Central limit theorem for mesoscopic eigenvalue statistics of deformed Wigner matrix

Yiting Li  
KTH Royal Institute of Technology  
yitingl@kth.se

Kevin Schnelli‡  
KTH Royal Institute of Technology  
schnelli@kth.se

Yuanyuan Xu†  
KTH Royal Institute of Technology  
yuax@kth.se

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 the characteristic function method in [33], local laws for the Green function of $X_N$ in [35, 32, 3] and analytic subordination properties of the free additive convolution [16, 29].

Date: September 30, 2019

1. Introduction

1.1. Linear eigenvalue statistics of Wigner matrix. 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 [49] proved the semicircle law stating that the empirical eigenvalue distribution of $H_N$ converges to the semicircle distribution with density

$$\rho_{sc}(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{[-2,2]}.$$ 

That is, for any test function $f \in C_c(\mathbb{R})$,

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

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

Johansson [30] obtained the corresponding Central Limit Theorem (CLT) for such linear eigenvalue statistics of the GUE, i.e.

$$\sum_{i=1}^N f(\lambda_i) - N \int_{\mathbb{R}} f(x) \rho_{sc}(x) dx$$

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 [4] used a martingale method to prove such CLTs for Wigner matrices and analytic test functions. Lytova and Pastur [39], and Shcherbina [43] improved these results by weakening the regularity conditions on the test functions. More recently, Sosoe and Wong [46] obtained CLTs for Wigner matrices with $H^{1+\epsilon}$ test functions using Littlewood–Paley decompositions.

†Supported by the Göran Gustafsson Foundation and the Swedish Research Council Grant VR-2017-05195.
‡Supported in parts by the Swedish Research Council Grant VR-2017-05195.
Boutet de Monvel and Khorunzhy initiated the study of mesoscopic linear eigenvalue statistics, i.e. the derivation of Gaussian fluctuations for the random variable
\[
\sum_{i=1}^N f\left( \frac{\lambda_i - E_0}{\eta_0} \right) - \mathbb{E} \left[ \sum_{i=1}^N f\left( \frac{\lambda_i - E_0}{\eta_0} \right) \right],
\]
with \( E_0 \in (-2, 2) \) on mesoscopic scales \( N^{-1} \ll \eta_0 \ll 1 \). In [12, 13], they obtained CLTs for the test function \((x - i)^{-1}\) on all mesoscopic scales for the GOE, and \( N^{-\frac{2}{3}} \ll \eta_0 \ll 1 \) for symmetric Wigner matrices, respectively. Lodhia and Simm [38] extended the CLT for arbitrary Wigner matrices and general test functions on scales \( N^{-1/3} \ll \eta_0 \ll 1 \). He and Knowles [27] used moment estimates for Green functions to prove the CLT for real symmetric and complex Hermitian Wigner matrices on the optimal scales \( N^{-1} \ll \eta_0 \ll 1 \). More recently, Landon and Sosoe [33] obtained similar CLT by means of the characteristic function.

Mesoscopic central limit theorems are important tools in the theory of the hologramization of Dyson Brownian motion introduced by Bourgade, Erdős, Yau and Yin [10] to prove fixed energy universality of local eigenvalue statistics of Wigner matrices. Landon, Sosoe and Yau [34] subsequently derived a mesoscopic CLT to show fixed energy universality of the Dyson Brownian motion. Mesoscopic central limit theorems were used by Landon and Sosoe [33] and by Bourgade and Mody [11] to derive Gaussian fluctuations of single eigenvalues, and in [11, 9] 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 \( N^{-\frac{3}{2}} \ll \eta_0 \ll 1 \). For the GUE, Basor and Widom [5] used asymptotics of the Airy kernel to prove mesoscopic CLTs at the edges. Min and Chen [40] 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 Dyson Brownian motion.

