Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2021, Vol. 57, No. 1, 506–546
https://doi.org/10.1214/20-AIHP1086

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES

Central limit theorem for mesoscopic eigenvalue statistics of deformed Wigner matrices and sample covariance matrices

Yiting Li[*], Kevin Schnelli[1,†] and Yuanyuan Xu[2,‡]

KTH Royal Institute of Technology, Stockholm, Sweden. E-mail: [*]yitingl@kth.se; [†]schnelli@kth.se; [‡]yuax@kth.se

Received 11 November 2019; revised 24 June 2020; accepted 3 July 2020

Abstract. We consider N by N deformed Wigner random matrices of the form $X_N = H_N + A_N$, where H_N is a real symmetric or complex Hermitian Wigner matrix and A_N is a deterministic real bounded diagonal matrix. We prove a universal Central Limit Theorem for the linear eigenvalue statistics of X_N for all mesoscopic scales both in the spectral bulk and at regular edges where the global eigenvalue density vanishes as a square root. The method relies on studying the characteristic function of the linear statistics (Landon and Sosoe (2018)) by using the cumulant expansion method, along with local laws for the Green function of X_N (*Ann. Probab.* **48** (2020) 963–1001; *Probab. Theory Related Fields* **169** (2017) 257–352; *J. Math. Phys.* **54** (2013) 103504) and analytic subordination properties of the free additive convolution (Dallaporta and Fevrier (2019); *Random Matrices Theory Appl.* **9** (2020) 2050011). We also prove the analogous results for high-dimensional sample covariance matrices.

Résumé. Nous considérons des matrices aléatoires de Wigner déformées de taille N de la forme $X_N = H_N + A_N$, où H_N est une matrice hermitienne de Wigner symétrique ou complexe réelle, et A_N est une matrice diagonale déterministe avec des entrées réelles et bornées. Nous prouvons un théorème de limite centrale universel pour les statistiques linéaires des valeurs propres de X_N pour toutes les échelles mésoscopiques à la fois dans le centre de spectre et aux bords réguliers où la densité globale des valeurs propres disparait sous forme de racine carrée. La méthode repose sur l'étude de la fonction caractéristique des statistiques linéaires (Landon and Sosoe (2018)) en utilisant la méthode des cumulants, ainsi que les lois locales pour la fonction de Green de X_N (*Ann. Probab.* **48** (2020) 963–1001; *Probab. Theory Related Fields* **169** (2017) 257–352; *J. Math. Phys.* **54** (2013) 103504) et les propriétés de subordination analytique de la convolution libre additive (Dallaporta and Fevrier (2019); *Random Matrices Theory Appl.* **9** (2020) 2050011). Nous prouvons également les résultats analogues pour des matrices de corrélation empirique de haute dimension.

MSC2020 subject classifications: 15B52; 60B20

Keywords: Linear eigenvalue statistics; Deformed Wigner matrices; Sample covariance matrices

1. Introduction

1.1. *Linear eigenvalue statistics of Wigner matrices*

A Wigner matrix H_N is an $N \times N$ real symmetric or complex Hermitian random matrix with independent entries up to the constraint $H_N = H_N^*$. In the case the entries are Gaussian random variables, these matrices belong to the Gaussian Orthogonal Ensemble (GOE), Gaussian Unitary Ensemble (GUE), respectively. Wigner [70] proved the semicircle law stating that the empirical eigenvalue distribution of H_N converges to the semicircle distribution with density $\rho_{\mathrm{sc}}(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}$. That is, for any test function $f \in C_c(\mathbb{R})$,

$$\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) \to \int_{\mathbb{R}} f(x)\rho_{\mathrm{sc}}(x)\,\mathrm{d}x \quad \text{as } N \to \infty,$$

in probability, which can be understood as a Law of Large Numbers.

[1]Supported in parts by the Swedish Research Council Grant VR-2017-05195.
[2]Supported by the Göran Gustafsson Foundation and the Swedish Research Council Grant VR-2017-05195.

Johansson [40] obtained the corresponding Central Limit Theorem (CLT) for such linear eigenvalue statistics of the GUE, i.e.,

$$\sum_{i=1}^{N} f(\lambda_i) - N \int_{\mathbb{R}} f(x)\rho_{\mathrm{sc}}(x)\,\mathrm{d}x \tag{1.1}$$

converges in distribution to a centered Gaussian random variable. Strikingly different from the classical CLT, the linear statistics need not be normalized by $N^{-\frac{1}{2}}$, which is a manifestation of the strong eigenvalue correlations. Bai and Yao [7] used a martingale method to prove such CLTs for Wigner matrices and analytic test functions. Lytova and Pastur [54], and Shcherbina [61] improved these results by weakening the regularity conditions on the test functions. More recently, Sosoe and Wong [65] obtained CLTs for Wigner matrices with $H^{1+\epsilon}$ test functions using Littlewood–Paley decompositions.

Boutet de Monvel and Khorunzhy initiated the study of mesoscopic linear eigenvalue statistics, i.e., the derivation of Gaussian fluctuations for the random variable

$$\sum_{i=1}^{N} f\left(\frac{\lambda_i - E_0}{\eta_0}\right) - \mathbb{E}\left[\sum_{i=1}^{N} f\left(\frac{\lambda_i - E_0}{\eta_0}\right)\right], \tag{1.2}$$

with fixed $E_0 \in (-2, 2)$ on mesoscopic scales $N^{-1} \ll \eta_0 \ll 1$. In [18,19], they obtained CLTs for the test function $(x - \mathrm{i})^{-1}$ on all mesoscopic scales for the GOE, and $N^{-\frac{1}{8}} \ll \eta_0 \ll 1$ for symmetric Wigner matrices, respectively. Lodhia and Simm [53] extended the CLT for complex Wigner matrices and general test functions on scales $N^{-1/3} \ll \eta_0 \ll 1$. He and Knowles [36] used moment estimates for Green functions to prove the CLTs for all arbitrary Wigner matrices on the optimal scales $N^{-1} \ll \eta_0 \ll 1$. More recently, Landon and Sosoe [44] obtained similar CLT by means of the characteristic function.

Mesoscopic central limit theorems are important tools in the theory of homogenization of Dyson Brownian motion (DBM) introduced by Bourgade, Erdős, Yau and Yin [16] to prove fixed energy universality of local eigenvalue statistics of Wigner matrices. Landon, Sosoe and Yau [45] subsequently derived a mesoscopic CLT to show fixed energy universality of the DBM. The dynamical approach using Dyson Brownian motion to prove the universality of the eigenvalue statistics on microscopic scale for all symmetry classes was initiated by Erdős, Schlein and Yau in [30]; we refer to the surveys [32,33] for further details. Mesoscopic central limit theorems combined with DBM were used by Landon and Sosoe [44] and by Bourgade and Mody [17] to derive Gaussian fluctuations of single eigenvalues, and in [15,17] to show Gaussian fluctuations of the determinant of Wigner matrices.

Mesoscopic CLTs can also be studied at the spectral edges, where the mesoscopic scales are $N^{-\frac{2}{3}} \ll \eta_0 \ll 1$. For the GUE, Basor and Widom [9] used asymptotics of the Airy kernel to prove mesoscopic CLTs at the edges. Min and Chen [56] subsequently considered edge CLTs for the GOE. Recently, Adhikari and Huang [1] obtained the mesoscopic CLT at the edges down to the optimal scale $\eta_0 \gg N^{-\frac{2}{3}}$ for the DBM.

1.2. *Deformed Wigner matrices*

In the present paper we are interested in deformed Wigner matrices. A deformed Wigner matrix is an $N \times N$ random matrix of the form

$$X_N = H_N + A_N, \tag{1.3}$$

where H_N is a real symmetric or complex Hermitian Wigner matrix and A_N is a real deterministic diagonal matrix. It is also known as the Rosenzweig–Porter model in the physics literature. Suppose the empirical eigenvalue distribution of A_N has a deterministic limiting measure, denoted by μ_α. It was shown by Pastur [60] that the empirical eigenvalue distribution of X_N converges weakly in probability to the free additive convolution of μ_{sc} and μ_α, denoted by $\mu_{\mathrm{fc}} = \mu_{\mathrm{sc}} \boxplus \mu_\alpha$; see also [66].

A CLT for the linear eigenvalue statistics with test functions in $C_c^2(\mathbb{R})$ was obtained by Ji and Lee [39] under a one-cut assumption on μ_{fc}. They also computed the expectation and variance in terms of μ_α. Dallaporta and Fevrier [22] obtained the CLT for general μ_{fc}. Their results are summarized in Theorem 2.6 below.

In the present paper, we study the fluctuations of the linear eigenvalue statistics (1.2) in the mesoscopic regime. We choose μ_α properly such that the free additive convolution μ_{fc} is supported on a single interval and vanishes as a square root at the end-points. This edge behavior of the limiting eigenvalue distribution is quite common in random matrix theory, and sometimes referred to as regular edge. Denoting $\kappa_0 = \kappa_0(E_0)$ the distance from E_0 to the closest edge of the free

convolution measure, we derive a CLT at energy E_0 on scales η_0 with $N^{-1} \ll \eta_0\sqrt{\eta_0+\kappa_0} \le 1$; see Theorem 2.8. This range of η_0 covers the global scale as well as all mesoscopic scales up to the spectral edges. For energies E_0 in the bulk and at the edges respectively, we compute the variances and biases explicitly on the mesoscopic scales, where we recover the formulas for the Gaussian ensembles. This shows the expected universality of the linear eigenvalue fluctuations on mesoscopic scales. The universality of the eigenvalue statistics on the microscopic scale was derived for the deformed GUE in the bulk [62] and at the edge [63]. For deformed Wigner matrices, the local bulk universality was obtained in [59] for a special class of A_N using moment matching, under a one cut assumption in [52] and in [31,46] for the general case using the DBM methods. The edge universality was derived in [49] using a Green function comparison method, and in [47,52] using DBM. More recently, quantum unique ergodicity for deformed Wigner matrices was derived in [11].

In the proof of the main results, we follow the idea of [44] to compute the characteristic function of (1.2) in combination with the Helffer–Sjöstrand formula and cumulant expansions; see (3.3) below. Cumulant expansions were used in e.g. [19, 36,44] to study the linear eigenvalue statistics of random matrices. We also rely on local laws for Green functions [3,43, 48] and analytic subordination for the free convolution measure, as used in [22,39,48].

On the global scale, the method used to derive the CLT for deformed Wigner matrices [22] is insensitive to the behavior of the free convolution measure μ_{fc}. An interesting aspect of the free additive convolution and deformed Wigner matrices is that the densities may show other edge behaviors than square roots. For such setups, one expects mesoscopic CLTs at the edges with different scalings, variances and biases. This is a main motivation for us to study linear eigenvalue statistics at spectral edges. The local eigenvalue statistics at such critical edges are only partly understood, see e.g. [41,50] for some results. At cusp points in the interior of the bulk spectrum the universality of the local eigenvalue fluctuations was recently proved in [21,28].

1.3. *Sample covariance matrix*

Sample covariance matrices form another class of archetypal random matrix models, with applications in multivariate statistical analysis. We consider the separable sample covariance matrices of the form $H = \Sigma^{1/2} X X^* \Sigma^{1/2}$, where X is a $M \times N$ matrix with independent random variables, and $\Sigma^{1/2}$ is the square root of the $M \times M$ diagonal and positive definite matrix Σ. The dimensionality M is chosen to be proportional to the sample size N. Assuming that the eigenvalue distribution of Σ has a deterministic limit μ_σ, it was proved by Marchenko and Pastur [55] that the spectral measure of H approaches a deterministic probability measure. In the null case $\Sigma = I$, the limiting measure is called the Marchenko–Pastur distribution, μ_{MP}. For the non-null case $\Sigma \neq I$, the limiting measure is given by the free multiplicative convolution of μ_{MP} and μ_σ, denoted by $\mu_{\mathrm{MP}} \boxtimes \mu_\sigma$, see [5,67,69]. A CLT for the fluctuations of the linear eigenvalue statistics was first studied by Jonsson [42] for Wishart matrices where X has Gaussian entries. CLTs for linear eigenvalue statistics with analytic test functions for general sample covariance matrices were then studied by Bai and Silverstein in [4]. The regularity condition on the test functions was weakened by [6,54,61] for the null case and [58] for the non-null case. In the second part of this paper, we extend the techniques to derive corresponding CLTs for the mesoscopic eigenvalue statistics of sample covariance matrices; see Theorem 8.7.

1.4. *Related models*

Deformed Wigner matrices are closely related to Dyson Brownian motion, for which mesoscopic CLTs were obtained inside the bulk [23,38,45] and at the regular edges [1]. The mesoscopic linear statistics were also studied for random band matrices [25,26], sparse Wigner matrices [35], mesoscopic eigenvalue density correlations for Wigner matrices [37], invariant β-ensembles [10] and orthogonal polynomial ensembles [20]. The global fluctuations of the deformed GOE/GUE can also be studied using the framework of second order freeness [57].

1.5. *Structure of this paper*

Section 2 contains the precise definitions, assumptions and the main results. The proof the main theorem is carried out in Section 3–5. In Sections 6 and 7, we compute the variances and the biases in the bulk and at the edges. In Section 8, we consider sample covariance matrices and obtain the corresponding results. Some auxiliary results are proved in the Appendices.

1.6. *Notation*

We denote the upper half-plane by $\mathbb{C}^+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ and the positive real line by $\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$. For any vector $v \in \mathbb{C}^N$, we use $\|v\|_2$ to denote the Euclidean norm. For a matrix $A \in \mathbb{C}^{N\times N}$, we denote by $\|A\|_{\mathrm{op}}$ its operator norm

induced by the Euclidean vector norm. We use c, k and C, K to denote strictly positive constants that are independent of N. Their values may change from line to line. We use standard big O and small o notations. For $X, Y \in \mathbb{R}$, we write $X \sim Y$ if there exist constants $c, C > 0$ such that $c|Y| \le |X| \le C|Y|$. We write $X \ll Y$ if there exists a small $\tau > 0$ such that $|X| \le N^{-\tau}|Y|$ for large N. We will use the following definition on high-probability estimates from [27].

Definition 1.1. Let $\mathcal{X} \equiv \mathcal{X}^{(N)}$ and $\mathcal{Y} \equiv \mathcal{Y}^{(N)}$ be two sequences of nonnegative random variables. We say $\mathcal{Y}$ stochastically dominates $\mathcal{X}$ if, for all (small) $\epsilon > 0$ and (large) $D > 0$,

$$\mathbb{P}\big(\mathcal{X}^{(N)} > N^{\epsilon}\mathcal{Y}^{(N)}\big) \le N^{-D}, \tag{1.4}$$

for sufficiently large $N \ge N_0(\epsilon, D)$, and we write $\mathcal{X} \prec \mathcal{Y}$ or $\mathcal{X} = O_{\prec}(\mathcal{Y})$.

We often use the notation $\prec$ also for deterministic quantities, then (1.4) holds with probability one. Stochastic domination has the following properties.

Lemma 1.2 (Proposition 6.5 in [33]).

(1) *$X \prec Y$ and $Y \prec Z$ imply $X \prec Z$;*
(2) *If $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then $X_1 + X_2 \prec Y_1 + Y_2$ and $X_1X_2 \prec Y_1Y_2$;*
(3) *If $X \prec Y$, $\mathbb{E}Y \ge N^{-c_1}$ and $|X| \le N^{c_2}$ almost surely with fixed constants c_1 and c_2, then we have $\mathbb{E}X \prec \mathbb{E}Y$.*

2. Model and main results

2.1. *Model and assumptions*

Let $H \equiv H_N$ be an $N \times N$ real or complex Wigner matrix satisfying the following assumption.

Assumption 2.1. For a real ($\beta = 1$) symmetric Wigner matrix H we assume that:

(1) $\{H_{ij} | i \le j\}$ are independent real-valued centered random variables with $H_{ij} = H_{ji}$.
(2) For $i \ne j$, $\mathbb{E}[(\sqrt{N}H_{ij})^2] = 1$; $\mathbb{E}[(\sqrt{N}H_{ii})^2] = m_2$ for some constant $m_2 > 0$. In addition, $\mathbb{E}[(\sqrt{N}H_{ij})^4] = W_4$ for some constant $W_4 > 0$.
(3) All entries have uniform sub-exponential decay, that is, there exist $C_0 > 0$ and $\theta > 1$ such that

$$\mathbb{P}\big(|\sqrt{N}H_{ij}| \ge x\big) \le C_0 e^{-x^{\frac{1}{\theta}}}, \quad \forall i, j. \tag{2.1}$$

In particular, we have

$$\mathbb{E}\big[|\sqrt{N}H_{ij}|^p\big] \le C(\theta p)^{\theta p} \quad (p \ge 3). \tag{2.2}$$

For complex ($\beta = 2$) Hermitian Wigner matrix we assume that:

(1) $\{\operatorname{Re} H_{ij}, \operatorname{Im} H_{ij} | i \le j\}$ are independent centered real-valued random variables with $H_{ij} = \overline{H_{ji}}$.
(2) For $i \ne j$, $\mathbb{E}[H_{ij}^2] = 0$ and $\mathbb{E}[(\sqrt{N}|H_{ij}|)^2] = 1$; $\mathbb{E}[(\sqrt{N}|H_{ii}|)^2] = m_2$ for some constant $m_2 > 0$. In addition, $\mathbb{E}[(\sqrt{N}|H_{ij}|)^4] = W_4$ for some constant $W_4 > 0$.
(3) The sub-exponential tail assumption in (2.1) holds.

Let $\{A_N\} = \operatorname{Diag}(a_i)$ be a sequence of real deterministic diagonal $N \times N$ matrices with $\|A\|_{\mathrm{op}}$ uniformly bounded in N. The empirical spectral measure of A_N is defined by $\mu_A := \frac{1}{N}\sum_{i=1}^{N} \delta_{a_i}$.

For a probability measure ν on $\mathbb{R}$ denote by m_ν its Stieltjes transform, i.e.

$$m_\nu(z) := \int_{\mathbb{R}} \frac{\mathrm{d}\nu(x)}{x - z}, \quad z \in \mathbb{C}^+. \tag{2.3}$$

Note that $m_\nu : \mathbb{C}^+ \to \mathbb{C}^+$ is analytic and can be analytically continued to the real line outside the support of ν. Moreover, m_ν satisfies $\lim_{\eta \nearrow \infty} \mathrm{i}\eta m_\mu(\mathrm{i}\eta) = -1$. Conversely, if $m : \mathbb{C}^+ \to \mathbb{C}^+$ is an analytic function with $\lim_{\eta \nearrow \infty} \mathrm{i}\eta m(\mathrm{i}\eta) = -1$, then m is the Stieltjes transform of a probability measure ν, *i.e.*, $m(z) = m_\nu(z)$, for all $z \in \mathbb{C}^+$; see *e.g.*, [2].

The following assumption ensures the existence of the weak limiting measure of μ_A.

Assumption 2.2. There exists a deterministic and compactly supported probability measure denoted as μ_α, such that μ_A converges weakly to μ_α. In addition, there exists $\alpha_0 > 0$ such that for any fixed compact set $D_0 \subset \mathbb{C}^+ \cup \mathbb{R}$ with $D_0 \cap \operatorname{supp}(\mu_\alpha) = \varnothing$,

$$\max_{z\in D_0}\left|\int_{\mathbb{R}} \frac{\mathrm{d}\mu_A(x)}{x-z} - \int_{\mathbb{R}} \frac{\mathrm{d}\mu_\alpha(x)}{x-z}\right| = O\big(N^{-\alpha_0}\big), \tag{2.4}$$

for sufficiently large N.

Define the deformed Wigner matrix as

$$X_N := A_N + H_N.$$

The eigenvalues of X_N are denoted as $\lambda_i \in \mathbb{R}$. The empirical spectral measure of X_N is defined by $\mu_N(x) = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$. For $z \in \mathbb{C}^+$, we introduce the Green function, $G(z)$, and its normalized trace as

$$G(z) := (X_N - zI)^{-1}, \qquad m_N(z) := \frac{1}{N}\operatorname{Tr} G(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_N(\lambda)}{\lambda - z},$$

i.e., $m(z) \equiv m_N(z)$ is the Stieltjes transform of μ_N.

The empirical spectral distribution μ_N converges as N tends to infinity to the free additive convolution of μ_α and the standard semicircle law, denoted by $\tilde{\mu}_{\mathrm{fc}} := \mu_\alpha \boxplus \mu_{\mathrm{sc}}$. The free convolution measure can be described by analytic subordination [13,68]: Its Stieltjes transform, $\tilde{m}_{\mathrm{fc}}$, is the unique solution to the Pastur equation

$$\tilde{m}_{\mathrm{fc}}(z) = \int_{\mathbb{R}} \frac{1}{a - z - \tilde{m}_{\mathrm{fc}}(z)}\,\mathrm{d}\mu_\alpha(a), \tag{2.5}$$

subject to the constraint $\operatorname{Im}\tilde{m}_{\mathrm{fc}}(z) > 0$, $z \in \mathbb{C}^+$.

Since the convergence speed in (2.4) can be very slow, we work with a finite N version of the free convolution measure. Let $\mu_{\mathrm{fc}} := \mu_A \boxplus \mu_{\mathrm{sc}}$. The Stieltjes transform of μ_{fc}, denoted by m_{fc}, is hence the unique solution to

$$m_{\mathrm{fc}}(z) = \frac{1}{N}\sum_{i=1}^N \frac{1}{a_i - z - m_{\mathrm{fc}}(z)}, \tag{2.6}$$

such that $\operatorname{Im} m_{\mathrm{fc}}(z) > 0$, $z \in \mathbb{C}^+$. Note that μ_{fc} depends on N, but is deterministic.

Biane [12] proved that $\widetilde{\mu}_{\mathrm{fc}}$ and μ_{fc} are absolutely continuous probability measures whose densities, are analytic wherever positive. We denote the density functions by $\widetilde{\rho}_{\mathrm{fc}}$ and ρ_{fc}. In general the measures ρ_{fc} and $\widetilde{\rho}_{\mathrm{fc}}$ are supported on several disjoint intervals and may have irregular edges where the densities do not vanish as a square root or have cusp points inside the support. The following assumption will rule out such scenarios.

Assumption 2.3. Let $\mathcal{I}$ be the smallest interval that contains the support of μ_α, and assume that

$$\inf_{x\in\mathcal{I}} \int_{\mathbb{R}} \frac{\mathrm{d}\mu_\alpha(a)}{(a-x)^2} \geq 1 + w,$$

for some constant $w > 0$ (the left side may be infinite). Similarly, let $\hat{\mathcal{I}}$ be the smallest interval that contains the support of μ_A, and assume that

$$\inf_{x\in\hat{\mathcal{I}}} \int_{\mathbb{R}} \frac{\mathrm{d}\mu_A(a)}{(a-x)^2} \geq 1 + w,$$

for sufficiently large N.

The above assumption ensures that the density functions ρ_{fc} and $\tilde{\rho}_{\mathrm{fc}}$ are supported on a single interval (for N sufficiently large) and vanish as square roots at the endpoints of the support.

Lemma 2.4 (Lemma 2.4, 3.2 and 3.5 in [52]). *Under Assumption 2.3, there exists $\tilde{L}_-$ and $\tilde{L}_+ \in \mathbb{R}$, such that* $\operatorname{supp}\tilde{\rho}_{\mathrm{fc}} = [\tilde{L}_-, \tilde{L}_+]$, *and $\tilde{\rho}_{\mathrm{fc}}$ is strictly positive in $(\tilde{L}_-, \tilde{L}_+)$. Moreover, there exists $C > 1$ such that*

$$C^{-1}\sqrt{\kappa} \leq \tilde{\rho}_{\mathrm{fc}}(E) \leq C\sqrt{\kappa}, \quad E \in [\tilde{L}_-, \tilde{L}_+],$$

where $\tilde{\kappa} := \min\{|E - \tilde{L}_-|, |E - \tilde{L}_+|\}$. *The endpoints* $\tilde{L}_\pm$ *are the two real solutions to the equation*

$$\int_{\mathbb{R}} \frac{\mathrm{d}\mu_\alpha(a)}{(a - \widetilde{L}_\pm - m_{\mathrm{fc}}(\widetilde{L}_\pm))^2} = 1. \tag{2.7}$$

The same holds true, for sufficiently large N, *if we replace* μ_α, $\tilde{\rho}_{\mathrm{fc}}$, $\tilde{L}_\pm$ *and* $\tilde{\kappa}$ *by* μ_A, ρ_{fc}, $L_\pm$ *and* κ, *respectively. Here* $[L_-, L_+]$ *is the support of* ρ_{fc} *and* $\kappa := \min\{|E - L_-|, |E - L_+|\}$.

2.2. *Local law for the deformed Wigner matrices*

We introduce the spectral domain,

$$D' := \big\{z = E + \mathrm{i}\eta : |E| \le M, N^{-1+c} \le \eta \le 3\big\}, \tag{2.8}$$

where $M > 1 + \max\{|\tilde{L}_-|, |\tilde{L}_+|\}$ and $c > 0$ is small. Define the deterministic control parameters

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_{\mathrm{fc}}(z)}{N|\eta|}} + \frac{1}{N|\eta|}, \qquad \Theta(z) := \frac{1}{N|\eta|}, \qquad z = E + \mathrm{i}\eta \in \mathbb{C} \setminus \mathbb{R}. \tag{2.9}$$

Using (4.1), (4.2) in Lemma 4.1 below, we have

$$CN^{-\frac{1}{2}} \le \Psi(z) \ll 1, \quad z \in D'.$$

The following local law for the Green function was proved in [48].

Theorem 2.5 (Local law for the deformed Wigner matrix, Theorem 2.10 in [48]). *Under Assumptions* 2.1–2.3, *the following holds*

$$\max_{ij} \left| G_{ij}(z) - \delta_{ij} \frac{1}{a_i - z - m_{\mathrm{fc}}(z)} \right| \prec \Psi(z), \qquad \left| N^{-1} \operatorname{Tr} G(z) - m_{\mathrm{fc}}(z) \right| \prec \Theta(z), \tag{2.10}$$

uniformly for $z \in D'$.

The local law gives strong rigidity estimates for the eigenvalues of X_N. It also gives an upper bound, up to factors of N^ϵ, on the size of the fluctuations $\operatorname{Tr} G(z) - \mathbb{E} \operatorname{Tr} G(z)$. It is hence natural to study the fluctuations of $\operatorname{Tr} G(z) - \mathbb{E} \operatorname{Tr} G(z)$. The CLT for the linear eigenvalue statistics for general test functions is proved in [22] and [39] on global scale when $\operatorname{Im} z$ is order one. Via the Helffer–Sjöstrand functional calculus, a CLT for the resolvent can be translated to a CLT for the linear statistics.

