frac{1}{2}|^q}\frac{|\theta|^q}{(N\eta)^q}\left(\frac{\eta^q}{|\theta|^{2q}} + [\operatorname{Im}(\Delta)]^{2q}\right)^{1/2}\\ &\leq \frac{(Cq)^{cq^2}|\theta|^q}{(N\eta)^q}\left[\left(\frac{\sqrt{\eta}}{|\theta||\Delta+\frac{1}{2}|}\right)^q + \left(\frac{\operatorname{Im}\Delta}{|\Delta+\frac{1}{2}|}\right)^q\right].\end{aligned}$$

To obtain the desired bound we now note that $\operatorname{Im}\Delta \leq |\Delta+\frac{1}{2}|$ and $\frac{\sqrt{\eta}}{|\theta||\Delta+\frac{1}{2}|} \leq C$ on our domain. The first one follows easily and for the second one we argue as follows:

$$\frac{\sqrt{\eta}}{|\theta||\Delta+\frac{1}{2}|} = \frac{2\sqrt{\eta}}{\sqrt{|\theta|}\sqrt{|\theta-4|}},$$

and by triangle inequality either $|\theta| \geq 2$ or $|\theta - 4| \geq 2$. Then in the first case, we use the bound $\sqrt{\eta} \leq \sqrt{|\theta-4|}$ and in the second case the bound $\sqrt{\eta} \leq \sqrt{|\theta|}$.
Overall, this implies that

$$\mathbb{P}\left(|\Delta_N - \Delta| \geq \frac{K}{N\eta}\right) \leq \frac{(N\eta)^q}{K^q}\mathbb{E}|\Lambda|^q \leq \frac{(Cq)^{cq^2}}{K^q}, \tag{6.21}$$

for $1 \leq q \leq c_0\left(\frac{N\eta}{|\sqrt{\theta}|}\right)^{1/8}$, $\frac{N\eta}{\sqrt{|\theta|}} \geq M$, $K > 0$, $\theta = E + i\eta \in S_{E,\eta}$.

□

7. Convergence of the counting function

In this section we prove Theorem 2.

Proof of Theorem 2. Let $0 < E \le 4$. We will use a Pleijel argument from [13], recently used in obtaining estimates on a measure from estimates on a Stieltjes transform in [4]. We start from the following equations (equations (13) and (14) in [4], following from equation (5) of [13]):

$$\mu(-K,E) = \frac{1}{2\pi i}\int_{L(z_0)} m_\mu(z)dz + \frac{\eta_0}{\pi}\,\mathrm{Re}\; m_\mu(z_0) + O(\eta_0\,\mathrm{Im}\; m_\mu(z_0)) \tag{7.1}$$

and

$$\mu(x,x') = \frac{1}{2\pi i}\int_{\gamma(x,x')} m_\mu(z)dz + O(\eta_0(|m_\mu(x+i\eta_0)| + |m_\mu(x'+i\eta_0)|)) \tag{7.2}$$

where m_μ is the Stieltjes transform of μ and $L(z_0)$ is a contour as in Figure 1 (see also [4] Fig 1A), namely connects with line segments the points $E-i\eta_0, E-iQ, -1-iQ, -1+iQ, E+iQ, E+i\eta_0$ in that order with an arbitrarily chosen constants -1 and Q, and $\gamma(x,x')$ is the contour connecting $x+i\eta_0, x+iQ, x'+iQ$, and $x'+i\eta_0$ in that order.

By Markov's inequality we obtain that

$$\mathbb{P}\left(|n_N(E) - n_{MP}(E)| \ge \frac{K\log N}{N}\right) \le \frac{N^q\mathbb{E}(|n_N(E)-n_{MP}(E)|^q)}{(K\log N)^q} \tag{7.3}$$