### 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
\[
X_N = H_N + A_N,
\]
where \( H_N \) is a real symmetric or complex Hermitian Wigner matrix and \( A_N \) is a real deterministic diagonal matrix. Suppose the empirical eigenvalue distribution of \( A_N \) has a deterministic limiting measure, denoted by \( \mu_a \). It was shown by Pastur [42] that the empirical eigenvalue distribution of \( X_N \) converges weakly in probability to the free additive convolution of \( \mu_{sc} \) and \( \mu_a \), denoted by \( \mu_{fc} = \mu_{sc} \boxplus \mu_a \); see also [47].

A CLT for the linear eigenvalue statistics with test functions in \( C^2_c(\mathbb{R}) \) was obtained by Ji and Lee [29] under a one-cut assumption on \( \mu_{fc} \). They also computed the expectation and variance in terms of \( \mu_a \). Dallaporta and Fevrier [16] obtained the CLT for general \( \mu_{fc} \). Their results are summarized in Theorem 2.7 below.

In the present paper, we study the fluctuations of the linear eigenvalue statistics (1.2) in the mesoscopic regime. We assume that the free convolution measure \( \mu_{fc} \) 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 \( \eta_0 \) with \( N^{-1} \ll \eta_0 \sqrt{\eta_0 + \kappa_0} \leq 1 \); see Theorem 2.9. This range of \( \eta_0 \) 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 mesoscopic linear eigenvalue fluctuations.

We follow the idea of [33] to compute the characteristic function of (1.2) in combination with the Helffer-Sjöstrand formula. We also rely on local laws for Green functions [3, 32, 35] and analytic subordination for the free convolution measure, as used in [16, 29, 35].

On the global scale, the derivation of the linear eigenvalue statistics is insensitive [16] to the precise behavior of the free convolution measure \( \mu_{fc} \). An interesting aspect of the free additive convolution measure and deformed Wigner matrices is that the densities may show other edge behaviors than square roots. For such setups, one expects mesoscopic CLTs 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. [31, 36] for some results. At cusp
points in the interior of the bulk spectrum the universality of the local eigenvalue fluctuations was recently proved in [24, 17].

1.3. Related models. Deformed Wigner matrices are closely related to Dyson Brownian motion, for which mesoscopic CLTs were obtained inside the bulk [18, 34, 28] and at the regular edges [1]. The mesoscopic linear statistics were also studied for random band matrices [19, 20], sparse Wigner matrices [25], mesoscopic eigenvalue density correlations for Wigner matrices [26], invariant β-ensembles [6] and orthogonal polynomial ensembles [14]. The global fluctuations of the deformed GOE/GUE can also be studied using the framework of second order freeness [41].

1.4. 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 Section 6 and 7, we compute the results. The proof the main theorem is carried out in Section 3-5. In Section 6 and 7, we compute the

2. Model and main results

2.1. Model and assumptions. Let $H_N \equiv H$ 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. $\{H_{ij} | i \leq j\}$ are independent real-valued centered random variables with $H_{ij} = H_{ji}$.
2. For $i \neq j$, $\mathbb{E}[(\sqrt{N}H_{ij})^2] = 1$; $\mathbb{E}[(\sqrt{N}H_{ii})^2] = m_2$ 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
   \[
   P\left(\{|\sqrt{N}H_{ij}| \geq x\}\right) \leq C_0 e^{-x^\theta}, \quad \forall i, j.
   \] (2.1)

   In particular, we have
   \[
   \mathbb{E}[|\sqrt{N}H_{ij}|^p] \leq C(\theta p)^{\theta p} (p \geq 3).
   \] (2.2)

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

1. $\{\text{Re}H_{ij}, \text{Im}H_{ij} | i \leq j\}$ are independent centered real-valued random variables with $H_{ij} = \overline{H_{ji}}$.
2. For $i \neq j$, $\mathbb{E}[H_{ii}^2] = 0$ 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 $\{A_N\} = \text{Diag}(a_i)$ be a sequence of real deterministic diagonal $N \times N$ matrices with $\|A\|_{op}$ 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 $\nu$ on $\mathbb{R}$ denote by $m_\nu$ its Stieltjes transform, i.e.