Theorem 2.6 (Theorem 2.15 of [39]). *Under Assumptions* 2.1–2.3, *for any* $\varphi \in C_c(\mathbb{R})$ *which is analytic on a neighborhood of* $[\tilde{L}_-, \tilde{L}_+]$, *the random variable* $\sum_{i=1}^N \varphi(\lambda_i) - N \int_{\mathbb{R}} \varphi(x) \rho_{\mathrm{fc}}(x)\, \mathrm{d}x$ *converges in distribution to the Gaussian random variable with mean* $M(\varphi) = -\frac{1}{2\pi \mathrm{i}} \int_\Gamma \varphi(z) \tilde{b}(z)\, \mathrm{d}z$, *and variance* $V(\varphi) = \frac{1}{(2\pi \mathrm{i})^2} \int_\Gamma \int_\Gamma \varphi(z_1)\varphi(z_2) \tilde{K}(z_1, z_2)\, \mathrm{d}z_1\, \mathrm{d}z_2$, *where*

$$\tilde{b}(z) := \frac{\tilde{m}''_{\mathrm{fc}}(z)}{2(1 + \tilde{m}'_{\mathrm{fc}}(z))^2} \left((m_2 - 1) + \tilde{m}'_{\mathrm{fc}}(z) + (W_4 - 3) \frac{\tilde{m}'_{\mathrm{fc}}(z)}{1 + \tilde{m}'_{\mathrm{fc}}(z)} \right),$$

and $\tilde{K}(z_1, z_2) := (m_2 - 2) \frac{\partial^2 \tilde{I}}{\partial z_1 \partial z_2} + (W_4 - 3)(\tilde{I} \frac{\partial^2 \tilde{I}}{\partial z_1 \partial z_2} + \frac{\partial \tilde{I}}{\partial z_1} \frac{\partial \tilde{I}}{\partial z_2}) + \frac{2}{(1 - \tilde{I})^2} (\frac{\partial \tilde{I}}{\partial z_1} \frac{\partial \tilde{I}}{\partial z_2} + (1 - \tilde{I}) \frac{\partial^2 \tilde{I}}{\partial z_1 \partial z_2})$, *with*

$$\tilde{I}(z_1, z_2) := \int_{\mathbb{R}} \frac{1}{(x - z_1 - \tilde{m}_{\mathrm{fc}}(z_1))(x - z_2 - \tilde{m}_{\mathrm{fc}}(z_2))} \mathrm{d}\mu_\alpha(x).$$

Here Γ *is a rectangular contour with vertices* $(a_\pm \pm \mathrm{i}v_0)$ *so that* $\pm(a_\pm - \tilde{L}_\pm) > 0$ *and* Γ *lies within the analytic domain of* φ.

Using ideas of M. Shcherbina [61], the above result can be extended to $C_c^2(\mathbb{R})$ test functions. In [22], the corresponding result was obtained for the multi-cut regime.

2.3. *Main results*

Choose $E_0 \in [-1+\tilde{L}_-, 1+\tilde{L}_+]$ and $N^{-1} \ll \eta_0 \ll 1$. Consider a test function $g \in C_c^2(\mathbb{R})$ and set

$$f_N(x) := g\left(\frac{x-E_0}{\eta_0}\right). \tag{2.11}$$

We will write f_N as f for notational simplicity. Define

$$\kappa_0 := \operatorname{dist}\big(\operatorname{supp}(f), \{L_+, L_-\}\big). \tag{2.12}$$

Following [44,54], we study the characteristic function

$$\phi(\lambda) := \mathbb{E}\big[e(\lambda)\big], \quad \text{where} \quad e(\lambda) := \exp\Big\{\mathrm{i}\lambda\big(\operatorname{Tr} f(X_N) - \mathbb{E}\operatorname{Tr} f(X_N)\big)\Big\}, \quad \lambda \in \mathbb{R}. \tag{2.13}$$

Let $\tau > 0$ be an arbitrary small constant and define

$$\Omega_0 := \big\{x + \mathrm{i}y \in \mathbb{C} : |y| \geq N^{-\tau}\eta_0\big\}. \tag{2.14}$$

A key observation in [44] is that working on Ω_0 instead of all $\mathbb{C}$, effectively removes the ultra-local scales without affecting the mesoscopic linear statistics.

Proposition 2.7. *Let X_N be a deformed Wigner matrix satisfying Assumptions 2.1–2.3. Let $\eta_0\sqrt{\kappa_0+\eta_0} \geq N^{-1+c_0}$ for some $c_0 > 0$. Then there exists a small $0 < \tau < \frac{c_0}{16}$, such that the characteristic function (2.13) satisfies*

$$\phi'(\lambda) = -\lambda\phi(\lambda)V(f) + \tilde{\mathcal{E}}, \qquad V(f) := \frac{1}{\pi^2}\int_{\Omega_0}\int_{\Omega_0} \frac{\partial}{\partial\overline{z_1}}\tilde{f}(z_1)\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)K(z_1,z_2)\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2, \tag{2.15}$$

where $\tilde{f}$ is an almost analytic extension of f given in (3.2) below and $\beta = 1, 2$ is the symmetry parameter. The kernel K is given by

$$K(z_1,z_2) := \frac{\partial^2}{\partial z_1 \partial z_2}\left(\left(m_2 - \frac{2}{\beta}\right)I + \frac{(W_4 - 1 - \frac{2}{\beta})}{2}I^2\right) + \frac{2}{\beta}\frac{\partial}{\partial z_1}\left(\frac{1}{1-I}\frac{\partial I}{\partial z_2}\right), \tag{2.16}$$

with

$$I(z_1,z_2) := \int_{\mathbb{R}} \frac{1}{(x-z_1-m_{\mathrm{fc}}(z_1))(x-z_2-m_{\mathrm{fc}}(z_2))}\,\mathrm{d}\mu_A(x), \tag{2.17}$$

and the error $\tilde{\mathcal{E}}$ is bounded by

$$|\tilde{\mathcal{E}}| = O_\prec\big(|\lambda|\log N N^{-\tau}\big) + O_\prec\left(\frac{(1+|\lambda|^4)N^{3\tau}}{N\eta_0\sqrt{\kappa_0+\eta_0}}\right) + O_\prec\left(\frac{(1+|\lambda|^4)N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}\right),$$

provided that $V(f) = O(1)$.

Proposition 2.7 implies the following result.

Theorem 2.8. *Under the same assumptions as in Proposition 2.7, if we further assume that there exist $c, C > 0$ such that $c \leq V(f) \leq C$ for sufficiently large N, then $\frac{\operatorname{Tr} f(X_N) - \mathbb{E}\operatorname{Tr} f(X_N)}{\sqrt{V(f)}}$ converges in distribution to a standard Gaussian random variable.*

We remark that Theorem 2.8 applies to the global scale as well as the mesoscopic scales. The expectation of $\operatorname{Tr} f(X_N)$ has the following asymptotic expansion, which matches the result in [22,39] on the global scale.

Proposition 2.9. *Under the same assumptions as in Proposition 2.7, the so-called bias is given by*

$$\mathbb{E}\operatorname{Tr} f(X_N) - N\int_{\mathbb{R}} f(x)\rho_{\mathrm{fc}}(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)b(z)\,\mathrm{d}^2z + O\big(N^{-\tau}\big) + O_\prec\left(\frac{N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}\right), \tag{2.18}$$

where $\tilde{f}$ is given in (3.2) below, and

$$b(z) := \left(\frac{2}{\beta} - 1\right)\frac{1}{1 - I_s(z)}\frac{\mathrm{d}I_s(z)}{\mathrm{d}z} + \left(m_2 - \frac{2}{\beta}\right)\frac{\mathrm{d}I_s(z)}{\mathrm{d}z} + \left(W_4 - 1 - \frac{2}{\beta}\right)I_s(z)\frac{\mathrm{d}I_s(z)}{\mathrm{d}z}, \tag{2.19}$$

with

$$I_s(z) := \int_{\mathbb{R}} \frac{1}{(x - z - m_{\mathrm{fc}}(z))^2}\,\mathrm{d}\mu_A(x). \tag{2.20}$$

Note that the variance $V(f)$ in (2.15) and the bias in (2.18) are N-dependent and their formulas depend explicitly on the free convolution measures. We will compute their limits on the mesoscopic scales as N goes to infinity in order to obtain the following universal CLTs in the bulk and at the regular edges respectively.

Theorem 2.10 (Mesoscopic CLT in the bulk). *Let X_N be a deformed Wigner matrix satisfying Assumptions 2.1–2.3. Let $N^{-1+c} \le \eta_0 \le N^{-c}$ with some small $c > 0$, fix $E_0 \in (\tilde{L}_-, \tilde{L}_+)$ such that $\kappa_0 > c_0$, for some $c_0 > 0$ and large N. Then, for any test function $g \in C_c^2(\mathbb{R})$, the linear statistics*

$$\sum_{i=1}^{N} g\left(\frac{\lambda_i - E_0}{\eta_0}\right) - N\int_{\mathbb{R}} g\left(\frac{x - E_0}{\eta_0}\right)\rho_{\mathrm{fc}}(x)\,\mathrm{d}x \tag{2.21}$$

converges in distribution to a Gaussian random variable with mean zero and variance

$$\frac{1}{2\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2)^2}\,\mathrm{d}x_1\,\mathrm{d}x_2 = \frac{1}{\beta\pi}\int_{\mathbb{R}} |\xi|\left|\hat{g}(\xi)\right|^2\mathrm{d}\xi, \tag{2.22}$$

where $\hat{g}(\xi) := (2\pi)^{-1/2}\int_{\mathbb{R}} g(x)e^{-\mathrm{i}\xi x}\,\mathrm{d}x$. In particular, the bias vanishes in the bulk regime.

Theorem 2.11 (Mesoscopic CLT at the edge). *Let X_N be a deformed Wigner matrix satisfying Assumptions 2.1–2.3. Let $N^{-\frac{2}{3}+c} \le \eta_0 \le N^{-c}$ with some small $c > 0$. For any function $g \in C_c^2(\mathbb{R})$, the linear statistics (2.21) with $E_0 = L_+$ converges in distribution to a Gaussian random variable with mean $(\frac{2}{\beta} - 1)\frac{g(0)}{4}$ and variance*

$$\frac{1}{4\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{g(-x^2) - g(-y^2)}{x - y}\right)^2\mathrm{d}x\,\mathrm{d}y = \frac{1}{2\beta\pi}\int_{\mathbb{R}} |\xi|\left|\hat{h}(\xi)\right|^2\mathrm{d}\xi, \tag{2.23}$$

where $h(x) = g(-x^2)$ and $\hat{h}(\xi) := (2\pi)^{-1/2}\int_{\mathbb{R}} h(x)e^{-\mathrm{i}\xi x}\,\mathrm{d}x$. At the left edge $E_0 = L_-$, we obtain a similar CLT with $h(x) = g(x^2)$.

Remark. The bulk variance (2.22) agrees with the GOE/GUE. For the edges, the bias and variance in (2.23) coincide with those of the GUE/GOE obtained in [9,56] and the Dyson Brownian motion in [1].

Remark. We remark that our assumption that the fourth moments of the off-diagonal entries are identical can easily be relaxed in the above theorems. The regularity condition we impose on the test function g is clearly not optimal, and we expect results can be extended to $C^{1,r,s}(\mathbb{R})$ functions; see [36]. The CLTs also hold true if we consider the resolvent test function $g(x) = \frac{1}{x-\mathrm{i}}$.

Finally, for test functions in $C_c^2(\mathbb{R})$, we can relax the single support condition for μ_{fc} by assuming instead that the cuts of the support of μ_{fc} are separated by order one and the density ρ_{fc} has square-root decay near the edges.

3. Proof of Proposition 2.7

In this section, we prove Proposition 2.7 by reducing it to the main technical result Lemma 3.4. Recall the scaled test function f on scale η_0 from (2.11). There are constants such that

$$\|f\|_1 \le C\eta_0; \qquad \left\|f'\right\|_1 \le C'; \qquad \left\|f''\right\|_1 \le \frac{C''}{\eta_0}. \tag{3.1}$$

We use the Helffer–Sjöstrand formula to link $f(X_N)$ to the Green function of X_N.

Lemma 3.1 (Helffer–Sjöstrand formula). *Let $f \in C_c^2(\mathbb{R})$ and $\chi(y)$ be a smooth cutoff function with support in $[-2,2]$, with $\chi(y)=1$ for $|y|\le 1$. Define its almost-analytic extension*

$$\tilde f(x+\mathrm{i}y) := \big(f(x)+\mathrm{i}yf'(x)\big)\chi(y). \tag{3.2}$$

Then we have

$$f(\lambda) = \frac{1}{\pi}\int_{\mathbb{C}} \frac{\frac{\partial}{\partial \overline{z}}\tilde f(z)}{\lambda - z}\,\mathrm{d}^2 z = \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{\mathrm{i}yf''(x)\chi(y)+\mathrm{i}(f(x)+\mathrm{i}yf'(x))\chi'(y)}{\lambda - x - \mathrm{i}y}\,\mathrm{d}x\,\mathrm{d}y, \tag{3.3}$$

where $z=x+\mathrm{i}y$, $\frac{\partial}{\partial\overline{z}} = \frac{1}{2}(\frac{\partial}{\partial x}+\mathrm{i}\frac{\partial}{\partial y})$, and $\mathrm{d}^2 z$ is the Lebesgue measure on $\mathbb{C}$.

Therefore, we write

$$\operatorname{Tr} f(X_N) - \mathbb{E}\operatorname{Tr} f(X_N) = \frac{1}{\pi}\int_{\mathbb{C}} \frac{\partial}{\partial\overline{z}}\tilde f(z)\big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E}\operatorname{Tr} G(z)\big)\,\mathrm{d}^2 z. \tag{3.4}$$

Plugging the above equation in $e(\lambda)$ given by (2.13), we have

$$e(\lambda) = \exp\left\{\frac{\mathrm{i}\lambda}{\pi}\int_{\mathbb{C}} \frac{\partial}{\partial\overline{z}}\tilde f(z)\big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E}\operatorname{Tr} G(z)\big)\,\mathrm{d}^2 z\right\}. \tag{3.5}$$

Taking the derivative of the characteristic function given in (2.13), and applying (3.5), we get

$$\phi'(\lambda) = \frac{\mathrm{i}}{\pi}\int_{\mathbb{C}} \frac{\partial}{\partial\overline{z}}\tilde f(z)\mathbb{E}\big[e(\lambda)\big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E}\operatorname{Tr} G(z)\big)\big]\,\mathrm{d}^2 z. \tag{3.6}$$

Following [44], we restrict the domain of the spectral parameter to Ω_0, as the very local scales do not contribute to $\phi(\lambda)$. We write

$$\operatorname{Tr} f(X_N) - \mathbb{E}\operatorname{Tr} f(X_N) = \frac{1}{\pi}\left(\int_{\Omega_0} + \int_{\Omega_0^c}\right)\frac{\partial\tilde f(z)}{\partial\overline{z}}\big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E}\operatorname{Tr} G(z)\big)\,\mathrm{d}^2 z. \tag{3.7}$$

Recall $\tilde f$ in (3.2) and the definition of Ω_0 in (2.14). Since $\chi(y)=1$ for $|y|\le 1$, we can write the second integral in (3.7) with $z=x+\mathrm{i}y$ as

$$\frac{N}{\pi}\int_{\mathbb{R}}\int_0^{\frac{\eta_0}{N^\tau}} \mathrm{i}yf''(x)\big(m_N(z)-\mathbb{E}m_N(z)\big)\,\mathrm{d}x\,\mathrm{d}y = -\frac{N}{\pi}\int_{\mathbb{R}}\int_0^{\frac{\eta_0}{N^\tau}} yf''(x)\operatorname{Im}\big(m_N(z)-\mathbb{E}m_N(z)\big)\,\mathrm{d}x\,\mathrm{d}y, \tag{3.8}$$

where we used the fact that $m_N(\overline{z}) = \overline{m_N(z)}$. We now choose a small $\tau>0$ such that $N^{-1} \ll y_0 := \sqrt{\frac{\eta_0}{N^{1+\tau}}} \le N^{-\tau}\eta_0$. In the regime $y\in[y_0, N^{-\tau}\eta_0]$, the integral can be estimated using the local law (2.10), (3.1) and Lemma 1.2, i.e.,

$$\left|\frac{N}{\pi}\int_{\mathbb{R}}\int_{y_0}^{N^{-\tau}\eta_0} yf''(x)\operatorname{Im}\big(m_N(z)-\mathbb{E}m_N(z)\big)\,\mathrm{d}x\,\mathrm{d}y\right| \prec \left|\int_{\mathbb{R}} f''(x)\,\mathrm{d}x\int_{y_0}^{N^{-\tau}\eta_0}\mathrm{d}y\right| = O_\prec\big(N^{-\tau}\big). \tag{3.9}$$

In the regime $y\in[0,y_0]$, the local law is not sharp but instead we use the fact that $y\to \operatorname{Im} m_N(x+\mathrm{i}y)y$ is increasing. That is,

$$\left|\frac{N}{\pi}\int_{\mathbb{R}}\int_0^{y_0} yf''(x)\operatorname{Im}\big(m_N(z)-\mathbb{E}m_N(z)\big)\,\mathrm{d}x\,\mathrm{d}y\right| = O_\prec\left(\frac{Ny_0^2}{\eta_0}\right) = O_\prec\big(N^{-\tau}\big). \tag{3.10}$$

Therefore, we have from (3.7) that

$$\operatorname{Tr} f(X_N) - \mathbb{E}\operatorname{Tr} f(X_N) = \frac{1}{\pi}\int_{\Omega_0} \frac{\partial}{\partial\overline{z}}\tilde f(z)\big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E}\operatorname{Tr} G(z)\big)\,\mathrm{d}^2 z + O_\prec\big(N^{-\tau}\big). \tag{3.11}$$

Using the same argument, since $|e(\lambda)|=1$, we have

$$\phi'(\lambda) = \frac{\mathrm{i}}{\pi}\int_{\Omega_0} \frac{\partial}{\partial\overline{z}}\tilde f(z)\mathbb{E}\big[e(\lambda)\big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E}\operatorname{Tr} G(z)\big)\big]\,\mathrm{d}^2 z + O_\prec\big(N^{-\tau}\big). \tag{3.12}$$

Similarly, we restrict the integration domain of $e(\lambda)$ in (3.5) to Ω_0. Let

$$e_0(\lambda) := \exp\left\{ \frac{\mathrm{i}\lambda}{\pi} \int_{\Omega_0} \frac{\partial}{\partial \overline{z}} \tilde{f}(z) \big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E} \operatorname{Tr} G(z) \big) \, \mathrm{d}^2 z \right\}. \tag{3.13}$$

In addition, (3.11) implies that $|e(\lambda) - e_0(\lambda)| = O_{\prec}(|\lambda| N^{-\tau})$. We also have $|e_0(\lambda)| = 1$, using $|e(\lambda)| = 1$ and $\operatorname{Tr} G(\overline{z}) = \overline{\operatorname{Tr} G(z)}$. If we further replace $e(\lambda)$ by $e_0(\lambda)$ in (3.12), then we get

$$\phi'(\lambda) = \frac{\mathrm{i}}{\pi} \int_{\Omega_0} \frac{\partial}{\partial \overline{z}} \tilde{f}(z) \mathbb{E}\big[(e_0(\lambda) \big(\operatorname{Tr}\big(G(z)\big) - \mathbb{E} \operatorname{Tr} G(z) \big) \big] \, \mathrm{d}^2 z + O_{\prec}\big(|\lambda| \log N N^{-\tau} \big). \tag{3.14}$$

The last error term on the right side, and many error terms below, are estimated using the following lemma, which is a variant of Lemma 4.4 in [44]. The proof is provided in Appendix B.

Lemma 3.2. *Suppose $h(z)$ is a holomorphic function on Ω_0 and $|h(z)| \le \frac{K}{|\operatorname{Im} z|^s}$ for some constants $s, K \ge 0$, then there exists some constant C such that*

$$\left| \int_{\Omega_0} \frac{\partial}{\partial \overline{z}} \tilde{f}(z) h(z) \, \mathrm{d}^2 z \right| \le C K N^{\tau s} \eta_0^{1-s}.$$

For $1 \le s \le 2$, the bound is sharpened to $C K \log(N) \eta_0^{1-s}$.

Thus, in order to study $\phi'(\lambda)$, it is sufficient to estimate $\mathbb{E}[e_0(\lambda)(\operatorname{Tr}(G(z)) - \mathbb{E} \operatorname{Tr} G(z))]$. The key input is the following cumulant expansion formula.

Lemma 3.3 (Cumulant expansion formula). *Let h be a real-valued random variable with finite moments, and f is a complex-valued smooth function on $\mathbb{R}$ with bounded derivatives. Let c_k be the k-th cumulant of h, given by $c_k(h) := (-\mathrm{i})^k \frac{\mathrm{d}}{\mathrm{d}t} \log \mathbb{E} e^{\mathrm{i}th}|_{t=0}$. Then for any fixed $l \in \mathbb{N}$, we have*

$$\mathbb{E}\big[h f(h) \big] = \sum_{k=0}^{l} \frac{1}{k!} c_{k+1}(h) \mathbb{E}\big[f^{(k)}(h) \big] + R_{l+1},$$

where the error term satisfies

$$|R_{l+1}| \le C_l \mathbb{E}|h|^{l+2} \sup_{|x| \le M} \big| f^{(l+1)}(x) \big| + C_l \mathbb{E}\big[|h|^{l+2} 1_{|h|>M} \big] \big\| f^{(l+1)} \big\|_{\infty},$$

and $M > 0$ is an arbitrary fixed cutoff.

For reference, we refer e.g. to Lemma 3.1 in [36]. We give the proof the following lemma in Section 5.

Lemma 3.4. *For any $z := E + \mathrm{i}\eta \in \Omega_0 \cap D'$, see (2.8), and $\kappa := \min\{|E - L_-|, |E - L_+|\}$. we have*

$$\mathbb{E}\big[e_0(\lambda) \big(\operatorname{Tr} G(z) - \mathbb{E} \operatorname{Tr} G(z) \big) \big] = \frac{\mathrm{i}\lambda}{\pi} \mathbb{E}\big[e_0(\lambda) \big] \int_{\Omega_0} \frac{\partial}{\partial \overline{z'}} \tilde{f}\big(z'\big) K\big(z, z'\big) \, \mathrm{d}^2 z' + \mathcal{E}(z),$$

where K is given in (2.16) and $\mathcal{E}(z)$ is analytic in Ω_0 and satisfies

$$\mathcal{E}(z) = O_{\prec}\left(\frac{1 + |\lambda|^4}{\sqrt{\kappa + \eta}} \right) \left(\frac{(\kappa + \eta)^{1/4}}{\sqrt{N\eta^3}} + \frac{1}{\sqrt{N\eta^2}} + \frac{1}{\sqrt{N\eta_0 \eta}} + \frac{1}{N\eta_0 \eta} + \frac{1}{N\eta^2} \right). \tag{3.15}$$

Admitting Lemma 3.4 and plugging in (3.14), we have

$$\phi'(\lambda) = -\lambda \mathbb{E}\big[e_0(\lambda) \big] V(f) + O_{\prec}\big(|\lambda| \log N N^{-\tau} \big) + \tilde{\mathcal{E}},$$

where

$$V(f) = \frac{1}{\pi^2} \int_{\Omega_0} \int_{\Omega_0} \frac{\partial}{\partial \overline{z}} \tilde{f}(z) \frac{\partial}{\partial \overline{z'}} \tilde{f}\big(z'\big) K\big(z, z'\big) \, \mathrm{d}^2 z \, \mathrm{d}^2 z', \qquad \tilde{\mathcal{E}} = \frac{\mathrm{i}}{\pi} \int_{\Omega_0} \frac{\partial}{\partial \overline{z}} \tilde{f}(z) \mathcal{E}(z) \, \mathrm{d}^2 z.$$

By the definition of κ_0 in (2.12), $\kappa \geq \kappa_0$. Moreover $|\eta| \geq N^{-\tau}\eta_0$, for $z \in \Omega_0$. Using Lemma 3.2, we hence obtain the estimate

$$\tilde{\mathcal{E}} = O_{\prec}\Big(\frac{(1+|\lambda|^4)N^{3\tau}}{N\eta_0\sqrt{\kappa_0+\eta_0}}\Big) + O_{\prec}\Big(\frac{(1+|\lambda|^4)N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}\Big).$$

Assuming $V(f) \prec O(1)$, we can replace $e_0(\lambda)$ by $e(\lambda)$ with error $O_{\prec}(|\lambda|N^{-\tau})$. Thus we have completed the proof of Proposition 2.7.

4. Properties of the free convolution

4.1. *Properties of m_{fc} and $\tilde{m}_{\mathrm{fc}}$*

In this subsection, we recall some properties of the Stieltjes transforms m_{fc} and $\tilde{m}_{\mathrm{fc}}$ of the free convolution measures. Let $\kappa = \kappa(E)$ be the distance from E to the closest spectral edge, i.e.,

$$\kappa := \min\big\{|E-L_-|, |E-L_+|\big\}.$$

Similarly define $\tilde{\kappa} := \min\{|E-\tilde{L}_-|, |E-\tilde{L}_+|\}$. Define the spectral domain

$$D := \big\{z = E+\mathrm{i}\eta : |E| < M, 0 < \eta \leq 3\big\}.$$

Lemma 4.1 (Lemma 3.5, Lemma A.1 in [52]).

(1) *For all $z \in D$, there exists $C > 1$ such that*

$$C^{-1}\sqrt{\tilde{\kappa}+\eta} \leq \big|\mathrm{Im}\,\tilde{m}_{\mathrm{fc}}(z)\big| \leq C\sqrt{\tilde{\kappa}+\eta}, \tag{4.1}$$

if $E \in [\tilde{L}_-, \tilde{L}_+]$. If $E \in [\tilde{L}_-, \tilde{L}_+]^c$, then

$$C^{-1}\frac{\eta}{\sqrt{\tilde{\kappa}+\eta}} \leq \big|\mathrm{Im}\,\tilde{m}_{\mathrm{fc}}(z)\big| \leq C\frac{\eta}{\sqrt{\tilde{\kappa}+\eta}}. \tag{4.2}$$

(2) (*Stability bound*) *There exists $C > 1$, such that*

$$C^{-1} \leq \big|a - z - \tilde{m}_{\mathrm{fc}}(z)\big| \leq C, \tag{4.3}$$

uniformly for $z \in D$ and $a \in \mathrm{supp}(\mu_\alpha)$.