Then using (7.1) and taking $z_0 := E + i\eta_0$ with $\eta_0 := \frac{M\sqrt{E}}{N}$ with M as in Theorem 1 we obtain that

$$\begin{aligned}\mathbb{E}(|n_N(E)-n_{MP}(E)|^q) &= \mathbb{E}\left|\frac{1}{2\pi i}\int_{L(z_0)}\Lambda(z)dz + \frac{\eta_0}{\pi}\,\mathrm{Re}\;\Lambda(z_0) + O\left(\eta_0(\mathrm{Im}\;\Delta_N(z_0) + \mathrm{Im}\;\Delta_{MP}(z_0))\right)\right|^q \\ &\le C^q\left(\mathbb{E}\left|\int_{L(z_0)}\Lambda(z)dz\right|^q + O\left(\eta_0^q\mathbb{E}|\Lambda(z_0)|^q + \eta_0^q\,\mathrm{Im}\;\Delta_{MP}(z_0)^q\right)\right),\end{aligned} \tag{7.4}$$

noting that the constant in the O comes from the Pleijel formula and is uniform in the matrix randomness. We study the above expression one term at a time. For $E \le 4$ we can bound the second term as follows

$$\eta_0^q\mathbb{E}|\Lambda(z_0)|^q \le \eta_0^q\frac{Cq^{q^2}}{(N\eta_0)^q} \le \frac{Cq^{q^2}}{N^q}. \tag{7.5}$$

The third term is bounded using the above inequality (7.5) on Λ as well as

$$\eta_0\,\mathrm{Im}\;\Delta_{MP} \le \frac{C\eta_0}{\sqrt{E}} \le \frac{CM}{N}. \tag{7.6}$$

Now for the integral, we note that it suffices to study the part of the contour where $\mathrm{Im}\; z > 0$ since $\Lambda(\bar{z}) = \overline{\Lambda(z)}$. Thus we obtain

$$\mathbb{E}\left|\int_{L(z_0)}\Lambda(z)dz\right|^q \le C^q\left(\mathbb{E}\left|\int_0^{\eta_0}\Lambda(-1+iy)dy\right|^q + \mathbb{E}\left|\int_{\eta_0}^{Q}\Lambda(-1+iy)-\Lambda(E+iy)dy\right|^q + \mathbb{E}\left|\int_{-1}^{E}\Lambda(x+iQ)dx\right|^q\right) \tag{7.7}$$

Since all eigenvalues are positive we bound Λ for $-1 < 0$ by $\Lambda(-1+i\eta) < 2$ which yields

$$\left|\int_0^{\eta_0}\Lambda(-1+iy)dy\right|^q \le \left(\int_0^{\eta_0}|\Lambda(-1+iy)|dy\right)^q \le C^q\eta_0^q. \tag{7.8}$$

Next we note that

$$\mathbb{E}\left|\int_{-1}^{E}\Lambda(x+iQ)dx\right|^q \le \frac{(Cq)^{q^2}}{(NQ)^q} \tag{7.9}$$

Now we can bound the expected value of the integrals $\mathbb{E}\left(\int_{\eta_0}^{Q}|\Lambda(E+iy)|dy\right)^q$ and $\mathbb{E}\left(\int_{\eta_0}^{Q}|\Lambda(-1+iy)|dy\right)^q$ for $E \le 4$, noting that the argument is identical at E and -1,

$$\mathbb{E}\left(\int_{\eta_0}^{Q}|\Lambda(E+iy)|dy\right)^q = \mathbb{E}\int_{\eta_0}^{Q}|\Lambda(E+iy_1)|dy_1\int_{\eta_0}^{Q}|\Lambda(E+iy_2)|dy_2\cdots\int_{\eta_0}^{Q}|\Lambda(E+iy_q)|dy_q$$
$$= \mathbb{E}\int_{\eta_0}^{Q}\cdots\int_{\eta_0}^{Q}\prod_{j=1}^{q}|\Lambda(E+iy_j)|\prod_{j=1}^{q}dy_j = \int_{\eta_0}^{Q}\cdots\int_{\eta_0}^{Q}\mathbb{E}\prod_{j=1}^{q}|\Lambda(E+iy_j)|\prod_{j=1}^{q}dy_j$$
$$\le \int_{\eta_0}^{Q}\cdots\int_{\eta_0}^{Q}\prod_{j=1}^{q}(\mathbb{E}|\Lambda(E+iy_j)|^q)^{\frac{1}{q}}\prod_{j=1}^{q}dy_j \le \frac{1}{N^q}\int_{\eta_0}^{Q}\cdots\int_{\eta_0}^{Q}\prod_{j=1}^{q}\frac{(Cq)^{cq}}{y_j}\prod_{j=1}^{q}dy_j$$
$$= \frac{(Cq)^{cq^2}}{N^q}\left(\int_{\eta_0}^{Q}\frac{1}{y}dy\right)^q \le (Cq)^{cq^2}\frac{(\log N)^q}{N^q}$$

where we can apply (6.21) inside the integral because our estimates on Λ are uniform on compact sets.