\[
\mu_\nu(z) := \int_{\mathbb{R}} \frac{d\nu(x)}{x - z}, \quad z \in \mathbb{C}^+.
\] (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 $\nu$. Moreover, $m_\nu$ satisfies $\lim_{\eta \to \infty} \text{Im} m_\nu(i\eta) = -1$. Conversely, if $m : \mathbb{C}^+ \to \mathbb{C}^+$ is an analytic function with $\lim_{\eta \to \infty} \text{Im} m(i\eta) = -1$, then $m$ is the Stieltjes transform of a probability measure $\nu$, 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 $\mu_A$.

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

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

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

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

The empirical spectral distribution $\mu_N$ converges as $N$ tends to infinity to the free additive convolution of $\mu_a$ and the standard semicircle law. The free convolution measure can be described by analytic subordination [8, 48]: Its Stieltjes transform, $\tilde{m}_{fc}$, is the unique solution to the Pastur equation

$$\tilde{m}_{fc}(z) = \int_{\mathbb{R}} \frac{1}{a - z - \tilde{m}_{fc}(z)} d\mu_a(a), \quad \text{(2.5)}$$

subject to the constraint $\text{Im} \tilde{m}_{fc}(z) > 0, z \in \mathbb{C}^+$. Since the convergence speed in (2.4) can be very slow, we work with a finite $N$ version of the free convolution measure. Let $\mu_{fc}$ denote the free additive convolution of the standard semicircle law and the spectral distribution $\mu_A$. The Stieltjes transform of $\mu_{fc}$, denoted by $m_{fc}$, is hence the unique solution to

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

such that $\text{Im} m_{fc}(z) > 0, z \in \mathbb{C}^+$. Note that $\mu_{fc}$ depends on $N$, but is a deterministic probability measure.

Biane [7] proved that $\tilde{\mu}_{fc}$ and $\mu_{fc}$ are absolutely continuous probability measures whose densities are analytic wherever positive. We denote the density functions by $\tilde{\rho}_{fc}$ and $\rho_{fc}$. In general the measures $\rho_{fc}$ and $\tilde{\rho}_{fc}$ 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 $I$ be the smallest interval that contains the support of $\mu_a$, and assume that

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

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

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

for sufficient large $N$.

The above assumption ensures that the density function $\rho_{fc}$ and $\tilde{\rho}_{fc}$ 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, 3.5 in [37]) Under the Assumption 2.3, there exists $L_- \in \mathbb{R}$ such that supp $\tilde{\rho}_{fc} = [L_-, L_+]$, and $\tilde{\rho}_{fc}$ is strictly positive in $(L_-, L_+)$. Moreover, there exists $C > 1$ such that

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

where $\kappa := \min\{|E - L_-|, |E - L_+|\}$. The points $L_\pm$ are the two real solutions to the equation

$$\int_{\mathbb{R}} \frac{d\mu_a(a)}{(a - L_\pm - m_{fc}(L_\pm))^2} = 1. \quad \text{(2.7)}$$

The same holds true, for sufficiently large $N$, if we replace $\mu_a$, $\tilde{\rho}_{fc}$, $L_\pm$ and $\kappa$ by $\mu_A$, $\rho_{fc}$, $L_\pm$ and $\kappa$, respectively. Here $[L_-, L_+]$ is the support of $\rho_{fc}$ and $\kappa := \min\{|E - L_-|, |E - L_+|\}$. 
2.2. Local law for the deformed Wigner matrices. We will use the following definition on high-probability estimates from [21].