(3) *For all $z \in D$, there exist $k, K > 0$ such that*

$$k\sqrt{\tilde{\kappa}+\eta} \leq \Big|1 - \int_{\mathbb{R}} \frac{1}{(x-z-\tilde{m}_{\mathrm{fc}}(z))^2}\,\mathrm{d}\mu_\alpha(x)\Big| \leq K\sqrt{\tilde{\kappa}+\eta}. \tag{4.4}$$

(4) *There exist $C > 0$ and $c_0 > 0$ such that for all $z \in D$ satisfying $\tilde{\kappa}+\eta \leq c_0$,*

$$C^{-1} \leq \Big|\int_{\mathbb{R}} \frac{1}{(x-z-\tilde{m}_{\mathrm{fc}}(z))^3}\,\mathrm{d}\mu_\alpha(x)\Big| \leq C; \tag{4.5}$$

moreover, there exists $C > 1$ such that for all $z \in D$,

$$\Big|\int_{\mathbb{R}} \frac{1}{(x-z-\tilde{m}_{\mathrm{fc}}(z))^3}\,\mathrm{d}\mu_\alpha(x)\Big| \leq C.$$

The following lemma implies that m_{fc} behaves similarly as $\tilde{m}_{\mathrm{fc}}$, for sufficiently large N.

Lemma 4.2 (Lemma 3.6 in [52]). *Under Assumptions 2.2 and 2.3, for sufficiently large N, statements 1–4 in Lemma 4.1 hold true with $\tilde{m}_{\mathrm{fc}}$, $\tilde{\kappa}$, μ_α and $\tilde{L}_\pm$ replaced by m_{fc}, κ, μ_A and $L_\pm$ respectively. Moreover, the constants in these inequalities can be chosen uniformly in N for sufficiently large N. Furthermore, there exists $c > 0$ such that*

$$\max_{z\in D}\big|\tilde{m}_{\mathrm{fc}}(z) - m_{\mathrm{fc}}(z)\big| \leq N^{-\frac{c\alpha_0}{2}}, \qquad |\tilde{L}_\pm - L_\pm| \leq N^{-c\alpha_0}, \tag{4.6}$$

for sufficiently large N.

Recall the function $I(z_1, z_2)$ given in (2.17) and $I_s(z)$ in (2.20). By direct computation, one proves the following lemma.

Lemma 4.3. *For $z_1 \neq z_2$, we have*

$$I(z_1, z_2) = \frac{m_{\mathrm{fc}}(z_1) - m_{\mathrm{fc}}(z_2)}{z_1 + m_{\mathrm{fc}}(z_1) - z_2 - m_{\mathrm{fc}}(z_2)}; \qquad I_s(z) = \frac{m'_{\mathrm{fc}}(z)}{1 + m'_{\mathrm{fc}}(z)}. \tag{4.7}$$

As a result of Lemmas 4.1, 4.2 and 4.3, we have the following lemma.

Lemma 4.4. *There exists $C > 1$ such that*

$$\big|I(z_1, z_2)\big| \le C; \qquad \big|I_s(z)\big| \le 1; \qquad C^{-1}\sqrt{\kappa + \eta} \le \big|1 - I_s(z)\big| \le C\sqrt{\kappa + \eta};$$
$$\big|m_{\mathrm{fc}}(z)\big| \le C; \qquad \big|m'_{\mathrm{fc}}(z)\big| \le \frac{C}{\sqrt{\kappa + \eta}}; \qquad \big|m''_{\mathrm{fc}}(z)\big| \le \frac{C}{\sqrt{(\kappa + \eta)^3}},$$

uniformly for $z, z_1, z_2 \in D$.

The proof of the above two lemmas can be found in Appendix B.

4.2. *Properties of the Green function*

As a more general version of the local law in Theorem 2.5, we introduce the anisotropic local law. Recall the control parameters Ψ and Θ from (2.9).

Theorem 4.5 (Theorem 12.2, 12.4 in [43]; Theorem 2.1, 2.2 in [29]; Theorem 2.6 in [3]). *For any deterministic vector $v, w \in \mathbb{C}^N$ and matrix $B \in \mathbb{C}^{N\times N}$, we have*

$$\big|\langle v, G(z)w\rangle - \langle v, \widehat{G}(z)w\rangle\big| \prec \|v\|_2\|w\|_2\Psi(z), \qquad \big|N^{-1}\operatorname{Tr}\big(B\big(G(z) - \widehat{G}(z)\big)\big)\big| \prec \|B\|_{\mathrm{op}}\Theta(z),$$

uniformly in $z \in \{z = E + \mathrm{i}\eta : |E| \le \rho^{-1}, N^{-1+\rho} \le \eta \le \rho^{-1}\}$, where ρ is small so that $\rho^{-1} \ge \|A\|_{\mathrm{op}}$, and $\widehat{G} = \operatorname{Diag}(\frac{1}{a_i - z - m_{\mathrm{fc}}(z)})$.

5. Proof of Lemma 3.4

For the simplicity of the presentation, we consider only the real symmetric case here. The complex case being similar is proved in Appendix A. For notational simplicity, let

$$g_i(z) := \frac{1}{a_i - z - m_{\mathrm{fc}}(z)}, \quad z \in \mathbb{C} \setminus \operatorname{supp}(\mu_{\mathrm{fc}}). \tag{5.1}$$

Before we proceed the proof of Lemma 3.4, we state a useful lemma.

Lemma 5.1. *For any i, j, we have*

$$\frac{\partial e_0(\lambda)}{\partial H_{ij}} = -\frac{\mathrm{i}(2 - \delta_{ij})\lambda}{\pi} e_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial \overline{z}} \tilde{f}(z) \frac{\mathrm{d}}{\mathrm{d}z} G_{ji}\, \mathrm{d}^2 z; \tag{5.2}$$

$$\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}} = \frac{\mathrm{i}(2 - \delta_{ij})\lambda}{\pi} e_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial \overline{z}} \tilde{f}(z) \frac{\mathrm{d}}{\mathrm{d}z} \big(g_i(z) g_j(z)\big)\, \mathrm{d}^2 z + O_{\prec}\bigg(\frac{(1 + |\lambda|)^2}{\sqrt{N\eta_0}}\bigg). \tag{5.3}$$

In general, for any integer $k \in \mathbb{N}$, we have

$$\bigg|\frac{\partial^k G_{ij}}{\partial H_{ij}^k}\bigg| \prec O(1); \qquad \bigg|\frac{\partial^k e_0(\lambda)}{\partial^k H_{ij}}\bigg| \prec O\big(\big(1 + |\lambda|\big)^k\big). \tag{5.4}$$

The above lemma follows from the relation

$$\frac{\partial G_{ij}}{\partial H_{ab}} = -\frac{G_{ia}G_{bj} + G_{ib}G_{aj}}{1+\delta_{ab}}. \tag{5.5}$$

The details are provided in Appendix B. Now we are ready to prove Lemma 3.4.

Proof of Lemma 3.4. By the definition of the resolvent function, we have

$$(z-a_i)G_{ii} = (HG)_{ii} - 1.$$

Thus we obtain that

$$(z-a_i)\mathbb{E}\big[e_0(\lambda)(G_{ii}-\mathbb{E}G_{ii})\big] = \sum_{j=1}^{N}\big(\mathbb{E}\big[H_{ij}G_{ji}e_0(\lambda)\big] - \mathbb{E}[H_{ij}G_{ji}]\mathbb{E}\big[e_0(\lambda)\big]\big).$$

Using the cumulant expansion Theorem 3.3, we obtain

$$(z-a_i)\mathbb{E}\big[e_0(\lambda)(G_{ii}-\mathbb{E}G_{ii})\big] = I_1 + I_2 + I_3 + O_{\prec}\big(N^{-\frac{3}{2}}\big(1+|\lambda|^4\big)\big), \tag{5.6}$$

where

$$I_1 := \frac{1}{N}\sum_{j=1}^{N} c_{ij}^{(2)}\left(\mathbb{E}\left[\frac{\partial e_0(\lambda)}{\partial H_{ij}}G_{ji}\right] + \mathbb{E}\left[\left(\frac{\partial G_{ji}}{\partial H_{ij}} - \mathbb{E}\left[\frac{\partial G_{ji}}{\partial H_{ij}}\right]\right)e_0(\lambda)\right]\right);$$

$$I_2 := \frac{1}{2!N^{\frac{3}{2}}}\sum_{j=1}^{N} c_{ij}^{(3)}\left(\mathbb{E}\left[\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}}G_{ji}\right] + 2\mathbb{E}\left[\frac{\partial e_0(\lambda)}{\partial H_{ij}}\frac{\partial G_{ji}}{\partial H_{ij}}\right] + \mathbb{E}\left[(1-\mathbb{E})\left(\frac{\partial^2 G_{ji}}{\partial^2 H_{ij}}\right)e_0(\lambda)\right]\right);$$

$$\begin{aligned} I_3 := \frac{1}{3!N^2}\sum_{j=1}^{N} c_{ij}^{(4)}\Bigg(&\mathbb{E}\left[\frac{\partial^3 e_0(\lambda)}{\partial^3 H_{ij}}G_{ji}\right] + 3\mathbb{E}\left[\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}}\frac{\partial G_{ji}}{\partial H_{ij}}\right] + 3\mathbb{E}\left[\frac{\partial e_0(\lambda)}{\partial H_{ij}}\frac{\partial^2 G_{ji}}{\partial^2 H_{ij}}\right] \\ &+ \mathbb{E}\left[\left(\frac{\partial^3 G_{ji}}{\partial^3 H_{ij}} - \mathbb{E}\left[\frac{\partial^3 G_{ji}}{\partial^3 H_{ij}}\right]\right)e_0(\lambda)\right]\Bigg). \end{aligned}$$

Here $c_{ij}^{(k)}$ denotes the k-th cumulant of $\sqrt{N}H_{ij}$. In particular,

$$c_{ij}^{(1)} = 0; \qquad c_{ij}^{(2)} = 1 + (m_2 - 1)\delta_{ij}; \qquad c_{ij}^{(4)} = W_4 - 3 \quad (i \neq j).$$

The last term on the right side of (5.6) is estimated by (5.4), (2.2) and Lemma 1.2. Note that for $z \in \Omega_0 \cap D'$, we have the deterministic bound $|G_{ij}| \leq \|G\|_{\mathrm{op}} \leq (\operatorname{Im} z)^{-1} = O(N^c)$. Combining with $|e_0(\lambda)| = 1$, we can use the last statement of Lemma 1.2. We will use this argument throughout the proof. The error terms in this section are all uniform in $z \in \Omega_0 \cap D'$. In the following, we estimate I_1, I_2 and I_3 respectively.

5.1. *Estimate on* I_1

Using (5.5), we have for each i,

$$\begin{aligned} I_1 =& -\frac{1}{N}\mathbb{E}\big[e_0(\lambda)\big(\big(G^2\big)_{ii} - \mathbb{E}\big(G^2\big)_{ii}\big)\big] - \frac{1}{N}\mathbb{E}\big[e_0(\lambda)(\operatorname{Tr}GG_{ii} - \mathbb{E}\operatorname{Tr}GG_{ii})\big] \\ &- \frac{m_2-2}{N}\mathbb{E}\big[e_0(\lambda)(G_{ii}G_{ii} - \mathbb{E}G_{ii}G_{ii})\big] + \frac{1}{N}\sum_{j=1}^{N}\big(1+(m_2-1)\delta_{ij}\big)\mathbb{E}\left[\frac{\partial e_0(\lambda)}{\partial H_{ij}}G_{ji}\right] \\ =:& A_1(i) + A_2(i) + A_3(i) + A_4(i). \end{aligned}$$

The first term can be written as

$$A_1(i) = -\frac{1}{N}\mathbb{E}\left[e_0(\lambda)(1-\mathbb{E})\frac{\mathrm{d}}{\mathrm{d}z}\big(G(z)\big)_{ii}\right] \prec \frac{\Psi(z)}{N\operatorname{Im} z},$$

with $\Psi(z)$ as in (2.9). The last step follows from the local law and the Cauchy integral formula. Similarly using the local law, the second term $A_2(i)$ can be written as

$$A_2(i) = -\frac{1}{N}\mathbb{E}\big[e_0(\lambda)\big(\operatorname{Tr} G(G_{ii} - \mathbb{E}G_{ii}) + \mathbb{E}G_{ii}(\operatorname{Tr} G - \mathbb{E}\operatorname{Tr} G) + \mathbb{E}\operatorname{Tr} G\mathbb{E}G_{ii} - \mathbb{E}\operatorname{Tr} GG_{ii}\big)\big]$$
$$= -m_{\mathrm{fc}}(z)\mathbb{E}\big[e_0(\lambda)(G_{ii} - \mathbb{E}G_{ii})\big] - \frac{1}{N}g_i(z)\mathbb{E}\big[e_0(\lambda)(\operatorname{Tr} G - \mathbb{E}\operatorname{Tr} G)\big] + O_{\prec}\big(\Theta(z)\Psi(z)\big),$$

with $\Theta(z)$ as in (2.9) and $g_i(z)$ in (5.1). Here the first term of A_2 will be moved to the left side of the equation (5.6). In addition, the local law also implies that $A_3(i) = O_{\prec}(\frac{\Psi(z)}{N})$.

Note that A_4 is a leading term of I_1. Using the local law, (5.2) and Lemma 5.4, we write

$$A_4(i) = A_{41}(i) + A_{42}(i) + O_{\prec}\big(\big(1 + |\lambda|\big)N^{-1}\Psi(z)\big),$$

where

$$A_{41}(i) = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial H_{ij}}(1+\delta_{ij})G_{ji}\bigg], \quad \text{and} \quad A_{42}(i) = \frac{m_2 - 2}{N}\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial H_{ii}}g_i(z)\bigg]. \tag{5.7}$$

We compute these two terms below in Section 5.4.

5.2. *Estimate on I_2*

In this subsection, we will show that I_2 is negligible, which can be written as

$$\frac{1}{2N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\bigg(\mathbb{E}\bigg[\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}}G_{ji}\bigg] + 2\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial H_{ij}}\frac{\partial G_{ji}}{\partial H_{ij}}\bigg] + \mathbb{E}\bigg[e_0(\lambda)\bigg(\frac{\partial^2 G_{ji}}{\partial^2 H_{ij}} - \mathbb{E}\frac{\partial^2 G_{ji}}{\partial^2 H_{ij}}\bigg)\bigg]\bigg)$$
$$=: B_1(i) + B_2(i) + B_3(i).$$

First, we study the last term $B_3(i)$. Using (5.5) and the local law, we have

$$B_3(i) = -\frac{1}{2N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\mathbb{E}\big[e_0(\lambda)\big(6G_{ii}G_{jj}G_{ij} + 2(G_{ij})^3 - 6\mathbb{E}[G_{ii}G_{jj}G_{ij}] - 2\mathbb{E}(G_{ij})^3\big)\big]$$
$$= -\frac{3}{N^{\frac{3}{2}}}\sum_{j=1}^{N}\mathbb{E}\big[e_0(\lambda)c_{ij}^{(3)}g_i(z)g_j(z)(G_{ij} - \mathbb{E}G_{ij})\big] + O_{\prec}\big(N^{-\frac{1}{2}}\Psi^2(z)\big).$$

Next, we estimate $\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{ij}^{(3)}g_j(z)G_{ij}$, using the anisotropic local law Theorem 4.5. Let $v_j = \delta_{ij}$ and $w_j = \frac{1}{\sqrt{N}}c_{ij}^{(3)}g_j(z)$. And $\|w\|_2$ is bounded because of the stability bound (4.3) and the moment condition (2.2).

Note that Theorem 4.5 holds for vector entries w_j and v_j that are deterministic constants. As in our setting w_j depend on z, we use a continuity argument to show that

$$\bigg|\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{ij}^{(3)}g_j(z)G_{ij}(z) - \frac{1}{\sqrt{N}}c_{ii}^{(3)}\big(g_i(z)\big)^2\bigg| \prec \Psi(z), \tag{5.8}$$

uniformly in $z \in D'$. Indeed, choose a lattice Δ of the domain D' in (2.8), with $|\Delta| = N^{100}$. Then for any $z \in D'$, there exists some point $p \in \Delta$, such that $|z - p| \le N^{-10}$. The anisotropic local law (4.5) combined with a union bound implies

$$\mathbb{P}\bigg(\exists p \in \Delta : \bigg|\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{ij}^{(3)}g_j(p)G_{ij}(p) - \frac{1}{\sqrt{N}}c_{ii}^{(3)}\big(g_i(p)\big)^2\bigg| \ge N^{\epsilon}\Psi(p)\bigg) \le N^{-D+100}. \tag{5.9}$$

Recall that $g_j(z) = \frac{1}{a_j - z - m_{\mathrm{fc}}(z)}$. Using (4.3), Lemma 4.4 and the fact that $|G_{ij}(z)| \le \frac{1}{\eta}$, the function $\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{ij}^{(3)}g_j(z) \times G_{ij}(z) - \frac{1}{\sqrt{N}}c_{ii}^{(3)}(g_i(z))^2$ as well as $\Psi(z)$ are Lipschitz continuous on D' with Lipschitz constant at most N^3. Thus we

obtain from (5.9) that

$$\mathbb{P}\left(\exists z\in D':\left|\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{ij}^{(3)}g_j(z)G_{ij}(z)-\frac{1}{\sqrt{N}}c_{ii}^{(3)}\big(g_i(z)\big)^2\right|\geq 2N^{\epsilon}\Psi(z)\right)\leq N^{-D+100}, \tag{5.10}$$

which implies (5.8).

Next, using (4.3) and $\Psi(z)\geq CN^{-\frac{1}{2}}$, we have $|\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{ij}^{(3)}g_j(z)G_{ij}(z)|\prec\Psi(z)$. Therefore, we obtain the upper bound

$$B_3(i)=O_{\prec}\big(N^{-1}\Psi(z)\big)+O_{\prec}\big(N^{-\frac{1}{2}}\Psi^2(z)\big)=O_{\prec}\big(N^{-\frac{1}{2}}\Psi^2(z)\big).$$

For the second term, by (5.5), (5.2), and the local law we have

$$\begin{aligned}B_2(i)&=-\frac{1}{N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\mathbb{E}\left[\frac{\partial e_0(\lambda)}{\partial H_{ij}}(G_{ji}G_{ji}+G_{ii}G_{jj})\right]\\&=-\frac{1}{N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\mathbb{E}\left[\frac{\partial e_0(\lambda)}{\partial H_{ij}}g_i(z)g_j(z)\right]+O_{\prec}\left(\frac{(1+|\lambda|)\Psi(z)}{N\sqrt{\eta_0}}\right)\\&=\frac{2\mathrm{i}\lambda}{\pi N^{\frac{3}{2}}}\mathbb{E}\left[e_0(\lambda)\sum_{j=1}^{N}c_{ij}^{(3)}\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}(z')\frac{\mathrm{d}}{\mathrm{d}z'}\big(G(z')\big)_{ji}\,\mathrm{d}^2z'g_i(z)g_j(z)\right]+O_{\prec}\left(\frac{(1+|\lambda|)\Psi(z)}{N\sqrt{\eta_0}}\right).\end{aligned}$$

By the same argument as in (5.8) and the Cauchy integral formula, we have

$$\left|\frac{\mathrm{d}}{\mathrm{d}z'}\frac{1}{\sqrt{N}}\left(\sum_{j=1}^{N}c_{ij}^{(3)}g_j(z)\big(G(z')\big)_{ji}\right)\right|=O_{\prec}\left(\frac{\Psi(z')}{|\operatorname{Im}z'|}\right).$$

Using the stability bound (4.3) and Lemma 3.2, we have

$$\big|B_2(i)\big|\prec\frac{1+|\lambda|}{N\sqrt{N\eta_0}}+\frac{(1+|\lambda|)\Psi(z)}{N\sqrt{\eta_0}}=O_{\prec}\left(\frac{(1+|\lambda|)\Psi(z)}{N\sqrt{\eta_0}}\right).$$

Similarly, by plugging (5.3) in the expression of B_1, we have

$$B_1(i)=\frac{\mathrm{i}\lambda}{\pi N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\mathbb{E}\left[e_0(\lambda)\left(\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}(z')\frac{\mathrm{d}}{\mathrm{d}z'}\big(g_i(z')g_j(z')\big)\,\mathrm{d}^2z'\right)G_{ji}\right]+O_{\prec}\left(\frac{(1+|\lambda|^2)\Psi(z)}{N\sqrt{\eta_0}}\right).$$

Using the anisotropic local law, we have

$$B_1(i)=O_{\prec}\big((1+|\lambda|^2)N^{-1}\Psi(z)\big)+O_{\prec}\left(\frac{(1+|\lambda|^2)\Psi(z)}{N\sqrt{\eta_0}}\right)=O_{\prec}\left(\frac{(1+|\lambda|^2)\Psi(z)}{N\sqrt{\eta_0}}\right).$$

5.3. *Estimate on I_3*

It is not hard to show that the diagonal terms for $i=j$ are negligible. Thus we can replace the fourth cumulants by W_4-3. There are four terms in I_3 and we denote them as $D_1(i)$, $D_2(i)$, $D_3(i)$ and $D_4(i)$ respectively.

First, we look at D_1. By the local law and (5.4), we have $|D_1(i)|\prec(1+|\lambda|^3)N^{-1}\Psi(z)$. Similarly, using (5.5), (5.2) and the local law, we have $|D_3(i)|\prec\frac{(1+|\lambda|)\Psi(z)}{N\sqrt{N\eta_0}}$. For the last term D_4, using (5.5) and the local law, we obtain that

$$D_4(i)=\frac{1}{6N^2}\sum_{j=1}^{N}\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\big(36G_{ii}G_{jj}(G_{ij})^2+6(G_{ii})^2(G_{jj})^2+6(G_{ij})^4\big)\big]\prec N^{-1}\Psi(z).$$

Finally, we look at the leading term $D_2(i)$. Using the local law and (5.4), we have

$$D_2(i) = -\frac{W_4-3}{2N^2}\sum_{j=1}^{N}\mathbb{E}\bigg[\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}}\big((G_{ji})^2 + G_{ii}G_{jj}\big)\bigg]$$
$$= -\frac{W_4-3}{2N^2}\sum_{j=1}^{N}\mathbb{E}\bigg[\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}} g_i(z)g_j(z)\bigg] + O_{\prec}\big(\big(1+|\lambda|^2\big)N^{-1}\Psi(z)\big). \tag{5.11}$$

5.4. *Adding up the contributions to* (5.6)

Summing up the contributions from the previous subsections, we write (5.6) as

$$(z-a_i+m_{\mathrm{fc}})\mathbb{E}\big[e_0(\lambda)(G_{ii}-\mathbb{E}G_{ii})\big] = -\frac{1}{N}g_i(z)\mathbb{E}\big[e_0(\lambda)(\operatorname{Tr}G-\mathbb{E}\operatorname{Tr}G)\big] + A_{41}(i)+A_{42}(i)+D_2(i)+\epsilon(i),$$

where D_2 is given in (5.11) and A_{41}, A_{42} in (5.7), and $\epsilon(i)$ is the error term obtained in the previous subsections. Thanks to the stability bound (4.3), we can divide both sides by $z-a_i+m_{\mathrm{fc}}$ to get

$$\mathbb{E}\big[e_0(\lambda)(G_{ii}-\mathbb{E}G_{ii})\big] = \frac{1}{N}\big(g_i(z)\big)^2\mathbb{E}\big[e_0(\lambda)(\operatorname{Tr}G-\mathbb{E}\operatorname{Tr}G)\big] - g_i(z)\big(A_{41}(i)+A_{42}(i)+D_2(i)+\epsilon(i)\big).$$

Summing over i and rearranging, we find

$$\big(1-I_s(z)\big)\mathbb{E}\big[e_0(\lambda)(\operatorname{Tr}G-\mathbb{E}\operatorname{Tr}G)\big] = -\sum_{i=1}^{N} g_i(z)\big(A_{41}(i)+A_{42}(i)+D_2(i)\big)+\mathcal{E}_1, \tag{5.12}$$

where $\mathcal{E}_1$ is the linear statistics of $\epsilon(i)$. By the argument in Sections 5.1–5.3, we get

$$\mathcal{E}_1 = O_{\prec}\big(\big(1+|\lambda|^4\big)N\Psi(z)\Theta(z)\big) + O_{\prec}\big(\big(1+|\lambda|^4\big)\sqrt{N}\Psi^2(z)\big) + O_{\prec}\bigg(\frac{(1+|\lambda|^4)\Psi(z)}{\sqrt{\eta_0}}\bigg).$$

Next, we study the leading terms of the right side of (5.12). Plugging (5.2) in (5.7), we have

$$\sum_{i=1}^{N}\frac{A_{41}(i)}{z-a_i+m_{\mathrm{fc}}} = \frac{2\mathrm{i}\lambda}{\pi N}\sum_{i=1}^{N} g_i(z)\mathbb{E}\bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}\big(z'\big)\frac{\partial}{\partial z'}\big(G\big(z'\big)G(z)\big)_{ii}\,\mathrm{d}^2z'\bigg].$$

By the resolvent identity,

$$G(z)G\big(z'\big) = \frac{G(z)-G(z')}{z-z'}, \quad z\neq z', \tag{5.13}$$

we can write

$$F\big(z,z'\big) := \frac{1}{N}\sum_{i=1}^{N} g_i(z)\big(G\big(z'\big)G(z)\big)_{ii} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z'-z}g_i(z)\big(G_{ii}\big(z'\big)-G_{ii}(z)\big).$$

We separate into two cases:

Case 1: If z and z' belong to different half-planes, then we have $\frac{1}{|z-z'|}\leq\frac{1}{|\operatorname{Im}z|}$. Thus by the anisotropic local law, we have

$$\bigg|F\big(z,z'\big)-\frac{1}{z'-z}\frac{1}{N}\sum_{i=1}^{N}g_i(z)\big(g_i\big(z'\big)-g_i(z)\big)\bigg|$$
$$= \frac{1}{|z'-z|}\bigg|\frac{1}{N}\sum_{i=1}^{N}g_i(z)\big(G_{ii}\big(z'\big)-g_i\big(z'\big)\big)\bigg| + \frac{1}{|z'-z|}\bigg|\frac{1}{N}\sum_{i=1}^{N}g_i(z)\big(G_{ii}(z)-g_i(z)\big)\bigg|$$
$$= O_{\prec}\bigg(\frac{\Theta(z)+\Theta(z')}{|\operatorname{Im}z|}\bigg).$$

Case 2: If z and z' are in the same half-plane, without loss of generality, we can assume they both belong to the upper half plane. If $|\operatorname{Im} z - \operatorname{Im} z'| \geq \frac{1}{2}\operatorname{Im} z$, then we can use the same argument as in Case 1. Thus it is sufficient to study when $|\operatorname{Im} z - \operatorname{Im} z'| \leq \frac{1}{2}\operatorname{Im} z$, which means $\frac{2}{3}\operatorname{Im} z' \leq \operatorname{Im} z \leq 2\operatorname{Im} z'$. Note that

$$\begin{aligned}&\left|F\big(z,z'\big)-\frac{1}{z'-z}\frac{1}{N}\sum_{i=1}^{N}g_i(z)\big(g_i\big(z'\big)-g_i(z)\big)\right|\\&\quad\leq\left|\frac{\frac{1}{N}\sum_{i=1}^{N}(g_i(z)-g_i(z'))(G_{ii}(z')-g_i(z'))}{z-z'}\right|\\&\qquad+\left|\frac{\frac{1}{N}\sum_{i=1}^{N}g_i(z')(G_{ii}(z')-g_i(z'))-\frac{1}{N}\sum_{i=1}^{N}g_i(z)(G_{ii}(z)-g_i(z))}{z-z'}\right|.\end{aligned}$$

Next, we look at the first term on the right side. By direct computation, we get

$$\left|\frac{g_i(z)-g_i(z')}{z-z'}\right|\leq\big|g_i(z)\big|\big|g_i\big(z'\big)\big|\left(1+\left|\frac{m_{\mathrm{fc}}(z)-m_{\mathrm{fc}}(z')}{z-z'}\right|\right).$$

When z, z' are in the same half plane, m_{fc} is analytic in the neighborhood of the segment connecting z and z', denoted as $L(z,z')$. Thus

$$\left|\frac{m_{\mathrm{fc}}(z)-m_{\mathrm{fc}}(z')}{z-z'}\right|\leq\sup_{\omega\in L(z,z')}\big|m'_{\mathrm{fc}}(\omega)\big|\leq\frac{C}{\operatorname{Im} z}.$$

Combining with (4.3), we have

$$\left|\frac{g_i(z)-g_i(z')}{z-z'}\right|\leq\frac{C'}{\operatorname{Im} z}.$$

Using the second statement of the anisotropic local law by letting $B=\operatorname{Diag}(\frac{g_i(z)-g_i(z')}{z-z'})$ and the continuity argument as in (5.8), we obtain that the first term is bounded as $O_{\prec}(\frac{\Theta(z')}{\operatorname{Im} z})$.