To prove the second part of (1.17), we use the (7.2) and study the interval $[-E,E]$, noting that $n_N(E) = \mathcal{N}([-E,E])/N$ and $n_{MP}(E) = n_{MP}(E) - n_{MP}(-E)$. The corresponding integral can be bounded similar to above

$$\mathbb{E}\left|\int_{-E}^{E}\Lambda(x+i\eta_0)-\Lambda(x-i\eta_0)dx\right|^q = \mathbb{E}\left|\int_{-E}^{E}2\,\mathrm{Im}\,\Lambda(x+i\eta_0)\right|^q$$
$$= \int_{-E}^{E}\cdots\int_{-E}^{E}\mathbb{E}\prod_{j=1}^{q}|2\,\mathrm{Im}\,\Lambda(x_j+i\eta_0)|dx_1\cdots dx_q \le \frac{(Cq)^{cq^2}E^q}{(N\eta_0)^q} \le \frac{(Cq)^{cq^2}(\sqrt{E})^q}{M^q} \tag{7.10}$$

and, similar to (7.6)

$$\max\{\eta_0\Delta_{MP}(-E),\eta_0\Delta_{MP}(E)\} \le \frac{\eta_0}{\sqrt{E}} \le \frac{M}{N} \tag{7.11}$$

which together with (7.5) yields the second part of (1.17) for $E < 4$.
To establish the (1.17) for $E > 4$, we use (1.17) for $E = 4$ to establish bounds on the number of eigenvalues outside the spectrum. Letting $\mathcal{N}_I$ be the number of eigenvalues in an interval I, we see that

$$\mathcal{N}_{(4,\infty)} = N - Nn(4) = N(n_{MP}(4) - n(4)) \tag{7.12}$$

which by (1.17) for $E = 4$ yields that

$$\mathbb{P}\left(\frac{\mathcal{N}_{(4,\infty)}}{N} > \frac{K\log N}{N}\right) \le \frac{(Cq)^{q^2}}{K^q} \tag{7.13}$$

and for $E > 4$,

$$\mathbb{P}\left(|n_N(E) - n_{MP}(E)| \ge \frac{K\log N}{N}\right) \le \mathbb{P}\left(\frac{\mathcal{N}_{(4,\infty)}}{N} > \frac{K\log N}{N}\right) \tag{7.14}$$

thus (7.13) gives the desired bound. □

8. Rigidity of the eigenvalues

The aim of this section is a proof of Theorem 3.

Proof of Theorem 3. Let $\alpha \le \frac{N}{2}$. We will make use of the following inequalities near the hard edge and away from the soft edge:

$$c\sqrt{x} \le n_{MP}(x) \le C\sqrt{x},$$

and

$$cn_{MP}(x)^{-1} \le \rho(x) \le Cn_{MP}(x)^{-1}.$$

valid for $x \in (0,3]$. The second inequality implies that

$$c\frac{N}{a} \le \rho(\gamma_a) \le C\frac{N}{a} \tag{8.1}$$

for any $a \leq \frac{N}{2}$.
For $\varepsilon > 0$, we have that

$$\begin{aligned}
&\mathbb{P}\left(|\lambda_a - \gamma_a| \geq K\epsilon \frac{a}{N}\right) \\
&\qquad \leq \mathbb{P}\left(|\lambda_a - \gamma_a| \geq K\varepsilon\frac{a}{N} \text{ and } \lambda_a \leq \gamma_a\right) + \mathbb{P}\left(|\lambda_a - \gamma_a| \geq K\varepsilon\frac{a}{N} \text{ and } \lambda_a > \gamma_a\right) \\
&\qquad = A + B\ .
\end{aligned}$$