**Definition 2.5.** 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) \( D > 0 \),

\[
\mathbb{P}(\mathcal{X}^{(N)} > N^\epsilon \mathcal{Y}^{(N)}) \leq N^{-D},
\]

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

We further introduce the spectral domain,

\[
D' := \{ z = E + i\eta : |E| \leq M, N^{-1+c} \leq \eta \leq 3 \},
\]

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

\[
\Psi(z) := \sqrt{\frac{\text{Im} m_{fc}(z)}{N \text{Im} z}} + \frac{1}{N \text{Im} z}, \quad \Theta(z) := -\frac{1}{N \text{Im} z}.
\]

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

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

The following local law for the Green function was proved in [35].

**Theorem 2.6.** (Local law for the deformed Wigner matrix, Theorem 2.10 in [35]) Under the Assumptions 2.1-2.3, the following holds

\[
\max_{ij} \left| G_{ij}(z) - \delta_{ij} \frac{1}{a_i - z - m_{fc}(z)} \right| < \Psi(z), \quad \left| N^{-1} \text{Tr} G(z) - m_{fc}(z) \right| < \Theta(z),
\]

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 \( N^c \), on the size of the fluctuations \( \text{Tr} G(z) - \mathbb{E} \text{Tr} G(z) \). It is hence natural to study the fluctuations of \( \text{Tr} G(z) - \mathbb{E} \text{Tr} G(z) \). The CLT for the linear eigenvalue statistics for general test functions is proved in [16] and [29] on global scale when \( \text{Im} z \) 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 2.7** (Theorem 2.15 of [29]). Under the 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 \( \sum_{i=1}^N \varphi(\lambda_i) - N \int_{\mathbb{R}} \varphi(x) \rho_{fc}(x) dx \) converges in distribution to the Gaussian random variable with mean \( M(\varphi) = \frac{1}{2\pi i} \int_{\Gamma} \varphi(z) b(z) dz \), and variance \( V(\varphi) = \frac{1}{(2\pi)^2} \int_{\Gamma} \int_{\Gamma} \varphi(z_1) \varphi(z_2) \Gamma(z_1, z_2) dz_1 dz_2 \), where

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

and \( \Gamma(z_1, z_2) =

\[
(m_2 - 2) \frac{\partial^2 \tilde{I}}{\partial z_1 \partial z_2} + (W_4 - 3) \left( \tilde{I} \frac{\partial^2 \tilde{I}}{\partial z_1 \partial z_2} + \frac{\partial \tilde{I}}{\partial z_1} \frac{\partial \tilde{I}}{\partial z_2} \right) + \frac{2}{(1 - \tilde{I})^2} \left( \frac{\partial \tilde{I}}{\partial z_1} \frac{\partial \tilde{I}}{\partial z_2} + (1 - \tilde{I}) \frac{\partial^2 \tilde{I}}{\partial z_1 \partial z_2} \right),
\]

with

\[
\tilde{I}(z_1, z_2) := \int_{\mathbb{R}} \frac{1}{(x - z_1 - \tilde{m}_{fc}(z_1))(x - z_2 - \tilde{m}_{fc}(z_2))} d\mu_n(x),
\]

and \( \Gamma \) is a rectangular contour with vertices \((a_{\pm} \pm iv_0)\) so that \( \pm (a_{\pm} - \tilde{L}_\pm) > 0 \) and \( \Gamma \) lies within the analytic domain of \( \varphi \).

Using ideas of M. Shcherbina [43], the above result can be extended to \( C_c^2(\mathbb{R}) \) test functions. In [16], 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
\[
f_N(x) := g \left( \frac{x - E_0}{\eta_0} \right).
\]
We will write $f_N$ as $f$ for notational simplicity. Define
\[
\kappa_0 := \text{dist}(\text{supp}(f), \{L_+, L_-\}).
\]

Following [39, 33], we study the characteristic function
\[
\phi(\lambda) := E[e^{\lambda X}], \quad \text{where } e(\lambda) = \exp \left\{ i\lambda(\text{Tr}f(X_N) - \mathbb{E}\text{Tr}f(X_N)) \right\}, \quad \lambda \in \mathbb{R}.
\]

Let $\tau > 0$ be an arbitrary small constant and define
\[
\Omega_0 := \{x + iy \in \mathbb{C} : |y| \geq N^{-\tau} \eta_0 \}.
\]

A key observation in [33] is that working on $\Omega_0$ instead of all $\mathbb{C}$, effectively removes the ultra-local scales without affecting the mesoscopic linear statistics.