For the second term, we write it as $\frac{h(z)-h(z')}{z-z'}$, where

$$h(z):=\frac{1}{N}\sum_{i=1}^{N}g_i(z)\big(G_{ii}(z)-g_i(z)\big).$$

Since h is analytic in the neighborhood of $L(z,z')$, we have

$$\left|\frac{h(z)-h(z')}{z-z'}\right|\leq\sup_{\omega\in L(z,z')}\left|\frac{\mathrm{d}}{\mathrm{d}\omega}h(\omega)\right|.$$

The anisotropic local law implies that $\sup_{w\in L(z,z')}|h(w)|\prec\Theta(z)$. Using the Cauchy integral formula, the second term is $O_{\prec}(\frac{\Theta(z)}{\operatorname{Im} z})$. Then we obtain the same upper bound as in Case 1.

Therefore, in both cases, we have

$$F\big(z,z'\big)=\frac{1}{z'-z}\left(\frac{1}{N}\sum_{i=1}^{N}g_i(z)g_i\big(z'\big)-\frac{1}{N}\sum_{i=1}^{N}g_i^2(z)\right)+O_{\prec}\left(\frac{\Theta(z)}{\operatorname{Im} z}\right)+O_{\prec}\left(\frac{\Theta(z')}{\operatorname{Im} z}\right).$$

Taking the derivative and using the Cauchy integral formula, we have

$$\frac{\partial}{\partial z'}F\big(z,z'\big)=\frac{\partial}{\partial z'}\left(\frac{1}{1-I(z,z')}\frac{\partial I(z,z')}{\partial z}\right)\big(1-I_s(z)\big)+O_{\prec}\left(\frac{\Theta(z)}{\operatorname{Im} z|\operatorname{Im} z'|}\right)+O_{\prec}\left(\frac{\Theta(z')}{\operatorname{Im} z|\operatorname{Im} z'|}\right).$$

Then by using Lemma 3.2, we have

$$\sum_{i=1}^{N}\frac{A_{41}(i)}{z-a_i+m_{\mathrm{fc}}}=\frac{2\mathrm{i}\lambda}{\pi}\mathbb{E}\big[e_0(\lambda)\big]\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}\big(z'\big)\frac{\partial}{\partial z'}\bigg(\frac{1}{1-I(z,z')}\frac{\partial I(z,z')}{\partial z}\big(1-I_s(z)\big)\bigg)\mathrm{d}^2z'$$
$$+O_{\prec}\bigg(\frac{\Theta(z)}{\operatorname{Im}z}\bigg)+O_{\prec}\bigg(\frac{1}{N\eta_0\operatorname{Im}z}\bigg).$$

Similarly, plugging (5.2) in (5.7), we have

$$\sum_{i=1}^{N}\frac{A_{42}(i)}{z-a_i+m_{\mathrm{fc}}}=\frac{(m_2-2)\mathrm{i}\lambda}{\pi}\mathbb{E}\big[e_0(\lambda)\big]\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}\big(z'\big)\frac{\partial}{\partial z'}\bigg(\frac{\partial I(z,z')}{\partial z}\big(1-I_s(z)\big)\bigg)\mathrm{d}^2z'+O_{\prec}\bigg(\frac{1}{\sqrt{N\eta_0}}\bigg).$$

Finally, plugging (5.3) in the leading term of (5.11) we have

$$\sum_{i=1}^{N}\frac{D_2(i)}{z-a_i+m_{\mathrm{fc}}}\frac{(W_4-3)\mathrm{i}\lambda}{\pi}\mathbb{E}\big[e_0(\lambda)\big]\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}\big(z'\big)\frac{\partial}{\partial z'}\bigg(\frac{\partial I(z,z')}{\partial z}\big(1-I_s(z)\big)I\big(z,z'\big)\bigg)\mathrm{d}^2z'+O_{\prec}\bigg(\frac{1+|\lambda|^2}{\sqrt{N\eta_0}}\bigg).$$

Therefore, we have

$$\big(1-I_s(z)\big)\mathbb{E}\big[e_0(\lambda)(\operatorname{Tr}G-\mathbb{E}\operatorname{Tr}G)\big]=\big(1-I_s(z)\big)\frac{\mathrm{i}\lambda}{\pi}\mathbb{E}\big[e_0(\lambda)\big]\cdot\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}\big(z'\big)K\big(z,z'\big)\mathrm{d}^2z'+\mathcal{E}_2,$$

where $K(z,z')$ is given by (2.16) and

$$\mathcal{E}_2=\big(1+|\lambda|^4\big)\bigg[O_{\prec}\big(N\Psi(z)\Theta(z)\big)+O_{\prec}\big(\big(\sqrt{N}\Psi^2(z)\big)+O_{\prec}\bigg(\frac{\Psi(z)}{\sqrt{\eta_0}}\bigg)+O_{\prec}\bigg(\frac{\Theta(z)}{\eta_0}\bigg)+O_{\prec}\bigg(\frac{\Theta(z)}{\eta}\bigg)\bigg].\tag{5.14}$$

Dividing both sides by $1-I_s(z)$, recalling from Lemma 4.4 that $|\frac{1}{1-I_s(z)}|\sim\frac{1}{\sqrt{\kappa+\eta}}$, and using (4.1), (4.2), we have completed the proof of Lemma 3.4. □

6. Proof of Theorem 2.10 and Theorem 2.11

In this section, we compute the variances of the mesoscopic CLT in the bulk and at the edges.

6.1. *In the bulk*

We compute the variance $V(f)$ defined in (2.15) with f given in (2.11).

Lemma 6.1. *Under the assumptions and notations of Theorem 2.8, we have*

$$\lim_{N\to\infty}V(f)=\frac{1}{2\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(g(x_1)-g(x_2))^2}{(x_1-x_2)^2}\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

Assuming that we have proved the above lemma, $V(f)$ converges to some positive constant since $g\in C_c^2(\mathbb{R})$. Theorem 2.10 is a direct result of Proposition 2.7 after integrating $\phi'(\lambda)$ and using the Arzelá-Ascoli theorem and the Lévy continuity theorem.

Proof of Lemma 6.1. Recall that

$$V(f)=\frac{1}{\pi^2}\int_{\Omega_0}\int_{\Omega_0}\frac{\partial}{\partial\overline{z_1}}\tilde{f}(z_1)\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)(K_1+K_2+K_3)\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2,$$

where

$$K_1 = \left(m_2 - \frac{2}{\beta}\right)\frac{\partial^2}{\partial z_1 \partial z_2} I; \qquad K_2 = \left(W_4 - 1 - \frac{2}{\beta}\right)\left(I\frac{\partial^2}{\partial z_1 \partial z_2} I + \frac{\partial}{\partial z_1} I \frac{\partial}{\partial z_2} I\right); \tag{6.1}$$

$$K_3 = \frac{2}{\beta}\frac{\partial}{\partial z_1}\left(\frac{1}{1-I}\frac{\partial I}{\partial z_2}\right) = \frac{2}{\beta}\left(\frac{1}{1-I(z_1,z_2)}\frac{\partial^2}{\partial z_1 \partial z_2} I + \frac{1}{(1-I(z_1,z_2))^2}\frac{\partial}{\partial z_1} I \frac{\partial}{\partial z_2} I\right). \tag{6.2}$$

Using Lemma 4.4 and the stability bound (4.3), we have

$$\frac{\partial}{\partial z_1} I(z_1,z_2) = \frac{1}{N}\sum_{i=1}^{N} \frac{1+m'_{\mathrm{fc}}(z_1)}{(a_i - z_1 - m_{\mathrm{fc}}(z_1))^2 (a_i - z_2 - m_{\mathrm{fc}}(z_2))} = O\left(\frac{1}{\sqrt{\kappa_1+\eta_1}}\right); \tag{6.3}$$

$$\frac{\partial}{\partial z_2} I(z_1,z_2) = \frac{1}{N}\sum_{i=1}^{N} \frac{1+m'_{\mathrm{fc}}(z_2)}{(a_i - z_1 - m_{\mathrm{fc}}(z_1))(a_i - z_2 - m_{\mathrm{fc}}(z_2))^2} = O\left(\frac{1}{\sqrt{\kappa_2+\eta_2}}\right); \tag{6.4}$$

$$\frac{\partial^2}{\partial z_1 \partial z_2} I(z_1,z_2) = \frac{1}{N}\sum_{i=1}^{N} \frac{(1+m'_{\mathrm{fc}}(z_1))(1+m'_{\mathrm{fc}}(z_2))}{(a_i - z_1 - m_{\mathrm{fc}}(z_1))^2 (a_i - z_2 - m_{\mathrm{fc}}(z_2))^2} = O\left(\frac{1}{\sqrt{(\kappa_1+\eta_1)(\kappa_2+\eta_2)}}\right). \tag{6.5}$$

In addition, recalling (4.7), for $z_1 \neq z_2$, we have

$$\frac{1}{1-I(z_1,z_2)} = 1 + \frac{m_{\mathrm{fc}}(z_1) - m_{\mathrm{fc}}(z_2)}{z_1 - z_2}. \tag{6.6}$$

If z_1 and z_2 are in the same half plane, m_{fc} is analytic in a neighborhood of the segment connecting z_1 and z_2, denoted as $L(z_1,z_2)$. By Lemma 4.4, then we have

$$\left|\frac{1}{1-I(z_1,z_2)}\right| \leq 1 + \left|\frac{m_{\mathrm{fc}}(z_1) - m_{\mathrm{fc}}(z_2)}{z_1 - z_2}\right| \leq \sup_{z\in L(z_1,z_2)} \left|m'_{\mathrm{fc}}(z)\right| \leq C \sup_{z\in L(z_1,z_2)} \left(\frac{1}{\sqrt{\kappa+\eta}}\right). \tag{6.7}$$

If z_1, z_2 belong to different half planes, using Lemma 4.4, then we have

$$\left|\frac{1}{1-I(z_1,z_2)}\right| \leq 1 + \left|\frac{m_{\mathrm{fc}}(z_1) - m_{\mathrm{fc}}(z_2)}{z_1 - z_2}\right| \leq \frac{C}{|z_1 - z_2|} \leq \frac{C}{|\eta_1| + |\eta_2|}. \tag{6.8}$$

Now, we are ready to compute $V(f)$. Since $\frac{\partial}{\partial \bar{z}} K_i(z,z') = \frac{\partial}{\partial z'} K_i(z,z') = 0$, $(i = 1,2,3)$, and by Stokes' formula, we have

$$V(f) = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2} \tilde{f}(z_1)\tilde{f}(z_2)(K_1 + K_2 + K_3)\,\mathrm{d}z_1\,\mathrm{d}z_2 := V_1 + V_2 + V_3,$$

where $\Gamma_1 = \{x_1 + \mathrm{i}y_1 : |y_1| = N^{-\tau}\eta_0\}$ and $\Gamma_2 = \{x_2 + \mathrm{i}y_2 : |y_2| = \frac{1}{2}N^{-\tau}\eta_0\}$. We choose the orientation of both contours to be counterclockwise. The parts on the upper half plane are denoted as Γ_1^+, Γ_2^+, while the parts on the lower half plane are Γ_1^-, Γ_2^-.

Using (6.3)–(6.5), since $\kappa \geq \kappa_0 \geq c_0$ for some positive constant $c_0 > 0$, we have $|K_1 + K_2| = O(1)$. Combining with (3.1), by direct computation, we have $|V_1 + V_2| = O(\eta_0^2)$. It then suffices to estimate V_3. We consider two cases.

Case 1: If z_1, z_2 are in the same half plane, by (6.7) and (6.3)–(6.5), we have $|K_3| = O(1)$. Therefore,

$$\left(\int_{\Gamma_1^+}\int_{\Gamma_2^+} + \int_{\Gamma_1^-}\int_{\Gamma_2^-}\right)\tilde{f}(z_1)\tilde{f}(z_2)K_3(z_1,z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2 = O\left(\eta_0^2\right).$$

Case 2: Consider z_1, z_2 are in different half planes. For notational simplicity, we define $m_1 = m_{\mathrm{fc}}(z_1)$ and $m_2 = m_{\mathrm{fc}}(z_2)$. Differentiating I given in (4.7), we have

$$\frac{\partial}{\partial z_1} I = \frac{(z_1 - z_2)m'_1 - m_1 + m_2}{(z_1 + m_1 - z_2 - m_2)^2}; \qquad \frac{\partial}{\partial z_2} I = \frac{(z_2 - z_1)m'_2 + m_1 - m_2}{(z_1 + m_1 - z_2 - m_2)^2}. \tag{6.9}$$

Using (6.8) (6.6), (6.3)–(6.5) and Lemma 4.4, we have

$$K_3 = \frac{2}{\beta}\frac{1}{(z_1 - z_2)^2}\frac{((z_1 - z_2)m'_1 - m_1 + m_2)((z_2 - z_1)m'_2 + m_1 - m_2)}{(z_1 + m_1 - z_2 - m_2)^2} + O\left(\eta_0^{-1}N^{\tau}\right).$$

Note that if $z \in \mathbb{C}^+$ and in the bulk, then there exists $k, K > 0$ such that $k \le \operatorname{Im} m_{\mathrm{fc}}(z) \le K$. If z_1, z_2 are in different half planes, there exists some constant $c > 0$ such that $|z_1 + m_1 - z_2 - m_2| > c$. Combining with Lemma 4.4, we have

$$K_3 = -\frac{2}{\beta}\frac{(m_1 - m_2)^2}{(z_1 - z_2)^2(z_1 + m_1 - z_2 - m_2)^2} + O_{\prec}\big(\eta_0^{-1}N^{\tau}\big) + O(1) = -\frac{2}{\beta}\frac{1}{(z_1 - z_2)^2} + O\big(\eta_0^{-1}N^{\tau}\big).$$

Therefore, recalling the definition of $\tilde{f}$ in (3.2), by symmetry and (3.1), we have

$$V_3 = \frac{1}{\beta\pi^2}\int_{\Gamma_1^+}\int_{\Gamma_2^-}\frac{\tilde{f}(z_1)\tilde{f}(z_2)}{(z_1 - z_2)^2}\,\mathrm{d}z_1\,\mathrm{d}z_2 + O\big(\eta_0 N^{\tau}\big), \tag{6.10}$$

with opposite integral directions on the contours. Since Γ_1^+ and Γ_2^- are disjoint and $\tilde{f}$ has compact support, we obtain from Cauchy's integral theorem that

$$\frac{1}{\beta\pi^2}\int_{\Gamma_1^+}\int_{\Gamma_2^-}\frac{\tilde{f}(z_2)^2}{(z_1 - z_2)^2}\,\mathrm{d}z_1\,\mathrm{d}z_2 = \frac{1}{\beta\pi^2}\int_{\Gamma_2^-}\tilde{f}(z_2)^2\bigg(\int_{\Gamma_1^+}\frac{1}{(z_1 - z_2)^2}\,\mathrm{d}z_1\bigg)\mathrm{d}z_2 = 0.$$

The integral of $\frac{\tilde{f}(z_1)^2}{(z_1 - z_2)^2}$ vanishes similarly. Thus, we have from (6.10) and (3.2) that

$$\begin{aligned} V_3 &= -\frac{1}{2\beta\pi^2}\int_{\Gamma_1^+}\int_{\Gamma_2^-}\frac{(\tilde{f}(z_1) - \tilde{f}(z_2)^2}{(z_1 - z_2)^2}\,\mathrm{d}z_1\,\mathrm{d}z_2 + O\big(\eta_0 N^{\tau}\big) \\ &= \frac{1}{2\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(f(x_1) - f(x_2) + \mathrm{i}N^{-\tau}\eta_0(f'(x_1) - f'(x_2)))^2}{(x_1 - x_2 + \frac{3\mathrm{i}}{2}N^{-\tau}\eta_0)^2}\,\mathrm{d}x_1\,\mathrm{d}x_2 + O\big(\eta_0 N^{\tau}\big). \end{aligned} \tag{6.11}$$

Changing the variable

$$\tilde{x}_1 = \frac{x_1 - E_0}{\eta_0}; \qquad \tilde{x}_2 = \frac{x_2 - E_0}{\eta_0}, \tag{6.12}$$

we hence obtain from (6.11) that

$$V_3 = \frac{1}{2\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(g(\tilde{x}_1) - g(\tilde{x}_2) + \mathrm{i}N^{-\tau}(g'(\tilde{x}_1) - g'(\tilde{x}_2)))^2}{(\tilde{x}_1 - \tilde{x}_2 + \frac{3\mathrm{i}}{2}N^{-\tau})^2}\,\mathrm{d}\tilde{x}_1\,\mathrm{d}\tilde{x}_2 + O\big(\eta_0 N^{\tau}\big).$$

Note that the integrand can be bounded uniformly by

$$\bigg|\frac{(g(\tilde{x}_1) - g(\tilde{x}_2) + \mathrm{i}N^{-\tau}(g'(\tilde{x}_1) - g'(\tilde{x}_2)))^2}{(\tilde{x}_1 - \tilde{x}_2 + \frac{3\mathrm{i}}{2}N^{-\tau})^2}\bigg| \le \frac{(g(\tilde{x}_1) - g(\tilde{x}_2))^2}{(\tilde{x}_1 - \tilde{x}_2)^2} + \frac{(g'(\tilde{x}_1) - g'(\tilde{x}_2))^2}{(\tilde{x}_1 - \tilde{x}_2)^2}. \tag{6.13}$$

Since $g \in C_c^2(\mathbb{R})$, we then conclude the proof using dominated convergence. □

6.2. *Near the edge*

Theorem 2.11 is a result of the following lemma and Proposition 2.7.

Lemma 6.2. *Under the assumptions and notations of Theorem 2.8, we have*

$$\lim_{N\to\infty} V(f) = \frac{1}{2\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\bigg(\frac{g(-x^2) - g(-y^2)}{x - y}\bigg)^2\,\mathrm{d}x\,\mathrm{d}y.$$

Proof of Lemma 6.2. Similarly as in the bulk, using Stokes' formula, we have

$$V(f) = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2}\tilde{f}(z_1)\tilde{f}(z_2)(K_1 + K_2 + K_3)\,\mathrm{d}z_1\,\mathrm{d}z_2 := V_1 + V_2 + V_3,$$

where K_1, K_2, K_3 are given by (6.1) and (6.2), and we use the same notations and definitions as in the previous subsection. Using (6.3)–(6.5) and Lemma 4.4, we have

$$|K_1 + K_2| = O\bigg(\frac{1}{\sqrt{|\operatorname{Im} z_1 \operatorname{Im} z_2|}}\bigg) = O\big(N^{\tau}\eta_0^{-1}\big).$$

Using (3.1), one can show $|V_1 + V_2| = O(\eta_0 N^{\tau})$. Thus it is sufficient to study the integral involving K_3. Using (6.9), (6.6), and

$$\frac{\partial^2}{\partial z \partial z'} I\big(z, z'\big) = \frac{\partial^2}{\partial z' \partial z} I\big(z, z'\big) = \frac{(m_1' + m_2' + 2m_1' m_2')(z_1 - z_2) - (m_1' + m_2' + 2)(m_1 - m_2)}{(z_1 + m_1 - z_2 - m_2)^3},$$

by direct computation, we have

$$K_3 = \frac{2}{\beta}\bigg(\frac{(1 + m_1')(1 + m_2')}{(z_1 + m_1 - z_2 - m_2)^2} - \frac{1}{(z_1 - z_2)^2}\bigg).$$

For the second integrand, using the similar arguments as in the previous subsection, we have

$$\begin{aligned}\lim_{N\to\infty} \int_{\Gamma_1^{\pm}} \int_{\Gamma_2^{\pm}} \frac{\tilde f(z_1)\tilde f(z_2)}{(z_1 - z_2)^2}\,\mathrm{d}z_1\,\mathrm{d}z_2 &= -\frac{1}{2}\lim_{N\to\infty} \int_{\Gamma_1^{\pm}} \int_{\Gamma_2^{\pm}} \frac{(\tilde f(z_1) - \tilde f(z_2))^2}{(z_1 - z_2)^2}\,\mathrm{d}z_1\,\mathrm{d}z_2 \\ &= -\frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2)^2}\,\mathrm{d}x_1\,\mathrm{d}x_2. \end{aligned} \tag{6.14}$$

Due to the opposite integral directions, we have

$$\lim_{N\to\infty} \int_{\Gamma_1^{\pm}} \int_{\Gamma_2^{\mp}} \frac{\tilde f(z_1)\tilde f(z_2)}{(z_1 - z_2)^2}\,\mathrm{d}z_1\,\mathrm{d}z_2 = \frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2)^2}\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

The whole integral with respect to the second term of K_3 will hence vanish when $N \to \infty$. Thus it is sufficient to study the integral of the first term $\frac{(1+m_1')(1+m_2')}{(z_1+m_1-z_2-m_2)^2}$, that is,

$$\begin{aligned} V_3(f) &= -\frac{1}{2\beta\pi^2}\bigg(\int_{\Gamma_1^+}\int_{\Gamma_2^+} + \int_{\Gamma_1^-}\int_{\Gamma_2^-} + \int_{\Gamma_1^+}\int_{\Gamma_2^-} + \int_{\Gamma_1^-}\int_{\Gamma_2^+}\bigg)\tilde f(z_1)\tilde f(z_2)\frac{(1 + m_1')(1 + m_2')}{(z_1 + m_1 - z_2 - m_2)^2}\,\mathrm{d}z_1\,\mathrm{d}z_2 \\ &:= V_3^{++} + V_3^{--} + V_3^{+-} + V_3^{-+}. \end{aligned}$$

Let $\zeta = z + m_{\mathrm{fc}}(z)$ and $\zeta_{\pm} = L_{\pm} + m_{\mathrm{fc}}(L_{\pm}) \in \mathbb{R}$. Define $F(\zeta) := \zeta - \frac{1}{N}\sum_{i=1}^N \frac{1}{a_i - \zeta}$ so that (2.6) is equivalent to $z = F(\zeta)$. Assumptions 2.2 and 2.3 imply that there exists some constant c_0 independent of N such that

$$\operatorname{dist}\big(\{\zeta_{\pm}, \hat{\mathcal{I}}\}\big) \geq c_0,$$

for all sufficiently large N, where $\hat{\mathcal{I}}$ is the smallest interval that contains the support of μ_A; see (4.3). Hence, $F(\zeta)$ is analytic in a neighborhood of ζ_+, where we write

$$F(\zeta) = F(\zeta_+) + F'(\zeta_+)(\zeta - \zeta_+) + \frac{F''(\zeta_+)}{2}(\zeta - \zeta_+)^2 + O\big(|\zeta - \zeta_+|^3\big).$$