We consider first the term A. We set

$$\ell = K\varepsilon\frac{a}{N}.$$

From $\lambda_a \leq \gamma_a$ and $|\lambda_a - \gamma_a| \geq \ell$ we find that $\lambda_a \leq \gamma_a - \ell$. This implies that $n_N(\gamma_a - \ell) \geq \frac{a}{N} = n_{MP}(\gamma_a)$. By the mean value theorem for the function n_{MP}, there exists a point $x^* \in [\gamma_a - \ell, \gamma_a]$ such that $n_{MP}(\gamma_a) - n_{MP}(\gamma_a - \ell) = \rho(x^*)\ell$, yielding that

$$n_N(\gamma_a - \ell) - n_{MP}(\gamma_a - \ell) \ = \ n_N(\gamma_a - \ell) - n_{MP}(\gamma_a) + \rho(x^*)\ell \ \geq \ \rho(x^*)\ell \ \geq \ \rho(\gamma_a)K\varepsilon\frac{a}{N} \ \geq \ cK\varepsilon, \quad (8.2)$$

because ρ is non-increasing, $a < N/2$, and from (8.1). Setting $\varepsilon = \frac{\log N}{N}$ we deduce from Theorem 2 that

$$A \leq \mathbb{P}\left(|n_N(\gamma_a - \ell) - n_{MP}(\gamma_a - \ell)| \geq \frac{cK\log N}{N}\right) \leq \frac{(Cq)^{cq^2}}{K^q} \quad (8.3)$$

For $a \leq \log N$, set $\varepsilon = \frac{a}{N} \geq c\sqrt{\gamma_a}$ to obtain

$$A \leq \mathbb{P}\left(|n_N(\gamma_a - \ell) - n_{MP}(\gamma_a - \ell)| \geq cK\sqrt{(\gamma_a - \ell)_+}\right) \leq \frac{(Cq)^{cq^2}}{K^q}. \quad (8.4)$$

We now estimate the term B. From the estimate $n_{MP}(x) \sim \sqrt{x}$ near the hard edge, we have that

$$\gamma_a \leq C\left(\frac{a}{N}\right)^2,$$

for some constant $C > 0$ for all $a < N/2$. We consider the number

$$y = 2C\left(\frac{a}{N}\right)^2$$

and we further consider the cases that $\gamma_a + \ell \leq y$ or $\gamma_a + \ell > y$.

In the first case since $\lambda_a > \gamma_a$ and $|\lambda_a - \gamma_a| \geq \ell$, we have that $\lambda_a > \gamma_a + \ell$ and so $n_N(\gamma_a + \ell) \leq \frac{a}{N} = n_{MP}(\gamma_a)$. Hence, from the mean value theorem, we find $x^* \in [\gamma_a, \gamma_a + \ell] \subset [\gamma_a, y]$ such that $n_{MP}(\gamma_a + \ell) - n_{MP}(\gamma_a) = \rho(x^*)\ell$, yielding that

$$n_{MP}(\gamma_a + \ell) - n_N(\gamma_a + \ell) = n_{MP}(\gamma_a) - n_N(\gamma_a + \ell) + \rho(x^*)\ell \geq \rho(x^*)\ell = \rho(x^*)K\varepsilon\frac{a}{N} \geq \rho(y)K\varepsilon\frac{a}{N} \geq cK\varepsilon,$$

where we used that ρ is nonincreasing and that $\rho(y) \geq \frac{c}{\sqrt{y}}$ near the hard edge. Setting $\varepsilon = \frac{\log N}{N}$ and using Theorem 2, we conclude that

$$B \leq \mathbb{P}\left(|n_{MP}(\gamma_a + \ell) - n_N(\gamma_a + \ell)| \geq cK\frac{\log N}{N}\right) \leq \frac{(Cq)^{cq^2}}{K^q}, \quad (8.5)$$

as required. For rigidity at the hard edge equation (1.19), let $\varepsilon = \frac{a}{N}$ to obtain

$$\begin{aligned}
B &\leq \mathbb{P}\left(|n_{MP}(\gamma_a + \ell) - n_N(\gamma_a + \ell)| \geq \frac{cKa}{N}\right) \\
&\qquad \leq \mathbb{P}\left(|n_{MP}(\gamma_a + \ell) - n_N(\gamma_a + \ell)| \geq c\sqrt{K}\sqrt{\gamma_a + \ell}\right) \leq \frac{(Cq)^{cq^2}}{K^{q/2}},
\end{aligned} \quad (8.6)$$

where the second line follows as before because $\sqrt{\gamma_a} \le c\frac{a}{N}$ and $\ell = \frac{Ka^2}{N^2}$.