**Proposition 2.8.** Let $X_N$ be a deformed Wigner matrix satisfying the Assumptions 2.1, 2.2 and 2.3. Define
\[
I(z_1, z_2) := \frac{1}{\pi} \int_{\Omega_0} \frac{1}{(x - z_1 - m_{fc}(z_1))(x - z_2 - m_{fc}(z_2))} d\mu_A(x),
\]
and
\[
V(f) = \frac{1}{2\pi} \int_{\Omega_0} \int_{\Omega_0} \frac{\partial}{\partial z_1} f(z_1) \frac{\partial}{\partial z_2} f(z_2) \left[ \frac{\partial^2}{\partial z_1 \partial z_2} \left( m_2 - \frac{2}{\beta} I + \frac{W_4 - 1 - 2}{2} I^2 \right) \right]
+ \frac{2}{\beta} \frac{\partial}{\partial z_1} \left( \frac{1}{1 - I \partial z_2} \right) \bigg] d^2 z_1 d^2 z_2,
\]
where $\tilde{f}$ is an almost analytic extension of $f$ given in Lemma 3.1 below. Assuming that $V(f) < 1$, we have
\[
\phi'(\lambda) = -\lambda \phi(\lambda) V(f) + \tilde{\xi},
\]
where
\[
|\tilde{\xi}| = O_{\prec} \left( |\lambda| \log NN^{-\tau} \right) + O_{\prec} \left( \frac{1 + |\lambda|^4}{N^{\eta_0 \sqrt{\kappa_0} + \eta_0}} \right) + O_{\prec} \left( \frac{1 + |\lambda|^4}{\sqrt{NN\eta_0}} \right).
\]

Proposition 2.8 implies the following result.

**Theorem 2.9.** Under the same assumptions as in Proposition 2.8, if we further assume that there exist $c_1, c_2 > 0$ such that $V(f_N) \geq c_1$ and $\eta_0 \sqrt{\kappa_0} + \eta_0 \geq N^{-1+c_2}$ for sufficient large $N$, then $\frac{\text{Tr}f(X_N) - \mathbb{E}\text{Tr}f(X_N)}{\sqrt{V(f)}}$ converges in distribution to a standard Gaussian distribution.

We remark that the Theorem 2.9 applies to the global scale as well as the mesoscopic scales. The expectation of $\text{Tr}f(X_N)$ has the following asymptotic expansion, which matches the result in [16, 29] on the global scale.

**Proposition 2.10.** Define
\[
I_s(z) := \frac{1}{\pi} \int_{\mathbb{R}} \frac{1}{(x - z - m_{fc}(z))^2} d\mu_A(x).
\]
Then the bias is given by $\mathbb{E}\text{Tr}f(X_N) - N \int \tilde{f}(x) \rho_{fc}(x) dx =
\[
\frac{1}{2\pi} \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z) \left( \left( \frac{2}{\beta} - 1 \right) \frac{1}{1 - I_s(z)} \frac{1}{dz} I_s(z) \right) dz + \frac{m_2}{1 - I_s(z)} \frac{1}{dz} I_s(z) \left( W_4 - 1 - \frac{2}{\beta} \right) I_s(z) \frac{1}{dz} I_s(z) dz
+ O(N^{-\tau}) + O_{\prec} \left( \frac{1}{\sqrt{NN\eta_0 \sqrt{\kappa_0} + \eta_0}} \right).
\]