By (2.7), $F'(\zeta_+) = 1 - \frac{1}{N}\sum_{i=1}^N \frac{1}{(a_i - \zeta_+)^2} = 0$. Moreover, $F''(\zeta_+) = -\frac{2}{N}\sum_{i=1}^N \frac{1}{(a_i - \zeta_+)^3}$, and by (4.5) it is bounded uniformly from below. In general, we have $|F^{(k)}(\zeta_+)| = |\frac{k!}{N}\sum_{i=1}^N \frac{1}{(a_i - \zeta_+)^{(k+1)}}| = O(1)$ because of (4.4). Inverting $F(\zeta) = z$ in the neighborhood of ζ_+, we have the expansion

$$\zeta = z + m_{\mathrm{fc}}(z) = \zeta_+ + c_+\sqrt{z - L_+}\big(1 + A_+(\sqrt{z - L_+})\big), \tag{6.15}$$

where A_+ is an analytic function depending on N with $A_+(0) = 0$. This has been shown in the proof of Lemma 3.6 and Lemma A.1 in [52]. Note that $c_+ = (-\frac{1}{N}\sum_{i=1}^N \frac{1}{(a_i - \zeta_+)^3})^{-\frac{1}{2}}$ is some positive number depending on N but is uniformly bounded. Furthermore, the coefficients of the expansion of A_+ are also uniformly bounded. Thus

$$z + m_{\mathrm{fc}}(z) = \zeta_+ + c_+\sqrt{z - L_+} + O\big(|z - L_+|\big),$$

where the square root is taken in a branch cut such that $\operatorname{Im}\sqrt{z-L_+}>0$ as $\operatorname{Im} z>0$. Similarly, we have

$$1+m'_{\mathrm{fc}}(z)=\frac{c_+}{2\sqrt{z-L_+}}+d_+ +O\big(\sqrt{|z-L_+|}\big),$$

where d_+ is some number which depends on N but is uniformly bounded. Let $z=L_+ +\eta_0 x+\mathrm{i}N^{-\tau}\eta_0$. Then

$$z+m_{\mathrm{fc}}(z)=\zeta_+ +c_+\sqrt{\eta_0\big(x+\mathrm{i}N^{-\tau}\big)}+O(\eta_0);\qquad 1+m'_{\mathrm{fc}}(z)=\frac{c_+}{2\sqrt{\eta_0(x+\mathrm{i}N^{-\tau})}}+d_+ +O(\sqrt{\eta_0}).$$

Therefore, after changing the variable as in (6.12), we have

$$\begin{aligned}V_3^{++}&=-\frac{1}{8\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\tilde g(x_1)\tilde g(x_2)(\frac{1}{\sqrt{x_1+\mathrm{i}N^{-\tau}}}+O(\sqrt{\eta_0}))(\frac{1}{\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}}}+O(\sqrt{\eta_0}))}{(\sqrt{x_1+\mathrm{i}N^{-\tau}}-\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}}+O(\sqrt{\eta_0}))^2}\,\mathrm{d}x_1\,\mathrm{d}x_2\\&=-\frac{1}{8\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\tilde g(x_1)\tilde g(x_2)}{\sqrt{x_1+\mathrm{i}N^{-\tau}}\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}}(\sqrt{x_1+\mathrm{i}N^{-\tau}}-\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}})^2}\,\mathrm{d}x_1\,\mathrm{d}x_2+O\big(\sqrt{\eta_0}N^{3\tau}\big),\end{aligned}$$

where $\tilde g(x)=g(x)+\mathrm{i}N^{-\tau}g'(x)$. The last step follows from the fact that $|\sqrt{x_1+\mathrm{i}N^{-\tau}}-\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}}|\geq CN^{-\tau}$, when x_1, x_2 belong to some compact set. Let $\gamma_1^{\pm}:=\{x_1\pm\mathrm{i}N^{-\tau}:x_1\in\mathbb{R}\}$ and $\gamma_2^{\pm}:=\{x_2\pm\frac{\mathrm{i}}{2}N^{-\tau}:x_2\in\mathbb{R}\}$. Then we obtain

$$V_3^{++}=-\frac{1}{8\beta\pi^2}\int_{\gamma_1^+}\int_{\gamma_2^+}\frac{\tilde g(z_1)\tilde g(z_2)}{\sqrt{z_1}\sqrt{z_2}(\sqrt{z_1}-\sqrt{z_2})^2}\,\mathrm{d}z_1\,\mathrm{d}z_2+O\big(\sqrt{\eta_0}N^{3\tau}\big),$$

where $\tilde g(x+\mathrm{i}y)=g(x)+\mathrm{i}yg'(x)\chi(y)$. Since γ_1^+ and γ_2^+ are disjoint and $\tilde g$ has compact support, changing the variable $w=\sqrt{z}$ and using Cauchy's integral theorem, we have

$$\int_{\gamma_1^+}\int_{\gamma_2^+}\frac{\tilde g(z_1)^2}{\sqrt{z_1}\sqrt{z_2}(\sqrt{z_1}-\sqrt{z_2})^2}\,\mathrm{d}z_1\,\mathrm{d}z_2=0=\int_{\gamma_1^+}\int_{\gamma_2^+}\frac{\tilde g(z_2)^2}{\sqrt{z_1}\sqrt{z_2}(\sqrt{z_1}-\sqrt{z_2})^2}\,\mathrm{d}z_1\,\mathrm{d}z_2,$$

and thus

$$V_3^{++}=\frac{1}{16\beta\pi^2}\int_{\gamma_1^+}\int_{\gamma_2^+}\frac{(\tilde g(z_1)-\tilde g(z_2))^2}{\sqrt{z_1}\sqrt{z_2}(\sqrt{z_1}-\sqrt{z_2})^2}\,\mathrm{d}z_1\,\mathrm{d}z_2+O\big(\sqrt{\eta_0}N^{3\tau}\big).$$

Therefore, we get

$$\lim_{N\to\infty}V_3^{++}=\frac{1}{16\beta\pi^2}\lim_{N\to\infty}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(g(x_1)-g(x_2)+\mathrm{i}N^{-\tau}g'(x_1)-\frac{\mathrm{i}}{2}N^{-\tau}g'(x_2))^2}{\sqrt{x_1+\mathrm{i}N^{-\tau}}\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}}(\sqrt{x_1+\mathrm{i}N^{-\tau}}-\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}})^2}\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

We denote the integrand as $h_N(x_1,x_2)$. Next, we interchange the limit and the integral. One shows that there exists $C>0$ such that

$$\left|\sqrt{x_1+\mathrm{i}N^{-\tau}}-\sqrt{x_2+\frac{\mathrm{i}}{2}N^{-\tau}}\right|\geq C|\sqrt{x_1}-\sqrt{x_2}|.$$

Set

$$h(x_1,x_2):=C^{-2}\frac{(g(x_1)-g(x_2))^2+(g'(x_1)-g'(x_2))^2}{\sqrt{|x_1|}\sqrt{|x_2|}|\sqrt{x_1}-\sqrt{x_2}|^2},$$

and observe that $|h_N(x_1,x_2)|\leq h(x_1,x_2)$. Next, we will show that $h(x_1,x_2)$ is integrable.

Suppose $\operatorname{supp}(g)\subset[-M,M]$ for some $M>0$. Then if x_1 and x_2 are both in $[-2M,2M]$ then we have the following estimation.

Case 1: If x_1, x_2 have the same sign, then

$$\begin{aligned} h(x_1,x_2) &= \frac{(g(x_1)-g(x_2))^2+(g'(x_1)-g'(x_2))^2}{\sqrt{x_1}\sqrt{x_2}(\sqrt{x_1}-\sqrt{x_2})^2} \\ &= \frac{1}{\sqrt{|x_1|}\sqrt{|x_2|}}\left(\left(\frac{g(x_1)-g(x_2)}{x_1-x_2}\right)^2+\left(\frac{g'(x_1)-g'(x_2)}{x_1-x_2}\right)^2\right)\left(\sqrt{|x_1|}+\sqrt{|x_2|}\right)^2 \\ &\le \frac{8M}{\sqrt{|x_1|}\sqrt{|x_2|}}\left(\|g'\|_\infty^2+\|g''\|_\infty^2\right). \end{aligned}$$

Case 2: If x_1 and x_2 are of opposite signs, using $|x_1-x_2| = (\sqrt{|x_1|}-\mathrm{i}\sqrt{|x_2|})(\sqrt{|x_1|}+\mathrm{i}\sqrt{|x_2|})$, we have

$$\begin{aligned} h(x_1,x_2) &= \frac{(g(x_1)-g(x_2))^2+(g'(x_1)-g'(x_2))^2}{\sqrt{|x_1|}\sqrt{|x_2|}|\sqrt{|x_1|}-\mathrm{i}\sqrt{|x_2|}|^2} \\ &= \frac{1}{\sqrt{|x_1|}\sqrt{|x_2|}}\left(\left(\frac{g(x_1)-g(x_2)}{x_1-x_2}\right)^2+\left(\frac{g'(x_1)-g'(x_2)}{x_1-x_2}\right)^2\right)\left|\sqrt{|x_1|}+\mathrm{i}\sqrt{|x_2|}\right|^2 \\ &\le \frac{8M}{\sqrt{|x_1|}\sqrt{|x_2|}}\left(\|g'\|_\infty^2+\|g''\|_\infty^2\right). \end{aligned}$$

If $x_1 \notin [-2M, 2M]$, then $x_2 \in [-M, M]$ otherwise $h(x_1, x_2) = 0$. So for $(x_1, x_2) \in [-2M, 2M]^c \times [-M, M]$,

$$h(x_1,x_2) \le \frac{4\|g\|_\infty^2+4\|g'\|_\infty^2}{\sqrt{|x_1|}\sqrt{|x_2|}(\sqrt{|x_1|}-\sqrt{|x_2|})^2} \le \frac{4\|g\|_\infty^2+4\|g'\|_\infty^2}{\sqrt{|x_1|}\sqrt{|x_2|}(\sqrt{|x_1|}-\frac{1}{\sqrt{2}}\sqrt{|x_1|})^2} = \frac{C}{|x_1|^{3/2}|x_2|^{1/2}}.$$

Therefore, $h(x_1, x_2)$ is integrable. Thus by dominated convergence,

$$\begin{aligned} \lim_{N\to\infty} V_3^{++} &= \frac{1}{16\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{(g(x_1)-g(x_2))^2}{\sqrt{x_1+\mathrm{i}0}\sqrt{x_2+\mathrm{i}0}(\sqrt{x_1+\mathrm{i}0}-\sqrt{x_2+\mathrm{i}0})^2}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \frac{1}{4\beta\pi^2}\int_{\psi(\mathbb{R}+\mathrm{i}0)}\int_{\psi(\mathbb{R}+\mathrm{i}0)} \frac{(g(w_1^2)-g(w_2^2))^2}{(w_1-w_2)^2}\,\mathrm{d}w_1\,\mathrm{d}w_2, \end{aligned}$$

where we change the variable $\psi(z) := \sqrt{z}$; with branch cut such that $\psi : \mathbb{C}^+ \to \mathbb{C}^+$.

Similarly, we have

$$\begin{aligned} \lim_{N\to\infty} V_3^{--} &= \frac{1}{4\beta\pi^2}\int_{\psi(\mathbb{R}-\mathrm{i}0)}\int_{\psi(\mathbb{R}-\mathrm{i}0)} \frac{(g(w_1^2)-g(w_2^2))^2}{(w_1-w_2)^2}\,\mathrm{d}w_1\,\mathrm{d}w_2; \\ \lim_{N\to\infty} V_3^{+-} &= \frac{1}{4\beta\pi^2}\int_{\psi(\mathbb{R}+\mathrm{i}0)}\int_{\psi(\mathbb{R}-\mathrm{i}0)} \frac{(g(w_1^2)-g(w_2^2))^2}{(w_1-w_2)^2}\,\mathrm{d}w_1\,\mathrm{d}w_2; \\ \lim_{N\to\infty} V_3^{-+} &= \frac{1}{4\beta\pi^2}\int_{\psi(\mathbb{R}-\mathrm{i}0)}\int_{\psi(\mathbb{R}+\mathrm{i}0)} \frac{(g(w_1^2)-g(w_2^2))^2}{(w_1-w_2)^2}\,\mathrm{d}w_1\,\mathrm{d}w_2. \end{aligned}$$

The contours are shown in Figure 1. Note that the horizontal parts of the blue and the red lines of above graph will cancel because of the opposite integral direction. To sum up, we have

$$\lim_{N\to\infty} V_3 = \frac{1}{4\beta\pi^2}\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{(g(w_1^2)-g(w_2^2))^2}{(w_1-w_2)^2}\,\mathrm{d}w_1\,\mathrm{d}w_2 = \frac{1}{4\beta\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{g(-x_1^2)-g(-x_2^2)}{x_1-x_2}\right)^2\mathrm{d}x_1\,\mathrm{d}x_2.$$

This concludes the proof of Theorem 2.11. □

7. Proof of Proposition 2.9 and computation of the bias

In this section, we first prove Proposition 2.9, using the same technique as in Proposition 2.7. After this, we compute the bias on mesoscopic scales inside the bulk and at the edges.

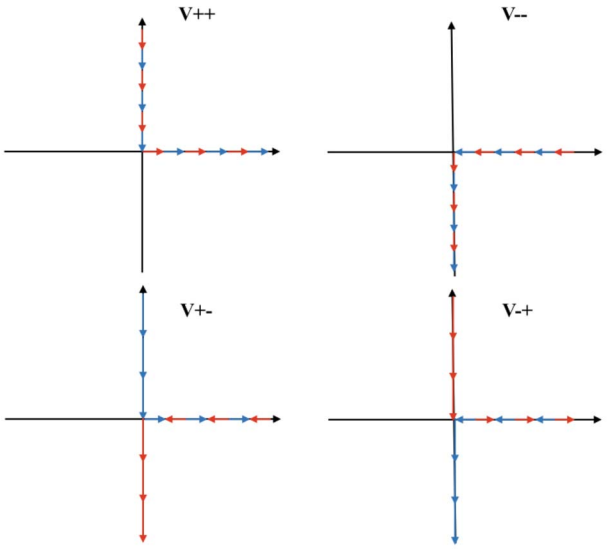

Fig. 1. Integration contours in the variance term V_3.

Proof of Proposition 2.9. We treat the expectation similarly using the cumulant expansion and (5.5):

$$\begin{aligned}(z-a_i)\mathbb{E}G_{ii} &= \mathbb{E}(HG)_{ii} - 1\\ &= \frac{1}{N}\mathbb{E}\sum_{j=1}^{N} c_{ij}^{(2)}\frac{\partial G_{ji}}{\partial H_{ij}} - 1 + \frac{1}{2!N^{\frac{3}{2}}}\sum_{j=1}^{N} c_{ij}^{(3)}\mathbb{E}\frac{\partial^2 G_{ji}}{\partial^2 H_{ij}} + \frac{1}{3!N^2}\sum_{j=1}^{N} c_{ij}^{(4)}\mathbb{E}\frac{\partial^3 G_{ji}}{\partial^3 H_{ij}} + O_{\prec}\big(N^{-\frac{3}{2}}\big)\\ &= -\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}G_{ii}G_{jj} - \frac{1}{N}\mathbb{E}\big(G^2\big)_{ii} - \frac{m_2-2}{N}\mathbb{E}(G_{ii})^2 - 1 + \frac{1}{2N^{\frac{3}{2}}}\sum_{j=1}^{N} c_{ij}^{(3)}\mathbb{E}\big(6G_{ii}G_{ij}G_{jj} + 2G_{ij}^3\big)\\ &\quad + \frac{1}{6N^2}\sum_{j=1}^{N}(W_4-3)\mathbb{E}\big(-36G_{ii}G_{jj}G_{ij}^2 - 6G_{ii}^2G_{jj}^2 - 6G_{ij}^4\big) + O_{\prec}\big(N^{-\frac{3}{2}}\big).\end{aligned}$$

Combining with the local law, we have

$$\begin{aligned}(z-a_i)\mathbb{E}G_{ii} &= -\frac{1}{N}\mathbb{E}G_{ii}\operatorname{Tr}G - \frac{1}{N}\frac{d}{dz}\frac{1}{a_i - z - m_{\mathrm{fc}}} - \frac{m_2-2}{N}\frac{1}{(a_i - z - m_{\mathrm{fc}})^2} - 1\\ &\quad + \frac{3}{N^{\frac{3}{2}}}\sum_{j=1}^{N}\frac{c_{ij}^{(3)}}{(a_i - z - m_{\mathrm{fc}})(a_j - z - m_{\mathrm{fc}})}G_{ij} - \frac{1}{N^2}\sum_{j=1}^{N}(W_4-3)\frac{1}{(a_i - z - m_{\mathrm{fc}})^2(a_j - z - m_{\mathrm{fc}})^2}\\ &\quad + O_{\prec}\left(\frac{\Psi(z)}{N\eta}\right).\end{aligned}$$

Using the anisotropic local law and the argument as in (5.8), one can show that the second term of the last line of above equation is $O_{\prec}(N^{-1}\Psi(z))$. Therefore, we have

$$\begin{aligned}(z-a_i)\mathbb{E}G_{ii} &= -\frac{1}{N}\mathbb{E}\left(G_{ii} - \frac{1}{a_i - z - m_{\mathrm{fc}}}\right)\operatorname{Tr}G - \frac{1}{N}\frac{1}{a_i - z - m_{\mathrm{fc}}}\mathbb{E}\operatorname{Tr}G - 1 - \frac{1}{N}\frac{1 + m_{\mathrm{fc}}'(z)}{(a_i - z - m_{\mathrm{fc}})^2}\\ &\quad - \frac{m_2-2}{N}\frac{1}{(a_i - z - m_{\mathrm{fc}})^2} - \frac{1}{N}I_s(z)\frac{W_4-3}{(a_i - z - m_{\mathrm{fc}})^2} + O_{\prec}\left(\frac{\Psi(z)}{N\eta}\right),\end{aligned}$$

and thus

$$(z - a_i + m_{\mathrm{fc}})\left(\mathbb{E}G_{ii} - \frac{1}{a_i - z - m_{\mathrm{fc}}}\right) = -\frac{1}{N}\frac{1}{a_i - z - m_{\mathrm{fc}}}(\mathbb{E}\operatorname{Tr}G - Nm_{\mathrm{fc}})$$

$$-\frac{1}{N}\frac{1+m_{\mathrm{fc}}'(z)}{(a_i-z-m_{\mathrm{fc}})^2}-\frac{m_2-2}{N}\frac{1}{(a_i-z-m_{\mathrm{fc}})^2}$$
$$-\frac{1}{N}I_s(z)\frac{W_4-3}{(a_i-z-m_{\mathrm{fc}})^2}+O_{\prec}\Big(\frac{\Psi(z)}{N\eta}\Big).$$

Dividing both sides by $a_i-z-m_{\mathrm{fc}}\sim O(1)$ and summing over i, we obtain

$$\big(1-I_s(z)\big)\mathbb{E}(\operatorname{Tr}G-Nm_{\mathrm{fc}})=\frac{1}{N}\sum_{i=1}^{N}\frac{1+m_{\mathrm{fc}}'(z)}{(a_i-z-m_{\mathrm{fc}})^3}+\frac{m_2-2}{N}\sum_{i=1}^{N}\frac{1}{(a_i-z-m_{\mathrm{fc}})^3}$$
$$+\frac{W_4-3}{N}\sum_{i=1}^{N}\frac{I_s(z)}{(a_i-z-m_{\mathrm{fc}})^3}+O_{\prec}\Big(\frac{\Psi(z)}{\eta}\Big).$$

Dividing both sides by $1-I_s(z)$ and using the relation $1-I_s(z)=\frac{1}{1+m_{\mathrm{fc}}'(z)}\sim\sqrt{\kappa+\eta}$, we obtain

$$\mathbb{E}(\operatorname{Tr}G-Nm_{\mathrm{fc}})=\frac{1}{1-I_s(z)}\frac{1}{2}\frac{\mathrm{d}I_s(z)}{\mathrm{d}z}+\frac{m_2-2}{2}\frac{\mathrm{d}I_s(z)}{\mathrm{d}z}+\frac{W_4-3}{2}I_s(z)\frac{\mathrm{d}I_s(z)}{\mathrm{d}z}+O_{\prec}\Big(\frac{1}{\eta\sqrt{N\eta\sqrt{\kappa+\eta}}}\Big).$$

Plugging into (3.11) (here we replace $\mathbb{E}\mu_N$ by μ_{fc}), using Lemma 3.2 and Stokes' formula, we have

$$\mathbb{E}\operatorname{Tr}f(X_N)-N\int_{\mathbb{R}}f(x)\rho_{\mathrm{fc}}(x)\,\mathrm{d}x=\frac{1}{4\pi\mathrm{i}}\int_{\partial\Omega_0}\tilde{f}(z)b(z)\,\mathrm{d}z+O_{\prec}\Big(\frac{N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}\Big)+O_{\prec}\big(N^{-\tau}\big),$$

where $b(z)$ is given by (2.19). Using the relation $I_s=\frac{m_{\mathrm{fc}}'}{1+m_{\mathrm{fc}}'}$, it coincides with the expectation that obtained in the global CLT given in Theorem 2.6. Thus we conclude the proof of Proposition 2.9. □

Next, we explicitly compute the bias in the bulk and at the edges, for the scaled test function in (2.11).

7.1. *Bias in the mesoscopic bulk*

Note that

$$\frac{dI_s}{dz}=\frac{2}{N}\sum_{i=1}^{N}\frac{1+m_{\mathrm{fc}}'}{(a_i-z-m_{\mathrm{fc}})^3}=O\Big(\frac{1}{\sqrt{\kappa+\eta}}\Big);\qquad 1-I_s(z)\sim\frac{1}{\sqrt{\kappa+\eta}};\qquad \big|I_s(z)\big|=O(1).\tag{7.1}$$

If $\kappa\geq\kappa_0>c>0$, then $|b(z)|=O(1)$. In combination with (3.1), we have

$$\mathbb{E}\operatorname{Tr}f(X_N)-N\int_{\mathbb{R}}f(x)\rho_{\mathrm{fc}}(x)\,\mathrm{d}x=O_{\prec}\Big(\eta_0+\frac{N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}+N^{-\tau}\Big),$$

hence we see that the bias vanishes as N goes to infinity.

7.2. *Bias at the mesoscopic edge*

Similarly, using (7.1) and (3.1), the last two terms of $b(z)$ will contribute $O_{\prec}(\sqrt{N^{\tau}\eta_0})$. We have

$$\mathbb{E}\operatorname{Tr}f(X_N)-N\int_{\mathbb{R}}f(x)\rho_{\mathrm{fc}}(x)\,\mathrm{d}x=\frac{1}{4\pi\mathrm{i}}\int_{\partial\Omega_0}\tilde{f}(z)\frac{m_{\mathrm{fc}}''}{1+m_{\mathrm{fc}}'}\,\mathrm{d}z+O_{\prec}\Big(N^{-\tau}+\frac{N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}+\sqrt{N^{\tau}\eta_0}\Big).$$

Using (6.15), we obtain the following expansions:

$$1+m_{\mathrm{fc}}'(z)=\frac{c_+}{2\sqrt{z-L_+}}+O(1),\qquad m_{\mathrm{fc}}''(z)=-\frac{c_+}{4(\sqrt{z-L_+})^3}+O\Big(\frac{1}{\sqrt{|z-L_+|}}\Big),$$

and then

$$\frac{m''_{\mathrm{fc}}}{1+m'_{\mathrm{fc}}} = -\frac{1}{2(z-L_+)} + O\bigg(\frac{1}{\sqrt{|z-L_+|}}\bigg).$$

Changing variables and using (3.1), we have

$$\begin{aligned}
&\mathbb{E}\,\mathrm{Tr}\, f(X_N) - N\int_{\mathbb{R}} f(x)\rho_{\mathrm{fc}}(x)\,\mathrm{d}x \\
&\quad = -\frac{1}{8\pi\mathrm{i}}\int_{\mathbb{R}}\big(g(x)+\mathrm{i}N^{-\tau}g'(x)\big)\frac{1}{x+\mathrm{i}N^{-\tau}}\,\mathrm{d}x \\
&\qquad + \frac{1}{8\pi\mathrm{i}}\int_{\mathbb{R}}\big(g(x)-\mathrm{i}N^{-\tau}g'(x)\big)\frac{1}{x-\mathrm{i}N^{-\tau}}\,\mathrm{d}x + O_{\prec}\bigg(N^{-\tau}+\frac{N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}+\sqrt{N^{\tau}\eta_0}\bigg) \\
&\quad = -\frac{1}{8\pi\mathrm{i}}\int_{\mathbb{R}}\frac{g(x)}{x+\mathrm{i}N^{-\tau}}\,\mathrm{d}x + \frac{1}{8\pi\mathrm{i}}\int_{\mathbb{R}}\frac{g(x)}{x-\mathrm{i}N^{-\tau}}\,\mathrm{d}x + O_{\prec}\bigg(N^{-\tau}+\frac{N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}+\sqrt{N^{\tau}\eta_0}\bigg).
\end{aligned}$$

Using the Sokhotski–Plemelj lemma, we have

$$\mathbb{E}\,\mathrm{Tr}\, f(X_N) - N\int_{\mathbb{R}} f(x)\rho_{\mathrm{fc}}(x)\,\mathrm{d}x = \frac{g(0)}{4} + O_{\prec}\bigg(N^{-\tau}+\frac{N^{2\tau}}{\sqrt{N\eta_0\sqrt{\kappa_0+\eta_0}}}+\sqrt{N^{\tau}\eta_0}\bigg),$$

where we used the regularity $g \in C_c^2(\mathbb{R})$. This finishes the computation of mesoscopic bias.

8. Sample covariance matrix

In this section, we use the previous arguments to derive the mesoscopic eigenvalue statistics of sample covariance matrix and prove similar CLTs in the bulk and at the regular edges. We start by introducing the model in detail.

8.1. *Setup, assumptions and main results*

Let $X_N = (X_{ij})$ be an $M \times N$ matrix satisfying the following assumption.

Assumption 8.1.

(1) $\{X_{ij} | 1 \le i \le M, 1 \le j \le N\}$ are independent real-valued centered random variables.
(2) For all i, j, we have $\mathbb{E}|\sqrt{N}X_{ij}|^2 = 1$. In addition, $\sqrt{N}X_{ij}$ has uniformly bounded moments, that is, there exists $C_p > 0$ independent of N such that for all i, j,

$$\mathbb{E}|\sqrt{N}X_{ij}|^p \le C_p. \tag{8.1}$$

(3) To simplify the statement, we also assume that there exists a constant K_4 such that

$$K_4 := \frac{1}{N}\sum_{j=1}^{N} c_{ij}^{(4)}, \quad \text{where } c_{ij}^{(4)} \text{ is the fourth cumulant of } \sqrt{N}X_{ij}. \tag{8.2}$$

Note that M depends on N and set

$$\gamma \equiv \gamma_N := \frac{M}{N} \to \gamma_0, \quad 0 < \gamma_0 < \infty. \tag{8.3}$$

We study the $M \times M$ sample covariance matrices

$$H_M := Y_N Y_N^*, \qquad Y_N := \Sigma^{1/2} X_N, \qquad \Sigma := \mathrm{Diag}(\sigma_i), \tag{8.4}$$

where Σ is an $M \times M$ positive definite, deterministic and diagonal matrix with

$$\infty > \sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_M > 0, \qquad \limsup_{N\to\infty}\sigma_1 < \infty, \qquad \liminf_{N\to\infty}\sigma_M > 0. \tag{8.5}$$

We denote by μ_Σ the empirical eigenvalue distribution of Σ, i.e. $\mu_\Sigma := \frac{1}{M}\sum_{j=1}^{M}\delta_{\sigma_j}$. The following assumption ensures that the limit of μ_Σ exists.