In the other case we have that $\gamma_a + \ell > y$ so the inequality $\lambda_a > \gamma_a + \ell$ implies that $\lambda_a > y$ and therefore $n_N(y) \le \frac{a}{N} = n_{MP}(\gamma_a)$. Hence from the mean value theorem there exists $x^* \in [\gamma_a, y]$ such that $n_{MP}(y) - n_{MP}(\gamma_a) = \rho(x^*)\ell$, which yields

$$n_{MP}(y) - n_N(y) = n_{MP}(\gamma_a) - n_N(y) + \rho(x^*)\ell \ge \rho(x^*)\ell = \rho(x^*)K\varepsilon\frac{a}{N} \ge \rho(y)K\varepsilon\frac{a}{N} \ge cK\varepsilon,$$

and we can conclude (1.18) and (1.19) as above. This finishes the proof of Theorem 3. □

Acknowledgements. Anastasis Kafetzopoulos and Anna Maltsev acknowledge the support of the Royal Society grant numbers RGF\R1\181001 and URF UF160569, respectively.

References

[1] Claudio Cacciapuoti, Anna Maltsev, and Benjamin Schlein. Local Marchenko-Pastur law at the hard edge of sample covariance matrices. *Journal of Mathematical Physics*, 54(4):043302, 2013.

[2] Claudio Cacciapuoti, Anna Maltsev, and Benjamin Schlein. Bounds for the Stieltjes transform and the density of states of wigner matrices. *Probability Theory and Related Fields*, 163(1-2):1–59, 2015.

[3] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin. The local relaxation flow approach to universality of the local statistics for random matrices. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1–46, 2012.

[4] László Erdős, Dominik Schröder, et al. Fluctuations of functions of Wigner matrices. *Electronic Communications in Probability*, 21, 2016.

[5] Friedrich Götze and Alexander Tikhomirov. Optimal bounds for convergence of expected spectral distributions to the semi-circular law. *Probability Theory and Related Fields*, 165(1-2):163–233, 2016.

[6] Friedrich Götze and AN Tikhomirov. Rate of convergence of the expected spectral distribution function to the Marchenko–Pastur law. *arXiv preprint arXiv:1412.6284*, 2014.

[7] Naumov A.A. Tikhomirov A.N. Gotze, A. Moment inequalities for linear and nonlinear statistics. *Theory Probab. Appl*, 65(1):1–16, 2020.

[8] Jonas Gustavsson. Gaussian fluctuations of eigenvalues in the GUE. In *Annales de l'IHP Probabilités et statistiques*, volume 41, pages 151–178, 2005.

[9] Benjamin Landon, Patrick Lopatto, and Philippe Sosoe. Single eigenvalue fluctuations of general Wigner-type matrices. *arXiv preprint arXiv:2105.01178*, 2021.

[10] Benjamin Landon, Philippe Sosoe, et al. Applications of mesoscopic CLTs in random matrix theory. *Annals of Applied Probability*, 30(6):2769–2795, 2020.

[11] Vladimir Alexandrovich Marchenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967.

[12] Natesh S Pillai, Jun Yin, et al. Universality of covariance matrices. *The Annals of Applied Probability*, 24(3):935–1001, 2014.

[13] Åke Pleijel. On a theorem by P. Malliavin. *Israel Journal of Mathematics*, 1(3):166–168, 1963.

[14] Zhonggen Su. Gaussian fluctuations in complex sample covariance matrices. *Electronic Journal of Probability*, 11:1284–1320, 2006.

[15] Terence Tao and Van Vu. Random matrices: The distribution of the smallest singular values. *Geometric And Functional Analysis*, 20(1):260–297, 2010.

[16] Terence Tao, Van Vu, et al. Random covariance matrices: Universality of local statistics of eigenvalues. *The Annals of Probability*, 40(3):1285–1315, 2012.

[17] Ke Wang. Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices: Theory and Applications*, 1(01):1150005, 2012.