Assumption 8.2. Together with (8.5), μ_Σ converges weakly to a deterministic measure μ_σ as $N \to \infty$ such that μ_σ is compactly supported in $(0, \infty)$.

The eigenvalues of $H \equiv H_M$ are denoted as $\lambda_i \in \mathbb{R}$, $1 \le i \le M$. The empirical spectral measure of H_M is defined by $\mu_M = \frac{1}{M}\sum_{i=1}^{M}\delta_{\lambda_i}$. The Stieltjes transform of μ_M is then given by

$$m_M(z) := M^{-1}\operatorname{Tr} G(z), \quad \text{where} \quad G(z) := (H_M - zI)^{-1}, \quad z \in \mathbb{C}^+. \tag{8.6}$$

We further introduce the $N \times N$ matrices

$$\mathcal{H}_N := Y_N^* Y_N, \qquad \mathcal{G} := (\mathcal{H} - z)^{-1}, \qquad \mathfrak{m}_N(z) := N^{-1}\operatorname{Tr}\mathcal{G}, \quad z \in \mathbb{C}^+. \tag{8.7}$$

The eigenvalues of $\mathcal{H} \equiv \mathcal{H}_N$ are denoted by $\{\mu_i\}_{i=1}^N$. It is straightforward that $\{\lambda_i\}_{i=1}^M$ differs from $\{\mu_i\}_{i=1}^N$ by $|N - M|$ zeros, hence we have the relation

$$\mathfrak{m}_N(z) = \gamma m_M + \frac{\gamma - 1}{z}. \tag{8.8}$$

In the null case $\Sigma = I$, the Marchenko–Pastur law states that the empirical eigenvalue distribution of $H = XX^*$ converges weakly to the Marchenko–Pastur distribution with aspect ratio γ_0, whose density is given by $\mathrm{d}\mu_{\mathrm{MP},\gamma_0} := \frac{1}{2\pi\gamma_0}\sqrt{\frac{[(x-\gamma_-)(\gamma_+-x)]_+}{x^2}}\,\mathrm{d}x + (1-\gamma_0^{-1})_+\delta_0$ with $\gamma_\pm = (1 \pm \sqrt{\gamma_0})^2$. Its Stieltjes transform m_{MP,γ_0}, or denoted by m_{γ_0} for short, is characterized as the unique solution of

$$1 + (z - 1 + \gamma_0)m(z) + \gamma_0 z m^2(z) = 0, \quad \text{or equivalently,} \quad m(z) = \frac{1}{1 - \gamma_0 - \gamma_0 z m(z) - z}, \tag{8.9}$$

such that $\operatorname{Im} m(z) > 0$, $z \in \mathbb{C}^+$. Because of (8.8), the Stieltjes transform of the limiting spectral measure of $\mathcal{H} = X^*X$, denoted by $\mathfrak{m}_{\gamma_0^{-1}}$, is then given by

$$\mathfrak{m}_{\gamma_0^{-1}}(z) = \gamma_0 m_{\gamma_0}(z) + \frac{\gamma_0 - 1}{z}. \tag{8.10}$$

In the non-null case $\Sigma \neq I$, under Assumption 8.2, the limiting spectral measure of $H = \Sigma^{1/2}XX^*\Sigma^{1/2}$ exists, henceforth referred to as the deformed Marchenko–Pastur law. Its Stieltjes transform, denoted by m_{fc,γ_0}, or m_{fc} for short, is the unique solution of

$$m(z) = \int_{\mathbb{R}} \frac{1}{t(1 - \gamma_0 - \gamma_0 z m(z)) - z}\,\mathrm{d}\mu_\sigma(t), \tag{8.11}$$

such that $\operatorname{Im} m(z) > 0$, $z \in \mathbb{C}^+$. The corresponding limiting measure, denoted by $\mu_{\mathrm{fc},\gamma_0}$ or μ_{fc} for short, is the free multiplicative convolution of μ_σ and the standard Marchenko–Pastur law with ratio γ_0, i.e., $\mu_{\mathrm{fc},\gamma_0} = \mu_\sigma \boxtimes \mu_{\mathrm{MP},\gamma_0}$; see [67,69]. It was proved in [64] that the free multiplicative convolution measure is absolutely continuous and its density function is analytic whenever positive in $(0, \infty)$.

According to (8.10), the Stieltjes transform of the limiting spectral measure of $\mathcal{H} = X^*\Sigma X$ is the unique solution to

$$\mathfrak{m}(z) = \frac{1}{-z + \gamma_0 \int_{\mathbb{R}} \frac{t}{t\mathfrak{m}(z)+1}\,\mathrm{d}\mu_\sigma(t)}, \quad \text{or equivalently,} \quad \gamma_0 - 1 - z\mathfrak{m}(z) = \int_{\mathbb{R}} \frac{\gamma_0}{1 + t\mathfrak{m}(z)}\,\mathrm{d}\mu_\sigma(t), \tag{8.12}$$

such that $\operatorname{Im}\mathfrak{m}(z) > 0$, $z \in \mathbb{C}^+$.

For general Σ, μ_{fc} could be supported on several disjoint intervals; we refer to [34,43,64] for discussions on the support of the density. The following assumption ensures that the free multiplicative convolution measure is supported on a single interval and the edges behave like square roots. It also rules out the possibilities of outliers.

Assumption 8.3. Let $[\sigma_-, \sigma_+] \subset \mathbb{R}^+$ be the smallest interval that contains the support of μ_σ and set $\mathcal{I} := [\sigma_+^{-1}, \sigma_-^{-1}]$. Assume that

$$\inf_{x\in\mathcal{I}} \int_{\mathbb{R}} \left(\frac{tx}{1-tx}\right)^2 \mathrm{d}\mu_\sigma(t) \geq \gamma_0^{-1} + w,$$

for some constant $w > 0$ (the left side may be infinite). Similarly, set $\hat{\mathcal{I}} := [\sigma_1^{-1}, \sigma_M^{-1}]$. Assume that

$$\inf_{x\in\hat{\mathcal{I}}} \int_{\mathbb{R}} \left(\frac{tx}{1-tx}\right)^2 \mathrm{d}\mu_\Sigma(t) \geq \gamma^{-1} + w,$$

for sufficiently large N.

Let $\xi = -\mathfrak{m}(z)$ so that (8.12) is equivalent to

$$z = F(\xi), \qquad F(\xi) := \frac{1}{\xi} + \gamma_0 \int_{\mathbb{R}} \frac{t}{1-t\xi}\, \mathrm{d}\mu_\sigma(t).$$

As an analogue of (2.7), it was argued in [64] that the edges of the support of μ_{fc}, denoted as $E_\pm$, are given by $E_\pm = F(\xi_\pm)$, where $\xi_\pm \in \mathbb{R}$ are the solutions to

$$H(\xi) := \int_{\mathbb{R}} \left(\frac{t\xi}{1-t\xi}\right)^2 \mathrm{d}\mu_\sigma(t) = \gamma_0^{-1}. \tag{8.13}$$

Under Assumption 8.3, we have at most two solutions of (8.13), since $H(\xi)$ is monotone outside $[\sigma_+^{-1}, \sigma_-^{-1}]$. Let ξ_+ be the unique solution of (8.13) in $(0, \sigma_+^{-1})$. The right boundary of the spectrum is given by $E_+ = F(\xi_+)$. As for the left edge, we split into three cases. If $0 < \gamma_0 < 1$, there is a unique solution of (8.13) in the interval (σ_-^{-1}, ∞), denoted by ξ_-, and the corresponding left edge is given by $E_- = F(\xi_-)$. If $\gamma_0 > 1$, similarly, there is a unique solution of (8.13) in the interval $(-\infty, 0)$. These edges are referred to as soft edges. For $\gamma_0 = 1$, the solution does not exist (or say $\xi_- = \infty$), corresponding to $E_- = 0$. This scenario is referred to as the hard edge and the density there goes to infinity at rate of $\kappa^{-1/2}$. In this paper, we only consider the right edge for all $0 < \gamma_0 < \infty$. Same discussion easily extends to the soft left edge when $\gamma_0 \neq 1$.

In addition, Assumption 8.3 implies that

$$\mathrm{dist}\big(\{\sigma_+^{-1}, \sigma_-^{-1}\}, \xi_\pm\big) \geq c_0 > 0, \tag{8.14}$$

provided that ξ_- is finite ($\gamma_0 \neq 1$). The above condition is crucial for the density of the limiting measure to have the square root behavior at the soft edges. It will be proved later in Lemma 8.5. Similar assumptions also appeared in [8,24,51] for the rightmost edge and [34,43] for all edges in multi-cuts.

Since the convergence rate of μ_Σ and γ could be very slow, from now on, we work with the finite-N version $\mu_{\mathrm{fc}} = \mu_{\mathrm{MP},\gamma} \boxtimes \mu_\Sigma$. The corresponding Stieltjes transforms are given by (8.11) and (8.12) replacing the limiting measure μ_σ by μ_Σ and γ_0 by γ. The following notations, e.g., m_{fc}, $\mathfrak{m}$, $E_\pm$, $\xi_\pm$ and κ are corresponding to $\mu_{\mathrm{MP},\gamma} \boxtimes \mu_\Sigma$ and are N-dependent. To be consistent with the previous sections, we will use the tilde sign to denote the ones with respect to $\mu_{\mathrm{MP},\gamma_0} \boxtimes \mu_\sigma$. The second condition of Assumption 8.3 ensures the same properties for the support of $\mu_{\mathrm{MP},\gamma_0} \boxtimes \mu_\Sigma$. In particular, for sufficiently large N we have

$$\min_i \big\{|1 - \xi_\pm \sigma_i|\big\} \geq c_0 > 0, \tag{8.15}$$

if $\gamma_0 \neq 1$. If $\gamma_0 = 1$, it only holds true with respect to ξ_+.

Next, we state the local law for the Green function of sample covariance matrix, which is an essential tool in our proof. Let $\mathfrak{m} \equiv \mathfrak{m}(z)$ be the unique solution of finite-N version of (8.12), i.e.,

$$\gamma - 1 - z\mathfrak{m}(z) = \frac{1}{M} \sum_{i=1}^{M} \frac{\gamma}{1 + \sigma_i \mathfrak{m}(z)}, \tag{8.16}$$

such that $\mathrm{Im}\,\mathfrak{m}(z) > 0$, $z \in \mathbb{C}^+$. Define the deterministic control parameters

$$\Psi(z) := \sqrt{\frac{\mathrm{Im}\,\mathfrak{m}(z)}{N|\eta|}} + \frac{1}{N|\eta|}, \qquad \Theta(z) := \frac{1}{N|\eta|}, \qquad z = E + \mathrm{i}\eta \in \mathbb{C} \setminus \mathbb{R}. \tag{8.17}$$

We also introduce the spectral domain, for some small $c > 0$,

$$S' := \{z = E + \mathrm{i}\eta : |E| \le c^{-1}, N^{-1+c} \le \eta \le c^{-1}, |z| \ge c\}. \tag{8.18}$$

We further introduce the $N + M$ by $N + M$ matrices

$$R := \begin{pmatrix} -\Sigma^{-1} & X \\ X^* & -z \end{pmatrix}^{-1}; \qquad \Pi := \begin{pmatrix} -\Sigma(1+\mathfrak{m}\Sigma)^{-1} & 0 \\ 0 & \mathfrak{m}(z) \end{pmatrix}; \qquad \Sigma' := \begin{pmatrix} \Sigma & 0 \\ 0 & I \end{pmatrix}, \quad z \in \mathbb{C}^+.$$

Using the Schur decomposition/Feshbach formula, we see that

$$R = \begin{pmatrix} z\Sigma^{1/2} G \Sigma^{1/2} & \Sigma X \mathcal{G} \\ \mathcal{G} X^* \Sigma & \mathcal{G} \end{pmatrix} = \begin{pmatrix} z\Sigma^{1/2} G \Sigma^{1/2} & \Sigma^{1/2} G Y \\ Y^* G \Sigma^{1/2} & \mathcal{G} \end{pmatrix}.$$

We are ready to state the (anisotropic) local law for such random matrix.

Theorem 8.4 (Theorem 2.4 in [14], Theorem 3.6 in [43]). *For any deterministic unit vector $v, w \in \mathbb{C}^N$, we have*

$$\left|\left\langle v, \Sigma'^{-1}\big(R(z) - \Pi(z)\big)\Sigma'^{-1} w\right\rangle\right| \prec \Psi(z),$$

uniformly in $z \in S'$. It also implies that

$$\left|\big(\mathcal{G}(z)\big)_{ij} - \mathfrak{m}(z)\delta_{ij}\right| \prec \Psi(z); \qquad \left|\big(G(z)\big)_{ij} + \frac{1}{z(1+\mathfrak{m}\sigma_i)}\delta_{ij}\right| \prec \Psi(z).$$

In addition, we have the averaged result

$$\left|N^{-1}\operatorname{Tr}\mathcal{G}(z) - \mathfrak{m}(z)\right| \prec \Theta(z), \qquad \left|M^{-1}\operatorname{Tr} G(z) + \int_{\mathbb{R}} \frac{1}{z(1+\mathfrak{m}t)}\,\mathrm{d}\mu_\Sigma(t)\right| \prec \Theta(z).$$

In the following, we state some properties of the Stieltjes transform $\mathfrak{m}$ in (8.12), whose proofs are given in Appendix B. Define the spectral domain, for some small $c > 0$,

$$S := \{z = E + \mathrm{i}\eta : |E| \le c^{-1}, 0 < \eta \le c^{-1}, |z| \ge c\}.$$

And set $\kappa \equiv \kappa(E) := \min\{|E_+ - E|, |E_- - E|\}$.

Lemma 8.5.

(1) *For $z \in S$ and sufficiently large N, we have*

$$\left|\mathfrak{m}(z)\right| \sim 1; \qquad \min_i \left|1 + \sigma_i \mathfrak{m}(z)\right| > c_0. \tag{8.19}$$

(2) *For $z \in S$ and sufficiently large N, we have*

$$\left|\operatorname{Im}\mathfrak{m}(z)\right| \sim \begin{cases} \sqrt{\kappa+\eta}, & \text{if } E \in [E_-, E_+], \\ \frac{\eta}{\sqrt{\kappa+\eta}}, & \text{if otherwise.} \end{cases} \tag{8.20}$$

(3) *For $z \in S$ with $E_- - c < E < E_+ + c$ and $\eta \le c$ for some small $c > 0$, we have*

$$1 - \frac{1}{M}\sum_{i=1}^{M} \frac{\gamma\sigma_i}{z(1+\mathfrak{m}(z)\sigma_i)^2} = -\frac{\mathfrak{m}(z)}{z\mathfrak{m}'(z)} \sim \sqrt{\kappa+\eta}. \tag{8.21}$$

(4) *Under the same condition as in* (3), *we have*

$$\mathfrak{m}'(z) = -\frac{\mathfrak{m}(z)}{z - \frac{1}{M}\sum_{i=1}^{M} \frac{\gamma\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2}} \sim \frac{1}{\sqrt{\kappa+\eta}}. \tag{8.22}$$

We are now prepared to state our main results for the sample covariance matrix.

Proposition 8.6. *Consider a sample covariance matrix satisfying Assumptions 8.1, 8.2 and 8.3 and E_0 is chosen to be away from zero, then Propositions 2.7 and 2.9 hold true with*

$$K(z_1,z_2)=2\bigg(\frac{\mathfrak{m}_1'\mathfrak{m}_2'}{(\mathfrak{m}_1-\mathfrak{m}_2)^2}-\frac{1}{(z_1-z_2)^2}\bigg)+K_4\gamma\frac{\partial^2}{\partial z_1\partial z_2}\bigg(\frac{1}{M}\sum_{i=1}^{M}\frac{1}{(1+\mathfrak{m}_1\sigma_i)(1+\mathfrak{m}_2\sigma_i)}\bigg), \tag{8.23}$$

where we use $\mathfrak{m}_1$ and $\mathfrak{m}_2$ to denote $\mathfrak{m}(z_1)$ and $\mathfrak{m}(z_2)$, and

$$b(z)=\bigg(\frac{(\mathfrak{m}'(z))^2}{\mathfrak{m}(z)}+K_4\mathfrak{m}(z)\mathfrak{m}'(z)\bigg)\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i^2}{(1+\mathfrak{m}(z)\sigma_i)^3}. \tag{8.24}$$

Proposition 8.6 implies that Theorems 2.10 and 2.11 hold true for sample covariance matrix. More specifically, we have the following theorem.

Theorem 8.7. *Let H_N be a sample covariance matrix of the form (8.4) satisfying Assumptions 8.1–8.3. Let $N^{-1+c_1}\le\eta_0\le N^{-c_1}$ with some $c_1>0$ and fix $E_0\in(E_-,E_+)$, such that $\kappa_0:=\operatorname{dist}(\operatorname{supp}(f_N),\{E_\pm\})>c_0$, for some $c_0>0$ and sufficiently large N. Then, for any function $g\in C_c^2(\mathbb{R})$, the linear eigenvalue statistics (2.21) converges in distribution to the Gaussian random variable $\mathcal{N}(0,\frac{1}{\pi}\int_{\mathbb{R}}|\xi||\hat g(\xi)|^2\,\mathrm{d}\xi)$, where $\hat g(\xi):=(2\pi)^{-1/2}\int_{\mathbb{R}}g(x)e^{-\mathrm{i}\xi x}\,\mathrm{d}x$.*

In addition, the linear statistics (2.21) with $E_0=E_+$ and $N^{-\frac{2}{3}+c_2}\le\eta_0\le N^{-c_2}$ for some $c_2>0$, converges in distribution to a Gaussian random variable $\mathcal{N}(\frac{g(0)}{4},\frac{1}{2\pi}\int_{\mathbb{R}}|\xi||\hat h(\xi)|^2\,\mathrm{d}\xi)$, where $h(x)=g(-x^2)$ and $\hat h(\xi):=(2\pi)^{-1/2}\int_{\mathbb{R}}h(x)e^{-\mathrm{i}\xi x}\,\mathrm{d}x$. Furthermore, if $\gamma_0\ne1$, a similar CLT for $E_0=E_-$ can be obtained with $h(x)=g(x^2)$.

Remark. We remark that (8.2) in Assumption 8.1 can be removed. In addition, we can relax the single support condition for μ_{fc} by assuming instead that the cuts of the support of μ_{fc} are separated by order one and the density has square root behaviors at the edges away from zero.

8.2. *Proof of the CLT and variance computation*

From the definition of the Green function and (8.4), we get

$$zG_{ii}=(HG)_{ii}-1=\big(\Sigma^{1/2}XX^*\Sigma^{1/2}G\big)_{ii}-1=\sqrt{\sigma_i}\sum_{j=1}^{N}X_{ij}(GY)_{ij}-1. \tag{8.25}$$

Similarly as (5.6), by the cumulant expansion formula, we have

$$z\mathbb{E}\big[e_0(\lambda)(G_{ii}-\mathbb{E}G_{ii})\big]=I_1+I_2+I_3+O_{\prec}\big(N^{-\frac{3}{2}}\big(1+|\lambda|^4\big)\big), \tag{8.26}$$

where

$$I_1:=\frac{\sqrt{\sigma_i}}{N}\sum_{j=1}^{N}c_{ij}^{(2)}\bigg(\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial X_{ij}}(GY)_{ij}\bigg]+\mathbb{E}\bigg[\bigg(\frac{\partial(GY)_{ij}}{\partial X_{ij}}-\mathbb{E}\bigg[\frac{\partial(GY)_{ij}}{\partial X_{ij}}\bigg]\bigg)e_0(\lambda)\bigg]\bigg),$$

$$I_2:=\frac{\sqrt{\sigma_i}}{2!N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\bigg(\mathbb{E}\bigg[\frac{\partial^2e_0(\lambda)}{\partial^2X_{ij}}(GY)_{ij}\bigg]+2\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial X_{ij}}\frac{\partial(GY)_{ij}}{\partial X_{ij}}\bigg]+\mathbb{E}\bigg[\bigg(\frac{\partial^2(GY)_{ij}}{\partial^2X_{ij}}-\mathbb{E}\bigg[\frac{\partial^2(GY)_{ij}}{\partial^2X_{ij}}\bigg]\bigg)e_0(\lambda)\bigg]\bigg),$$

$$\begin{aligned}I_3:=\frac{\sqrt{\sigma_i}}{3!N^2}\sum_{j=1}^{N}c_{ij}^{(4)}\bigg(&\mathbb{E}\bigg[\frac{\partial^3e_0(\lambda)}{\partial^3X_{ij}}(GY)_{ij}\bigg]+3\mathbb{E}\bigg[\frac{\partial^2e_0(\lambda)}{\partial^2X_{ij}}\frac{\partial(GY)_{ij}}{\partial X_{ij}}\bigg]+3\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial X_{ij}}\frac{\partial^2(GY)_{ij}}{\partial^2X_{ij}}\bigg]\\&+\mathbb{E}\bigg[(1-\mathbb{E})\bigg(\frac{\partial^3(GY)_{ij}}{\partial^3X_{ij}}\bigg)e_0(\lambda)\bigg]\bigg).\end{aligned}$$

The last term on the right side of (8.26) is estimated by (8.32), (8.1) and Lemma 1.2. The argument is similar as in (5.6). The only things to check are the deterministic bounds of $(Y^*GY)_{ii}$ and $(GY)_{ij}$. Note that from (8.4) and (8.7), $YG=\mathcal{G}Y$

and $|G_{ij}| = O(N^c)$, for $z \in \Omega_0 \cap S'$, thus we have

$$\big(Y^*GY\big)_{ii} = \big(Y^*Y\mathcal{G}\big)_{ii} = (\mathcal{H}\mathcal{G})_{ii} = (1 + z\mathcal{G})_{ii} = O\big(N^{c_1}\big); \tag{8.27}$$

$$\begin{aligned}\big|(GY)_{ij}\big| &\le \sqrt{N}\big(GYY^*G^*\big)_{ii}^{1/2} = \sqrt{N}\big(z\big(GG^*\big)_{ii} + G^*_{ii}\big)^{1/2} = \sqrt{N}\bigg(\frac{z}{2\operatorname{Im} z}\big(G_{ii} - G^*_{ii}\big) + G^*_{ii}\bigg)^{1/2} \\ &= O\big(N^{c_2}\big),\end{aligned} \tag{8.28}$$

where we use Cauchy–Schwarz inequality and the resolvent identity (5.13).

Using the formulas,

$$\frac{\partial G_{ab}}{\partial Y_{jk}} = -G_{aj}\big(Y^*G\big)_{kb} - (GY)_{ak}G_{jb}, \qquad \frac{\partial G_{ab}}{\partial X_{jk}} = \frac{\partial G_{ab}}{\partial Y_{jk}}\sqrt{\sigma_j}, \tag{8.29}$$

we obtain the analogue of Lemma 5.1:

Lemma 8.8. *For any i, j, we have*

$$\frac{\partial e_0(\lambda)}{\partial X_{ij}} = -\frac{\mathrm{i}2\sqrt{\sigma_i}\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\frac{\mathrm{d}}{\mathrm{d}z}(GY)_{ij}\,\mathrm{d}^2z; \tag{8.30}$$

$$\frac{\partial^2 e_0(\lambda)}{\partial^2 X_{ij}} = -\frac{\mathrm{i}2\sigma_i\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\frac{\mathrm{d}}{\mathrm{d}z}\bigg(\frac{\mathfrak{m}}{1+\mathfrak{m}\sigma_i}\bigg)\mathrm{d}^2z + O_{\prec}\bigg(\frac{(1+|\lambda|)^2}{\sqrt{N\eta_0}}\bigg). \tag{8.31}$$

In general, for any integer $k \in \mathbb{N}$, we have

$$\bigg|\frac{\partial^k(GY)_{ij}}{\partial X_{ij}^k}\bigg| \prec O(1); \qquad \bigg|\frac{\partial^k e_0(\lambda)}{\partial^k X_{ij}}\bigg| \prec O\big(\big(1+|\lambda|\big)^k\big). \tag{8.32}$$

We first look at I_1. Using (8.29) and (8.30), we have

$$\begin{aligned}I_1 &= \frac{\sigma_i}{N}\sum_{j=1}^N\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})(G_{ii})\big] - \frac{\sigma_i}{N}\sum_{j=1}^N\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\big(\big(Y^*GY\big)_{jj}G_{ii}\big)\big] \\ &\quad - \frac{\sigma_i}{N}\sum_{j=1}^N\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\big((GY)_{ji}(GY)_{ji}\big)\big] + \frac{\sqrt{\sigma_i}}{N}\sum_{j=1}^N\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial X_{ji}}(GY)_{ji}\bigg] =: A_1 + A_2 + A_3 + A_4.\end{aligned}$$

Note that

$$A_1 = \sigma_i\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})G_{ii}\big] = \sigma_i\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})G_{ii}\big].$$

In addition, using the definition of the resolvent and the local law in Theorem 8.4, we have

$$\begin{aligned}A_2 &= -\frac{\sigma_i}{N}\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\big(\operatorname{Tr}(GH)G_{ii}\big)\big] = -\gamma\sigma_i\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\big(zM^{-1}\operatorname{Tr}GG_{ii}\big)\big] - \gamma\sigma_i\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})(G_{ii})\big] \\ &= (-\gamma\sigma_i z m_{\mathrm{fc}} - \gamma\sigma_i)\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})(G_{ii})\big] + \frac{\gamma\sigma_i}{1+\mathfrak{m}\sigma_i}\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\big(M^{-1}\operatorname{Tr}G\big)\big] + O_{\prec}\big(\Psi(z)\Theta(z)\big).\end{aligned}$$

Next, we use the local law to estimate the third term,

$$A_3 = -\frac{1}{N}\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\big(Y^*G^2Y\big)_{ii}\big] = -\frac{1}{N}\mathbb{E}\bigg[e_0(\lambda)(1-\mathbb{E})\frac{\mathrm{d}}{\mathrm{d}z}\big(Y^*GY\big)_{ii}\bigg] = O_{\prec}\bigg(\frac{\Psi(z)}{N\eta}\bigg).$$

Finally, we study the last term A_4 using (8.30), which can be written as

$$\begin{aligned}A_4 &= -\frac{\mathrm{i}2\sigma_i\lambda}{\pi N}\mathbb{E}\bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)\frac{\partial}{\partial z_2}\big(G(z_2)HG(z_1)\big)_{ii}\,\mathrm{d}^2z_2\bigg] \\ &= -\frac{\mathrm{i}2\sigma_i\lambda}{\pi N}\mathbb{E}\bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)\frac{\partial}{\partial z_2}\big(z_2G(z_2)G(z_1)\big)_{ii}\,\mathrm{d}^2z_2\bigg].\end{aligned}$$

Using the resolvent identity (5.13), the local law Theorem 8.4 and Lemma 3.2, we obtain that

$$A_4 = -\frac{\mathrm{i}2\lambda}{\pi N}\mathbb{E}\bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)\frac{\partial}{\partial z_2}\bigg(\frac{\sigma_i z_2(\mathfrak{g}_i(z_1)-\mathfrak{g}_i(z_2))}{z_1-z_2}\bigg)\mathrm{d}^2 z_2\bigg] + e(i), \tag{8.33}$$

where $\mathfrak{g}_i(z) := -\frac{1}{z(1+\mathfrak{m}\sigma_i)}$ for simplicity, and $e(i)$ is the error term. If we consider the linear statistics of the error term $e(i)$, using the same argument as for deformed Wigner matrix in Section 5.4, we have

$$\bigg|\sum_{i=1}^{N}\mathfrak{g}_i(z)e_i(z)\bigg| = O_{\prec}\bigg(\frac{1}{N\eta_1^2}\bigg) + O_{\prec}\bigg(\frac{1}{N\eta_0\eta_1}\bigg).$$

As for the second cumulant expansion term I_2, we apply the same argument as in (5.8) and the anisotropic law Theorem 8.4 to find that

$$I_2 = O_{\prec}\bigg(\frac{(1+|\lambda|^2)\Psi(z)}{N\sqrt{\eta_0}}\bigg) + O_{\prec}\bigg(\frac{\Psi^2(z)}{\sqrt{N}}\bigg).$$

We compute the third cumulant expansion term I_3 similarly using (8.29), the local law Theorem 8.4 and (8.32). The leading term comes from the second term of I_3, denoted by D_2, i.e.,

$$I_3 = \frac{\sigma_i}{2N^2}\sum_{j=1}^{N}\mathbb{E}\bigg[\frac{\partial^2 e_0(\lambda)}{\partial^2 X_{ij}}c_{ij}^{(4)}\big(G_{ii}-\big(Y^*GY\big)_{jj}G_{ii}-(GY)_{ij}(GY)_{ij}\big)\bigg] + O_{\prec}\big(\big(1+|\lambda|^3\big)N^{-1}\Psi(z)\big).$$

Using (8.27) and (8.31), D_2 can be written as

$$\begin{aligned} D_2 &= \frac{\sigma_i}{2N^2}\sum_{j=1}^{N}\mathbb{E}\bigg[\frac{\partial^2 e_0(\lambda)}{\partial^2 X_{ij}}c_{ij}^{(4)}\frac{\mathfrak{m}}{1+\mathfrak{m}\sigma_i}\bigg] + O_{\prec}\big(\big(1+|\lambda|^2\big)N^{-1}\Psi(z)\big) \\ &= -\frac{\mathrm{i}K_4\sigma_i^2\lambda}{\pi N}\mathbb{E}\bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z'}}\tilde{f}\big(z'\big)\frac{\partial}{\partial z'}\bigg(\frac{\mathfrak{m}(z)}{1+\mathfrak{m}(z)\sigma_i}\frac{\mathfrak{m}(z')}{1+\mathfrak{m}(z')\sigma_i}\bigg)\mathrm{d}^2 z'\bigg] \\ &\quad + O_{\prec}\bigg(\frac{1+|\lambda|^2}{N\sqrt{N\eta_0}}\bigg) + O_{\prec}\big(\big(1+|\lambda|^2\big)N^{-1}\Psi(z)\big), \end{aligned} \tag{8.34}$$

where K_4 is given in (8.2) and is independent of the index i.

Summing up and rearranging terms in (8.26), the coefficient in front of $\mathbb{E}[e_0(\lambda)(1-\mathbb{E})G_{ii}]$ is given by

$$(z-\sigma_i+\gamma\sigma_i+\gamma\sigma_i z m_{\mathrm{fc}}) = z\big(1+\mathfrak{m}(z)\sigma_i\big),$$

which is away from zero for $z\in S'$. Dividing both sides of (8.26) by this coefficient, we have

$$\mathbb{E}\big[e_0(\lambda)(G_{ii}-\mathbb{E}G_{ii})\big] = \frac{\gamma\sigma_i}{Mz(1+\mathfrak{m}(z)\sigma_i)^2}\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})(\operatorname{Tr}G)\big] + \frac{A_4(i)}{z(1+\mathfrak{m}(z)\sigma_i)} + \frac{D_2(i)}{z(1+\mathfrak{m}(z)\sigma_i)} + \mathcal{E}(i),$$

where A_4 and D_2 are given in (8.33) and (8.34), and the error terms $\mathcal{E}(i)$ is analytic in Ω_0 and estimated as before. Summing over i and rearranging, we have

$$\bigg(1-\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i}{z(1+\mathfrak{m}(z)\sigma_i)^2}\bigg)\mathbb{E}\big[e_0(\lambda)(1-\mathbb{E})\operatorname{Tr}G\big] = \sum_{i=1}^{M}\frac{A_4(i)}{z(1+\mathfrak{m}(z)\sigma_i)} + \sum_{i=1}^{M}\frac{D_2(i)}{z(1+\mathfrak{m}(z)\sigma_i)} + \mathcal{E}, \tag{8.35}$$

where the error term $\mathcal{E}$ has an upper bound as in (5.14). Dividing by the coefficient $1-\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i}{z(1+\mathfrak{m}(z)\sigma_i)^2}$ and recalling (8.21), the first two terms on the right side of (8.35) are denoted as A and D respectively, and the error term is bounded by (3.15). Using (8.33) and (8.21), we have

$$\begin{aligned} A &= -\frac{z_1\mathfrak{m}_1'}{\mathfrak{m}_1}\frac{\mathrm{i}2\lambda\gamma}{\pi}\mathbb{E}\bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)\frac{\partial}{\partial z_2}\bigg(\frac{z_2}{z_1-z_2}\frac{1}{M}\sum_{i=1}^{M}\sigma_i\mathfrak{g}_i(z_1)\big(\mathfrak{g}_i(z_1)-\mathfrak{g}_i(z_2)\big)\bigg)\mathrm{d}^2 z_2\bigg] \\ &= \frac{\mathrm{i}2\lambda}{\pi}\mathbb{E}\bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)\bigg(\frac{\mathfrak{m}_1'\mathfrak{m}_2'}{(\mathfrak{m}_1-\mathfrak{m}_2)^2}-\frac{1}{(z_1-z_2)^2}\bigg)\mathrm{d}^2 z_2\bigg], \end{aligned}$$

where we set $\mathfrak{m}_1 := \mathfrak{m}(z_1)$ and $\mathfrak{m}_2 := \mathfrak{m}(z_2)$ for short. This follows from

$$z_1 z_2 \frac{1}{M}\sum_{i=1}^{M} \sigma_i \mathfrak{g}_i(z_1)\mathfrak{g}_i(z_2) = \frac{1}{M}\sum_{i=1}^{N} \frac{\sigma_i}{(1+\mathfrak{m}_1\sigma_i)(1+\mathfrak{m}_2\sigma_i)} = \frac{z_1\mathfrak{m}_1 - z_2\mathfrak{m}_2}{\gamma(\mathfrak{m}_1-\mathfrak{m}_2)};$$

$$z_1^2 \frac{1}{M}\sum_{i=1}^{M} \sigma_i \mathfrak{g}_i^2(z_1) = \frac{1}{M}\sum_{i=1}^{N} \frac{\sigma_i}{(1+\mathfrak{m}_1\sigma_i)^2} = \frac{(z_1\mathfrak{m}_1)'}{\gamma\mathfrak{m}_1'}.$$

Similarly, recalling (5.11), we obtain that

$$D = -\frac{z_1\mathfrak{m}_1'}{\mathfrak{m}_1}\sum_{i=1}^{N}\frac{D_2(i)}{z(1+\mathfrak{m}(z)\sigma_i)} = \frac{\mathrm{i}K_4\gamma\lambda}{\pi}\mathbb{E}\Bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)\frac{1}{M}\sum_{i=1}^{M}\Bigg(\sigma_i^2\frac{\mathfrak{m}_1'}{(1+\mathfrak{m}_1\sigma_i)^2}\frac{\partial}{\partial z_2}\frac{\mathfrak{m}_2}{1+\mathfrak{m}_2\sigma_i}\Bigg)\mathrm{d}^2 z_2\Bigg]$$

$$= \frac{\mathrm{i}K_4\gamma\lambda}{\pi}\mathbb{E}\Bigg[e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z_2}}\tilde{f}(z_2)\Bigg(\frac{\partial^2}{\partial z_1\partial z_2}\Bigg[\frac{1}{M}\sum_{i=1}^{M}\frac{1}{(1+\mathfrak{m}_1\sigma_i)(1+\mathfrak{m}_2\sigma_i)}\Bigg]\Bigg)\mathrm{d}^2 z_2\Bigg].$$

Therefore, we obtain an analogue of Lemma 3.4, where the kernel is given instead by (8.23).

Next, we compute the explicit formula for the variance $V(f)$ given by (2.15) with K in (8.23) and test function f in (2.11). Both in the bulk and at the edge, the second term of (8.23) will only contribute $O(\eta_0 N^{\tau})$, because of (8.22), (8.19) and (3.1). It is sufficient to look at the first term. In the bulk, the main contribution of $V(f)$ comes from the term $-\frac{2}{(z_1-z_2)^2}$ with z_1, z_2 in different half planes. Hence by similar arguments as in Lemma 6.1, we obtain that

$$\lim_{N\to\infty} V(f) = \frac{1}{2\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(g(x_1)-g(x_2))^2}{(x_1-x_2)^2}\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

At the edge, we recall the expansion of $\mathfrak{m}(z)$ near $z = E_+$ from (B.3),

$$\mathfrak{m}(z) = \mathfrak{m}(E_+) + c_+\sqrt{z-E_+}\big(1+A_+(\sqrt{z-E_+})\big) = \xi_+ + c_+\sqrt{z-E_+} + O\big(|z-E_+|\big),$$

where the square root is taken in a branch cut such that $\operatorname{Im}\mathfrak{m} > 0$ when $\operatorname{Im} z > 0$, and $c_+ > c_0 > 0$ for sufficiently large N. Differentiating it, we have

$$\mathfrak{m}'(z) = \frac{c_+}{2\sqrt{z-E_+}} + d_+ + O\big(\sqrt{|z-E_+|}\big).$$

Therefore, repeating the arguments in the proof of Lemma 6.2, we get the variance at the edge

$$\lim_{N\to\infty} V(f) = \frac{1}{4\pi^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\Bigg(\frac{g(-x_1^2)-g(-x_2^2)}{x_1-x_2}\Bigg)^2\mathrm{d}x_1\,\mathrm{d}x_2.$$

8.3. *Expectation and bias computation*

Starting from (8.25), using the cumulant expansion and (8.29), we obtain that

$$z\mathbb{E}G_{ii} = \sqrt{\sigma_i}\sum_{j=1}^{N}\mathbb{E}X_{ij}(GY)_{ij} - 1 = \frac{\sqrt{\sigma_i}}{N}\mathbb{E}\sum_{j=1}^{N}c_{ij}^{(2)}\frac{\partial(GY)_{ij}}{\partial X_{ij}} - 1$$

$$+ \frac{\sqrt{\sigma_i}}{2!N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\mathbb{E}\frac{\partial^2(GY)_{ij}}{\partial^2 X_{ij}} + \frac{\sqrt{\sigma_i}}{3!N^2}\sum_{j=1}^{N}c_{ij}^{(4)}\mathbb{E}\frac{\partial^3(GY)_{ij}}{\partial^3 X_{ij}} + O_{\prec}\big(N^{-\frac{3}{2}}\big)$$

$$= \frac{\sigma_i}{N}\sum_{j=1}^{N}\mathbb{E}G_{ii} - \frac{\sigma_i}{N}\sum_{j=1}^{N}\mathbb{E}\big(Y^*GY\big)_{jj}G_{ii} - \frac{\sigma_i}{N}\sum_{j=1}^{N}\mathbb{E}\big((GY)_{ij}\big)^2 - 1 \tag{8.36}$$

$$+\frac{\sigma_i^{3/2}}{2N^{\frac{3}{2}}}\sum_{j=1}^{N}c_{ij}^{(3)}\mathbb{E}\big(-6G_{ii}(GY)_{ij}+6G_{ii}(GY)_{ij}\big(Y^*GY\big)_{jj}+2\big((GY)_{ij}\big)^3\big) \tag{8.37}$$

$$+\frac{\sigma_i^2}{6N^2}\sum_{j=1}^{N}c_{ij}^{(4)}\mathbb{E}\big(-6(G_{ii})^2+12(G_{ii})^2\big(Y^*GY\big)_{jj}-6(G_{ii})^2\big(\big(Y^*GY\big)_{jj}\big)^2\big)+O_{\prec}\big(N^{-\frac{3}{2}}\big). \tag{8.38}$$

The first line (8.36) can be written as

$$\begin{aligned}
&\sigma_i\mathbb{E}G_{ii}-\frac{\sigma_i}{N}\mathbb{E}\big[\operatorname{Tr}\big(Y^*GY\big)G_{ii}\big]-\frac{\sigma_i}{N}\mathbb{E}(GHG)_{ii}-1\\
&\quad=\sigma_i(1-\gamma)\mathbb{E}G_{ii}-z\sigma_i\gamma\mathbb{E}\big[M^{-1}\operatorname{Tr}GG_{ii}\big]+\frac{\sigma_i}{N}\bigg(\frac{1}{1+\sigma_i\mathfrak{m}}\bigg)'-1+O_{\prec}\big(N^{-3/2}\big)\\
&\quad=\sigma_i(1-\gamma)\mathbb{E}G_{ii}-z\sigma_i\gamma m_{\mathrm{fc}}\mathbb{E}\bigg(G_{ii}+\frac{1}{z(1+\mathfrak{m}\sigma_i)}\bigg)+\frac{\sigma_i\gamma}{M(1+\mathfrak{m}\sigma_i)}\mathbb{E}\operatorname{Tr}G-\frac{\sigma_i^2}{N}\frac{\mathfrak{m}'}{(1+\sigma_i\mathfrak{m})^2}\\
&\qquad-1+O_{\prec}\big((N\eta)^{-\frac{3}{2}}\big).
\end{aligned}$$

Using the anisotropic local law and the same arguments as in (5.8), one shows that the second line (8.37) is $O_{\prec}(N^{-1}\Psi(z))$. The local law implies that the last line (8.38) becomes

$$\begin{aligned}
&\frac{\sigma_i^2}{N^2}\sum_{j=1}^{N}c_{ij}^{(4)}\mathbb{E}\big(-(G_{ii})^2+2(G_{ii})^2\big(Y^*GY\big)_{jj}-(G_{ii})^2\big(\big(Y^*GY\big)_{jj}\big)^2\big)+O_{\prec}\big(N^{-\frac{3}{2}}\big)\\
&\quad=-\frac{\sigma_i^2}{N^2}\sum_{j=1}^{N}c_{ij}^{(4)}\bigg(\frac{\mathfrak{m}^2}{(1+\mathfrak{m}\sigma_i)^2}\bigg)+O_{\prec}\big(N^{-\frac{3}{2}}\big)=-K_4\frac{\sigma_i^2}{N}\bigg(\frac{\mathfrak{m}^2}{(1+\mathfrak{m}\sigma_i)^2}\bigg)+O_{\prec}\big(N^{-\frac{3}{2}}\big).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
z(1+\mathfrak{m}\sigma_i)\bigg(\mathbb{E}G_{ii}+\frac{1}{z(1+\mathfrak{m}\sigma_i)}\bigg)&=\frac{\sigma_i\gamma}{1+\mathfrak{m}\sigma_i}\mathbb{E}\big[M^{-1}\operatorname{Tr}G-m_{\mathrm{fc}}\big]-\frac{1}{M}\frac{\gamma\sigma_i^2\mathfrak{m}'}{(1+\sigma_i\mathfrak{m})^2}\\
&\quad-K_4\frac{1}{M}\frac{\gamma\sigma_i^2\mathfrak{m}^2}{(1+\mathfrak{m}\sigma_i)^2}+O_{\prec}\big((N\eta)^{-\frac{3}{2}}\big).
\end{aligned}$$

Dividing both sides by $z(1+\mathfrak{m}\sigma_i)\sim O(1)$ from (8.19) and summing over i, we obtain

$$\begin{aligned}
&\bigg(1-\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i}{z(1+\mathfrak{m}(z)\sigma_i)^2}\bigg)\mathbb{E}(\operatorname{Tr}G-Mm_{\mathrm{fc}})\\
&\quad=-\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i^2\mathfrak{m}'}{z(1+\mathfrak{m}\sigma_i)^3}-K_4\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i^2\mathfrak{m}^2}{z(1+\mathfrak{m}\sigma_i)^3}+O_{\prec}\bigg(\frac{1}{\sqrt{N\eta^3}}\bigg).
\end{aligned}$$

Dividing both sides by the coefficient of $\mathbb{E}(\operatorname{Tr}G-Mm_{\mathrm{fc}})$ and using (8.21), the error becomes $O_{\prec}(\frac{1}{\eta\sqrt{N\eta\sqrt{\kappa+\eta}}})$ and the leading term on the right side becomes (8.24). Therefore, we obtained an analogue of Proposition 2.9 with the integral kernel $b(z)$ given instead by (8.24).

Finally, we compute the explicit formula for the bias. Using (8.22), (8.19) and (3.1), the second term of $b(z)$ will contribute $O(\sqrt{\eta_0 N^{\tau}})$ both in the bulk and at the edge. In addition, (8.16) implies that the first term of $b(z)$ can be written as

$$\frac{(\mathfrak{m}')^2}{\mathfrak{m}}\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i^2}{(1+\sigma_i\mathfrak{m})^3}=\frac{\mathfrak{m}''}{2\mathfrak{m}'}-\frac{\mathfrak{m}'}{\mathfrak{m}}.$$

The second term on the right side contributes $O(\sqrt{\eta_0 N^\tau})$. The first term vanishes if $\kappa_0 > c > 0$ and thus the bias in the bulk vanishes. At the edge, we use the expansion of $\mathfrak{m}(z)$ around E_+ from (B.3),

$$\mathfrak{m}'(z) = \frac{c_+}{2\sqrt{z-E_+}} + d_+ + O\big(\sqrt{|z-E_+|}\big), \qquad \mathfrak{m}''(z) = -\frac{c_+}{4(\sqrt{z-E_+})^3} + O\bigg(\frac{1}{\sqrt{|z-E_+|}}\bigg).$$

Hence we have

$$\frac{\mathfrak{m}''}{2\mathfrak{m}'} = -\frac{1}{4(z-E_+)} + O\bigg(\frac{1}{\sqrt{|z-E_+|}}\bigg),$$

and using the Sokhotski–Plemelj lemma, the bias at the edge becomes $\frac{g(0)}{4}$.

Appendix A: Complex case

In this appendix, we extend previous results from real symmetric to complex Hermitian matrices. We will use the complex analogue of Lemma 3.3.

Lemma A.1 (Complex cumulant expansion). *Let h be a complex-valued random variable with finite moments, and f is a complex-valued smooth function on $\mathbb{R}$ with bounded derivatives. Let $c_{p,q}$ be the (p,q) cumulant of h, which is defined as*

$$c_{p,q} := (-\mathrm{i})^{p+q}\bigg(\frac{\partial^{p+q}}{\partial s^p \partial t^q}\log \mathbb{E}e^{\mathrm{i}sh+\mathrm{i}t\bar{h}}\bigg)\bigg|_{s,t=0}.$$

Then for any fixed $l \in \mathbb{N}$, we have

$$\mathbb{E}\big[\bar{h} f(h,\bar{h})\big] = \sum_{p+q=0}^{l} \frac{1}{p!q!} c_{p,q+1}(h)\mathbb{E}\big[f^{(p,q)}(h)\big] + R_{l+1},$$

where the error term satisfies

$$|R_{l+1}| \le C_l \mathbb{E}|h|^{l+2} \max_{p+q=l+1}\Big\{\sup_{|x|\le M}\big|f^{(p,q)}(z,\bar{z})\big|\Big\} + C_l \mathbb{E}\big[|h|^{l+2}1_{|h|>M}\big]\max_{p+q=l+1}\big\|f^{(p,q)}(z,\bar{z})\big\|_\infty,$$

and $M > 0$ is an arbitrary fixed cutoff.

Instead of (5.5), we have

$$\frac{\partial G_{ij}}{\partial H_{ab}} = -G_{ia}G_{bj}, \tag{A.1}$$

from which we obtain the analogue of Lemma 5.1.

The assumption $\mathbb{E}H_{ij}^2 = 0$ implies that $c_{ij}^{(1,1)} = 1$, $c_{ij}^{(2,2)} = W_4 - 2$ for $i \ne j$. Using the anisotropic law and (A.1), one shows similarly that the expansion terms corresponding to $p+q=3$ are negligible. Using (A.1) and the analogue of Lemma 5.1, we obtain that

$$\begin{aligned}
&(z-a_i)\mathbb{E}\big[e_0(\lambda)(G_{ii}-\mathbb{E}G_{ii})\big] \\
&\quad= \frac{1}{N}\sum_{j=1}^{N} c_{ij}^{(1,1)}\mathbb{E}\bigg[\frac{\partial}{\partial H_{ji}}\big(e_0(\lambda)(G_{ji}-\mathbb{E}G_{ji})\big)\bigg] \\
&\qquad+ \frac{1}{2!N^2}\sum_{j=1}^{N} c_{ij}^{(2,2)}\mathbb{E}\bigg[\frac{\partial^3}{\partial^2 H_{ji}\partial H_{ij}}\big(e_0(\lambda)(G_{ji}-\mathbb{E}G_{ji})\big)\bigg] + \cdots
\end{aligned}$$

$$= \frac{1}{N}\mathbb{E}\bigg[e_0(\lambda)\Big(\frac{\partial G_{ji}}{\partial H_{ji}} - \mathbb{E}\frac{\partial G_{ji}}{\partial H_{ji}}\Big)\bigg] + \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial H_{ji}}G_{ji}\bigg] + \frac{m_2-1}{N}\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial H_{ii}}G_{ii}\bigg]$$

$$+ \frac{1}{N^2}\sum_{j=1}^{N}(W_4-2)\mathbb{E}\bigg[\frac{\partial e_0(\lambda)}{\partial H_{ji}\partial H_{ij}}\frac{\partial G_{ji}}{\partial H_{ji}}\bigg] + \cdots.$$

Thus Proposition 2.7 holds with modified variance, i.e. m_2-2 be replaced by m_2-1, W_4-3 be replaced by W_4-2, and the coefficient of the remaining term be 1 instead of 2. Similarly, as for the expectation,

$$(z-a_i)\mathbb{E}G_{ii} = \mathbb{E}(HG)_{ii} - 1 = \frac{1}{N}\sum_{j=1}^{N}c_{ij}^{(1,1)}\mathbb{E}\frac{\partial G_{ji}}{\partial H_{ji}} - 1 + \frac{1}{2!N^2}\sum_{j=1}^{N}c_{ij}^{(2,2)}\mathbb{E}\frac{\partial^3 G_{ji}}{\partial^2 H_{ji}\partial H_{ij}} + \cdots.$$

Thus the first term of $b(z)$ given in (2.19) vanishes, m_2-2 is replaced by m_2-1 and W_4-3 is replaced by W_4-2.

Appendix B: Proofs of auxiliary lemmas

Proof of Lemma 3.2. For $1 \le s \le 2$, the proof is given in Lemma 4.4 in [44]. Since $h(z)$ is holomorphic on Ω_0, $\frac{\partial}{\partial\overline{z}}\tilde{f}(z)h(z) = \frac{\partial}{\partial\overline{z}}(\tilde{f}(z)h(z))$. Using Stokes' formula, we have

$$\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)h(z)\,\mathrm{d}^2z = -\frac{\mathrm{i}}{2}\int_{\partial\Omega_0}\tilde{f}(z)h(z)\,\mathrm{d}z.$$

Since g is compactly support, $\tilde{f}(z)=0$ on $\partial\Omega_0$ except

$$\Gamma_0 := \big\{x+\mathrm{i}y : x\in\operatorname{supp}(f), |y| = N^{-\tau}\eta_0\big\}.$$

Using (3.1) we have

$$\bigg|\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)h(z)\,\mathrm{d}^2z\bigg| \le CK\int_{\Gamma_0}\big(|y|^{-s}\big|f(x)\big| + |y|^{1-s}\big|f'(x)\big|\big)\,\mathrm{d}z \le C'KN^{\tau s}\eta_0^{1-s}.$$

□

Proof of Lemma 4.3. Using the self-consistent equation of m_{fc} in (2.6), we have

$$m_{\mathrm{fc}}(z_1) - m_{\mathrm{fc}}(z_2) = \frac{1}{N}\sum_{i=1}^{N}\bigg(\frac{1}{a_i - z_1 - m_{\mathrm{fc}}(z_1)} - \frac{1}{a_i - z_2 - m_{\mathrm{fc}}(z_2)}\bigg)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\bigg(\frac{z_1 + m_{\mathrm{fc}}(z_1) - z_2 - m_{\mathrm{fc}}(z_2)}{(a_i - z_1 - m_{\mathrm{fc}}(z_1))(a_i - z_2 - m_{\mathrm{fc}}(z_2))}\bigg).$$

Dividing $z_1 + m_{\mathrm{fc}}(z_1) - z_2 - m_{\mathrm{fc}}(z_2)$ on both sides and we get the first identity. Taking the derivative of (2.6), we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1+m'_{\mathrm{fc}}(z)}{(a_i - z - m_{\mathrm{fc}})^2} = m'_{\mathrm{fc}}(z). \tag{B.1}$$

We treat $\tilde{I}$ and $\tilde{I}_s$ similarly. Thus we complete the proof. □

Proof of Lemma 4.4. Note that

$$\big|I_s(z)\big| \le \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|a_i - z - m_{\mathrm{fc}}(z)|^2} = \frac{\operatorname{Im} m_{\mathrm{fc}}(z)}{\operatorname{Im} m_{\mathrm{fc}}(z) + \eta} < 1.$$

By (B.1), we have $\frac{m'_{\mathrm{fc}}}{1+m'_{\mathrm{fc}}} = I_s(z)$ and thus $m'_{\mathrm{fc}}(z) = \frac{I_s(z)}{1-I_s(z)}$. Using Lemma 4.1, we have

$$\big|m'_{\mathrm{fc}}(z)\big| \le \frac{1}{|1-I_s(z)|} \sim \frac{1}{\sqrt{\kappa+\eta}}. \tag{B.2}$$

Differentiating (B.1) again, we obtain that

$$\frac{m_{\mathrm{fc}}''}{(1+m_{\mathrm{fc}}')^3} = \frac{2}{N}\sum_{i=1}^{N}\frac{1}{(a_i - z - m_{\mathrm{fc}})^3}.$$

Combining (4.3) and (B.2), we get the upper bound of m_{fc}''. The rest inequalities follow directly from Lemma 4.1. □

Proof of Lemma 5.1. Using (5.5), we have

$$\frac{\partial e_0(\lambda)}{\partial H_{ij}} = \frac{\mathrm{i}\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\left(\sum_{l=1}^{N}\frac{\partial G_{ll}}{\partial H_{ij}}\right)\mathrm{d}^2z = -\frac{\mathrm{i}(2-\delta_{ij})\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\big(G^2\big)_{ji}\,\mathrm{d}^2z.$$

Note that $(G^2)_{ji} = \frac{d}{dz}G_{ji}(z)$. Since G_{ij} is analytic in D', using the Cauchy integral formula and the local law, we have that for $i \neq j$, $(G^2)_{ji} \prec \frac{\Psi(z)}{\operatorname{Im} z}$. Combining with Lemma 3.2, we obtain that, for $i \neq j$,

$$\left|\frac{\partial e_0(\lambda)}{\partial H_{ij}}\right| = O_{\prec}\left(\frac{1+|\lambda|}{\sqrt{N\eta_0}}\right).$$

Similarly, if $i = j$, we have

$$\frac{\partial e_0(\lambda)}{\partial H_{ii}} = -\frac{\mathrm{i}\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\frac{\partial}{\partial z}\frac{1}{a_i - z - m_{\mathrm{fc}}(z)}\,\mathrm{d}^2z + O_{\prec}\left(\frac{1+|\lambda|}{\sqrt{N\eta_0}}\right).$$

Furthermore, we compute that

$$\begin{aligned}\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}} &= -\frac{\lambda^2(2-\delta_{ij})^2}{\pi^2}e_0(\lambda)\left(\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\big(G^2\big)_{ji}\,\mathrm{d}^2z\right)^2\\ &\quad + \frac{\mathrm{i}(2-\delta_{ij})\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\big(2\big(G^2\big)_{ji}G_{ij} + (1-\delta_{ij})\big(G^2\big)_{ii}G_{jj} + (1-\delta_{ij})\big(G^2\big)_{jj}G_{ii}\big)\,\mathrm{d}^2z.\end{aligned}$$

For $i \neq j$, combining the local law and Lemma 3.2, we have

$$\begin{aligned}\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}} &= \frac{\mathrm{i}2\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\frac{\mathrm{d}}{\mathrm{d}z}(G_{ji}G_{ij} + G_{ii}G_{jj})\,\mathrm{d}^2z + O_{\prec}\left(\frac{(1+|\lambda|)^2}{N\eta_0}\right)\\ &= \frac{\mathrm{i}2\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\frac{\partial}{\partial z}\frac{1}{(a_i - z - m_{\mathrm{fc}})(a_j - z - m_{\mathrm{fc}})}\,\mathrm{d}^2z + O_{\prec}\left(\frac{(1+|\lambda|)^2}{\sqrt{N\eta_0}}\right).\end{aligned}$$

Similarly, for $i = j$, we have

$$\frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ii}} = \frac{\mathrm{i}\lambda}{\pi}e_0(\lambda)\int_{\Omega_0}\frac{\partial}{\partial\overline{z}}\tilde{f}(z)\frac{\partial}{\partial z}\frac{1}{(a_i - z - m_{\mathrm{fc}})^2}\,\mathrm{d}^2z + O_{\prec}\left(\frac{(1+|\lambda|)^2}{\sqrt{N\eta_0}}\right).$$

In general, using the local law, (5.5) and Lemma 3.2, we complete the proof of (5.4). □

Proof of Lemma 8.5. We start by proving the first two statements, using which we will show (8.21). The last statement then follows directly from (8.21). Note that the first equation in (8.12) implies the first inequality in (8.19). To prove the rest, we divide the spectral domain S into three regimes, corresponding to the bulk, the edge and the outside. First, we consider z near the edge E_+, i.e., $z \in S^e := \{z = E + \mathrm{i}\eta \in S : E \in [E_+ - \tau', E_+ + \tau']\}$ for some small $\tau' > 0$. Let $\xi = -\mathfrak{m}(z)$ so that (8.16) is equivalent to

$$z = F(\xi), \qquad F(\xi) := \frac{1}{\xi} + \gamma\int_{\mathbb{R}}\frac{t}{1-t\xi}\,\mathrm{d}\mu_{\Sigma}(t).$$

Due to (8.15), $F(\xi)$ is analytic around ξ_+, and we have

$$F(\xi) = F(\xi_+) + F'(\xi_+)(\xi - \xi_+) + \frac{1}{2!}F''(\xi_+)(\xi - \xi_+)^2 + O\big(|\xi - \xi_+|^3\big),$$

where the linear term vanishes because of (8.13). It also yields

$$F''(\xi_+) = \frac{2}{\xi_+^3} + 2\gamma \int \frac{t^3}{(1-t\xi_+)^3}\,\mathrm{d}\mu_\Sigma(t) = 2\gamma \int \frac{t^2}{\xi_+(1-t\xi_+)^3}\,\mathrm{d}\mu_\Sigma(t) \geq c > 0.$$

The last step follows from the fact that $\xi_+ \geq c > 0$ and $1 - t\xi_+ \geq c > 0$ for sufficiently large N. Thus we have the expansion of $\mathfrak{m}$ near the edge $z = E_+$,

$$\mathfrak{m}(z) = \mathfrak{m}(E_+) + c_+\sqrt{z-E_+}\big(1 + A_+(\sqrt{z-E_+})\big) = \xi_+ + c_+\sqrt{z-E_+} + O\big(|z-E_+|\big), \tag{B.3}$$

where $c_+ > c_0 > 0$ for sufficiently large N, A_+ is an analytic function with $A_+(0) = 0$, and the square root is taken with the branch cut such that $\operatorname{Im}\mathfrak{m}(z) > 0$ when $\operatorname{Im} z > 0$. Hence the corresponding density has the square root behavior at the right edge. The left edge can be treated similarly when $\gamma_0 \neq 1$. Using the definition of Stieltjes transform, one shows (8.20). Similar arguments can be found in Lemma A.5 in [48].

If E is inside the bulk, i.e., $z \in S^b := \{z \in S : E \in [E_- + \tau', E_+ - \tau']\}$, then $\operatorname{Im}\mathfrak{m} \geq c > 0$. Thus (8.19) and (8.20) follows. Finally, for the outside spectral domain, if $z \in S^o := \{z \in S : \operatorname{dist}(E, [E_-, E_+]) \geq \tau'\}$, it follows from $\operatorname{Im}\mathfrak{m} \sim \eta$ and (8.15). Similar arguments can be found in Appendix A in [43].

Next, we will prove (8.21). We first prove the upper bound. Taking the real and imaginary part of (8.16), we have

$$E\operatorname{Im}\mathfrak{m} + \eta\operatorname{Re}\mathfrak{m} = \frac{\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i\operatorname{Im}\mathfrak{m}}{|1+\mathfrak{m}\sigma_i|^2}; \qquad E\operatorname{Re}\mathfrak{m} - \eta\operatorname{Im}\mathfrak{m} = -\frac{\gamma}{M}\sum_{i=1}^{M}\frac{1+\sigma_i\operatorname{Re}\mathfrak{m}}{|1+\mathfrak{m}\sigma_i|^2} + \gamma - 1, \tag{B.4}$$

with $\mathfrak{m} \equiv \mathfrak{m}(z)$. Then we have

$$\left|z - \frac{\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i}{|1+\mathfrak{m}(z)\sigma_i|^2}\right| = \left|\mathrm{i}\eta - \frac{\operatorname{Re}\mathfrak{m}}{\operatorname{Im}\mathfrak{m}}\eta\right| = \frac{|\mathfrak{m}|\eta}{\operatorname{Im}\mathfrak{m}}.$$

Using $|\mathfrak{m}(z)| \sim 1$ and (8.20). we obtain an upper bound of the right side as $C\sqrt{\kappa+\eta}$. In addition,

$$\begin{aligned}&\left|\frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_i}{|1+\mathfrak{m}(z)\sigma_i|^2} - \frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2}\right| \\ &\quad = 2\left|\frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_i(\operatorname{Im}(1+\mathfrak{m}(z)\sigma_i))^2 + \mathrm{i}\sigma_i\operatorname{Re}(1+\mathfrak{m}(z)\sigma_i)\operatorname{Im}(1+\mathfrak{m}(z)\sigma_i)}{|1+\mathfrak{m}(z)\sigma_i|^4}\right| \leq C\sqrt{\kappa+\eta},\end{aligned}$$

hence we obtain an upper bound for the left side of (8.21). Next, it is sufficient to show the lower bound. If $z \in S^e$, (8.15) implies that $\operatorname{Re}(1+\sigma_i\mathfrak{m}) > c$ if we choose τ' sufficiently small. We split into two cases. If $E \in [E_-, E_+]$, we have

$$\begin{aligned}\left|z - \frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2}\right| &\geq \left|\operatorname{Im}\left(z - \frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2}\right)\right| \geq \left|\eta + \frac{2\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i^2\operatorname{Im}\mathfrak{m}\operatorname{Re}(1+\sigma_i\mathfrak{m})}{|(1+\mathfrak{m}(z)\sigma_i)|^4}\right| \\ &\geq \frac{2\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i\operatorname{Im}\mathfrak{m}\operatorname{Re}(1+\sigma_i\mathfrak{m})}{|(1+\mathfrak{m}(z)\sigma_i)|^4} \geq C\sqrt{\kappa+\eta}.\end{aligned} \tag{B.5}$$

Otherwise, $E \in [E_-, E_+]^c$, we have

$$\begin{aligned}\left|z - \frac{\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2}\right| &\geq \left||z| - \frac{\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i}{|1+\mathfrak{m}(z)\sigma_i|^2}\right| = \left|\sqrt{E^2+\eta^2} - E - \frac{\operatorname{Re}\mathfrak{m}}{\operatorname{Im}\mathfrak{m}}\eta\right| \\ &\geq \frac{|\operatorname{Re}\mathfrak{m}|}{\operatorname{Im}\mathfrak{m}}\eta - \eta \geq C\sqrt{\kappa+\eta} - \eta \geq C'\sqrt{\kappa+\eta},\end{aligned}$$

if we choose τ' sufficiently small. The last second step follows from the fact that near the edge, $|\operatorname{Re}\mathfrak{m}| \geq C > 0$, because of the second equation of (B.4). Next, if $z \in S^b$, then we have $\operatorname{Im}\mathfrak{m} > c$. We also split into two cases. If $\operatorname{Re}\mathfrak{m} > 0$, we

repeat (B.5) to get (8.21). If $\operatorname{Re}\mathfrak{m} \le 0$, from (B.4) we have

$$\frac{\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i}{|1+\mathfrak{m}\sigma_i|^2} = E + \eta\frac{\operatorname{Re}\mathfrak{m}}{\operatorname{Im}\mathfrak{m}} \le E.$$

In addition, we have

$$\operatorname{Re}\frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2} = \frac{\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i[\operatorname{Re}(1+\mathfrak{m}(z)\sigma_i)^2 - \operatorname{Im}(1+\mathfrak{m}(z)\sigma_i)]^2]}{|1+\mathfrak{m}(z)\sigma_i|^4} \le E - C.$$

Therefore, we have

$$\left|z - \frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2}\right| \ge E - \operatorname{Re}\left(\frac{\gamma}{M}\sum_{i=1}^{M}\frac{\sigma_i}{(1+\mathfrak{m}(z)\sigma_i)^2}\right) \ge C \ge C\sqrt{\kappa+\eta}.$$

Finally, taking the derivative of (8.16), we obtain that

$$(z\mathfrak{m})' = \mathfrak{m} + z\mathfrak{m}' = \frac{1}{M}\sum_{i=1}^{M}\frac{\gamma\sigma_i\mathfrak{m}'}{(1+\sigma_i\mathfrak{m})^2}.$$

Hence, we finish the proof of (8.21), which directly implies (8.22). □

References

[1] A. Adhikari and J. Huang Dyson Brownian motion for general β and potential at the edge. Preprint, 2018. Available at arXiv:1810.08308.

[2] N. I. Akhiezer. *The Classical Moment Problem: And Some Related Questions in Analysis*. Hafner Publishing Co., New York, 1965. MR0184042

[3] J. Alt, L. Erdős, T. Krüger and D. Schröder. Correlated random matrices: Band rigidity and edge universality. *Ann. Probab.* **48** (2) (2020) 963–1001. MR4089499 https://doi.org/10.1214/19-AOP1379

[4] Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** (1A) (2004) 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

[5] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. *Mathematics Monograph Series* **2**. Science Press, Beijing, 2006. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

[6] Z. D. Bai, X. Wang and W. Zhou. Functional CLT for sample covariance matrices. *Bernoulli* **16** (4) (2010) 1086–1113. MR2759170 https://doi.org/10.3150/10-BEJ250

[7] Z. D. Bai and J. F. Yao. On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** (6) (2005) 1059–1092. MR2189081 https://doi.org/10.3150/bj/1137421640

[8] Z. Bao, G. Pan and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** (1) (2015) 382–421. MR3311864 https://doi.org/10.1214/14-AOS1281

[9] E. L. Basor and H. Widom. Determinants of airy operators and applications to random matrices. *J. Stat. Phys.* **96** (1999) 1–20. MR1706781 https://doi.org/10.1023/A:1004539513619

[10] F. Bekerman and A. Lodhia. Mesoscopic central limit theorem for general β-ensembles. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** (2018) 1917–1938. MR3865662 https://doi.org/10.1214/17-AIHP860

[11] L. Benigni Eigenvectors distribution and quantum unique ergodicity for deformed Wigner matrices. Preprint, 2018. Available at arXiv:1711.07103.

[12] P. Biane. On the free convolution with a semi-circular distribution. *Indiana Univ. Math. J.* **46** (1997) 705–718. MR1488333 https://doi.org/10.1512/iumj.1997.46.1467

[13] P. Biane. Processes with free increments. *Math. Z.* **227** (1) (1998) 143–174. MR1605393 https://doi.org/10.1007/PL00004363

[14] A. Bloemendal, L. Erdős, A. Knowles, H. T. Yau and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** (2014) 33. MR3183577 https://doi.org/10.1214/ejp.v19-3054

[15] P. Bourgade Extreme gaps between eigenvalues of Wigner matrices. Preprint, 2018. Available at arXiv:1812.10376.

[16] P. Bourgade, L. Erdős, H.-T. Yau and J. Yin. Fixed energy universality for generalized Wigner matrices. *Comm. Pure Appl. Math.* **69** (10) (2016) 1815–1881. MR3541852 https://doi.org/10.1002/cpa.21624

[17] P. Bourgade and K. Mody. Gaussian fluctuations of the determinant of Wigner matrices. *Electron. J. Probab.* **24** (2019) 96. MR4017114 https://doi.org/10.1214/19-ejp356

[18] A. Boutet de Monvel and A. Khorunzhy. Asymptotic distribution of smoothed eigenvalue density. I. Gaussian random matrices. *Random Oper. Stoch. Equ.* **7** (1) (1999) 1–22. MR1678012 https://doi.org/10.1515/rose.1999.7.1.1

[19] A. Boutet de Monvel and A. Khorunzhy. Asymptotic distribution of smoothed eigenvalue density. II. Wigner random matrices. *Random Oper. Stoch. Equ.* **7** (2) (1999) 149–168. MR1689027 https://doi.org/10.1515/rose.1999.7.2.149

[20] J. Breuer and M. Duits. Universality of mesoscopic fluctuations for orthogonal polynomial ensembles. *Comm. Math. Phys.* **342** (2) (2016) 491–531. MR3459158 https://doi.org/10.1007/s00220-015-2514-6

[21] G. Cipolloni, L. Erdős, T. Krüger and D. Schröder. Cusp universality for random matrices II: The real symmetric case. *Pure Appl. Anal.* **1** (4) (2019) 615–707. MR4026551 https://doi.org/10.2140/paa.2019.1.615

[22] S. Dallaporta and M. Fevrier Fluctuations of linear spectral statistics of deformed Wigner matrices. Preprint, 2019. Available at arXiv:1903.11324. MR4119597 https://doi.org/10.1142/S2010326320500112

[23] M. Duits and K. Johansson. On mesoscopic equilibrium for linear statistics in Dyson's Brownian motion. *Mem. Amer. Math. Soc.* **255** (2018) 1222. MR3852256 https://doi.org/10.1090/memo/1222

[24] N. El Karoui. Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** (2) (2007) 663–714. MR2308592 https://doi.org/10.1214/009117906000000917

[25] L. Erdős and A. Knowles. The Altshuler–Shklovskii formulas for random band matrices I: The unimodular case. *Comm. Math. Phys.* **333** (3) (2015) 1365–1416. MR3302637 https://doi.org/10.1007/s00220-014-2119-5

[26] L. Erdős and A. Knowles. The Altshuler–Shklovskii formulas for random band matrices II: The general case. *Ann. Henri Poincaré* **16** (3) (2015) 709–799. MR3311888 https://doi.org/10.1007/s00023-014-0333-5

[27] L. Erdős, A. Knowles and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** (8) (2013) 1837–1926. MR3119922 https://doi.org/10.1007/s00023-013-0235-y

[28] L. Erdős, T. Krüger and D. Schröder Cusp universality for random matrices I: Local law and the complex Hermitian case. Preprint, 2018. Available at arXiv:1809.03971. MR4134946 https://doi.org/10.1007/s00220-019-03657-4

[29] L. Erdős, T. Krüger and D. Schröder. Random matrices with slow correlation decay. *Forum Math. Sigma* **7** (2019) 8. MR3941370 https://doi.org/10.1017/fms.2019.2

[30] L. Erdős, B. Schlein and H.-T. Yau. Universality of random matrices and local relaxation flow. *Invent. Math.* **185** (1) (2011) 75–119. MR2810797 https://doi.org/10.1007/s00222-010-0302-7

[31] L. Erdős and K. Schnelli. Universality for random matrix flows with time-dependent density. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** (4) (2017) 1606–1656. MR3729630 https://doi.org/10.1214/16-AIHP765

[32] L. Erdős and H.-T. Yau. Universality of local spectral statistics of random matrices. *Bull. Amer. Math. Soc.* **49** (3) (2012) 377–414. MR2917064 https://doi.org/10.1090/S0273-0979-2012-01372-1

[33] L. Erdős and H.-T. Yau. *A Dynamical Approach to Random Matrix Theory*. *Courant Lecture Notes* **28**. American Mathematical Soc., Providence, 2017. MR3699468

[34] W. Hachem, A. Hardy and J. Najim. Large complex correlated Wishart matrices: Fluctuations and asymptotic independence at the edges. *Ann. Probab.* **44** (3) (2016) 2264–2348. MR3502605 https://doi.org/10.1214/15-AOP1022

[35] Y. He Bulk eigenvalue fluctuations of sparse random matrices. Preprint, 2019. Available at arXiv:1904.07140.

[36] Y. He and A. Knowles. Mesoscopic eigenvalue statistics of Wigner matrices. *Ann. Appl. Probab.* **27** (3) (2017) 1510–1550. MR3678478 https://doi.org/10.1214/16-AAP1237

[37] Y. He and A. Knowles. Mesoscopic eigenvalue density correlations of Wigner matrices. *Probab. Theory Related Fields* **177** (2020) 147–216. MR4095015 https://doi.org/10.1007/s00440-019-00946-w

[38] J. Huang and B. Landon. Rigidity and a mesoscopic central limit theorem for Dyson Brownian motion for general beta and potentials. *Probab. Theory Related Fields* **175** (1–2) (2019) 209–253. MR4009708 https://doi.org/10.1007/s00440-018-0889-y

[39] H. C. Ji and J. O. Lee. Gaussian fluctuations for linear spectral statistics of deformed Wigner matrices. *Random Matrices Theory Appl.* **9** (2020) 2050011. MR4119597 https://doi.org/10.1142/S2010326320500112

[40] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** (1) (1998) 151–204. MR1487983 https://doi.org/10.1215/S0012-7094-98-09108-6

[41] K. Johansson. From Gumbel to Tracy–Widom. *Probab. Theory Related Fields* **138** (1–2) (2007) 75–112. MR2288065 https://doi.org/10.1007/s00440-006-0012-7

[42] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** (1) (1982) 1–38. MR0650926 https://doi.org/10.1016/0047-259X(82)90080-X

[43] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** (1–2) (2017) 257–352. MR3704770 https://doi.org/10.1007/s00440-016-0730-4

[44] B. Landon and P. Sosoe Applications of mesoscopic CLTs in random matrix theory. Preprint, 2018. Available at arXiv:1811.05915.

[45] B. Landon, P. Sosoe and H.-T. Yau. Fixed energy universality of Dyson Brownian motion. *Adv. Math.* **346** (2019) 1137–1332. MR3914908 https://doi.org/10.1016/j.aim.2019.02.010

[46] B. Landon and H.-T. Yau. Convergence of local statistics of Dyson Brownian motion. *Comm. Math. Phys.* **355** (3) (2017) 949–1000. MR3687212 https://doi.org/10.1007/s00220-017-2955-1

[47] B. Landon and H.-T. Yau Edge statistics of Dyson Brownian motion. Preprint, 2017. Available at arXiv:1712.03881. MR3687212 https://doi.org/10.1007/s00220-017-2955-1

[48] J. O. Lee and K. Schnelli. Local deformed semicircle law and complete delocalization for Wigner matrices with random potential. *J. Math. Phys.* **54** (10) (2013) 103504. MR3134604 https://doi.org/10.1063/1.4823718

[49] J. O. Lee and K. Schnelli. Edge universality for deformed Wigner matrices. *Rev. Math. Phys.* **27** (8) (2015) 1550018. MR3405746 https://doi.org/10.1142/S0129055X1550018X

[50] J. O. Lee and K. Schnelli. Extremal eigenvalues and eigenvectors of deformed Wigner matrices. *Probab. Theory Related Fields* **164** (1) (2016) 165–241. MR3449389 https://doi.org/10.1007/s00440-014-0610-8

[51] J. O. Lee and K. Schnelli. Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** (6) (2016) 3786–3839. MR3582818 https://doi.org/10.1214/16-AAP1193

[52] J. O. Lee, K. Schnelli, B. Stetler and H.-T. Yau. Bulk universality for deformed Wigner matrices. *Ann. Probab.* **44** (3) (2016) 2349–2425. MR3502606 https://doi.org/10.1214/15-AOP1023

[53] A. Lodhia and N. J. Simm Mesoscopic linear statistics of Wigner matrices. Preprint, 2015. Available at arXiv:1503.03533. MR3678478 https://doi.org/10.1214/16-AAP1237

[54] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** (5) (2009) 1778–1840. MR2561434 https://doi.org/10.1214/09-AOP452

[55] V. A. Marchenko and L. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** (4) (1967) 457–483.

[56] C. Min and Y. C. Linear. Statistics of random matrix ensembles at the spectrum edge associated with the airy kernel. *Nuclear Phys. B* **950** (2020) 114836. MR4038239 https://doi.org/10.1016/j.nuclphysb.2019.114836
[57] J. A. Mingo and R. Speicher. Second order freeness and fluctuations of random matrices: I. Gaussian and Wishart matrices and cyclic Fock spaces. *J. Funct. Anal.* **235** (1) (2006) 226–270. MR2216446 https://doi.org/10.1016/j.jfa.2005.10.007
[58] J. Najim and J. Yao. Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Ann. Appl. Probab.* **26** (3) (2016) 1837–1887. MR3513608 https://doi.org/10.1214/15-AAP1135
[59] S. O'Rourke and V. Vu. Universality of local eigenvalue statistics in random matrices with external source. *Random Matrices Theory Appl.* **3** (2) (2014) 1450005. MR3208886 https://doi.org/10.1142/S2010326314500051
[60] L. Pastur. The spectrum of random matrices. *Theoret. and Math. Phys.* **10** (1972) 64–74. MR0475502
[61] M. Shcherbina. Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. *Zh. Mat. Fiz. Anal. Geom.* **7** (2) (2011) 176–192. MR2829615
[62] T. Shcherbina. On universality of local bulk regime for the deformed Gaussian unitary ensemble. *Zh. Mat. Fiz. Anal. Geom.* **5** (4) (2009) 396–433. MR2590774
[63] T. Shcherbina. On universality of local edge regime for the deformed Gaussian unitary ensemble. *J. Stat. Phys.* **143** (2011) 455–481. MR2799948 https://doi.org/10.1007/s10955-011-0196-9
[64] J. W. Silverstein and S. I. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54** (2) (1995) 295–309. MR1345541 https://doi.org/10.1006/jmva.1995.1058
[65] P. Sosoe and P. Wong. Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices. *Adv. Math.* **249** (20) (2013) 37–87. MR3116567 https://doi.org/10.1016/j.aim.2013.09.004
[66] D. Voiculescu. Addition of certain non-commuting random variables. *J. Funct. Anal.* **66** (3) (1986) 323–346. MR0839105 https://doi.org/10.1016/0022-1236(86)90062-5
[67] D. Voiculescu. Multiplication of certain non-commuting random variables. *J. Operator Theory* **18** (2) (1987) 223–235. MR0915507
[68] D. Voiculescu. The analogues of entropy and of Fisher's information theory in free probability theory, I. *Comm. Math. Phys.* **155** (1993) 71–92. MR1228526
[69] D. Voiculescu, K. J. Dykema and A. Nica. *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups*. American Mathematical Society, Providence, 1992. MR1217253
[70] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.* **62** (3) (1955) 548–564. MR0077805 https://doi.org/10.2307/1970079