On the mesoscopic scales, we obtain the following universal CLTs for the linear eigenvalue statistics in the bulk and at the regular edges.
Theorem 2.11. (Mesoscopic CLT in the bulk) Let $X_N$ be a deformed Wigner matrix satisfying the Assumptions 2.1, 2.2 and 2.3. Let $\eta_0 = N^{-1+a_1}$ with some small $a_1 > 0$, $E_0 \in (L_-, L_+)$ such that $\kappa_0 > c_0$, for some $c_0 > 0$. Then, for any function $g \in C^4_c(\mathbb{R})$, the linear statistics

$$
\sum_{i=1}^{N} g \left( \frac{\lambda_i - E_0}{\eta_0} \right) - N \int_{\mathbb{R}} g \left( \frac{x - E_0}{\eta_0} \right) \rho_{f_c}(x) dx
$$

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

$$
\frac{1}{2^{\beta} \pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \frac{g(x_1) - g(x_2)}{(x_1 - x_2)^2} \right)^2 dx_1 dx_2.
$$

(2.19)

In particular, the bias vanishes in the bulk regime.

Theorem 2.12. (Mesoscopic CLT at the edge) Let $X_N$ be a deformed Wigner matrix satisfying the Assumptions 2.1, 2.2 and 2.3. Let $\eta_0 = N^{-\frac{2}{3}+a_2}$ with some small $a_2 > 0$. For any function $g \in C^4_c(\mathbb{R})$, the linear statistics

$$
\sum_{i=1}^{N} g \left( \frac{\lambda_i + L_+}{\eta_0} \right) - N \int_{\mathbb{R}} g \left( \frac{x + L_+}{\eta_0} \right) \rho_{f_c}(x) dx - \left( \frac{2}{\beta} - 1 \right) \frac{g(0)}{4}
$$

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

$$
\frac{1}{4^{\beta} \pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \frac{g(-x^2) - g(-y^2)}{x - y} \right)^2 dx dy.
$$

(2.20)

At the left edge $L_-$, we obtain a similar CLT with variance

$$
\frac{1}{4^{\beta} \pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \frac{g(x^2) - g(y^2)}{x - y} \right)^2 dx dy.
$$

Remark: The bulk variance (2.19) agrees with the GOE/GUE. For the edges, the bias and variance in (2.20) coincide with those of the GUE/GOE obtained in [5, 40] 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 $C^{1+,\alpha}(\mathbb{R})$ functions; see [27]. Finally, for test functions in $C^4_c(\mathbb{R})$, we can relax the single support condition for $\mu_{f_c}$ by assuming instead that the cuts of the support of $\mu_{f_c}$ are separated by order one.

3. Proof of Proposition 2.8

In this section, we prove Proposition 2.8 by reducing it to the main technical result Lemma 3.4. Recall the scaled test function $f$ from (2.12). There are constants such that

$$
\|f\|_1 \leq C_{\eta_0}; \quad \|f\|_1 \leq C'; \quad \|f''\|_1 \leq \frac{C''}{\eta_0}.
$$

(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 $f \in C^2_c(\mathbb{R})$ and $\chi(y) = 1$ for $|y| \leq 1$. Define its almost-analytic extension

$$
\tilde{f}(x + iy) := (f(x) + i y f'(x)) \chi(y).
$$

Then we have

$$
f(\lambda) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial}{\partial \lambda} \tilde{f}(z) d^2 z = \frac{1}{2\pi} \int_{\mathbb{R}^2} \frac{iyf''(x) \chi(y) + i f(x) + iyf'(x) \chi'(y)}{\lambda - x - iy} dxdy,
$$

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

Therefore, we write

$$
\text{Tr} f(X_N) - \mathbb{E} \text{Tr} f(X_N) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial}{\partial \sigma} \tilde{f}(z)(\text{Tr}(G(z)) - \mathbb{E} \text{Tr} G(z)) d^2 z.
$$

(3.2)

Plugging the above equation in $e(\lambda)$ given by (2.14), we have

$$
e(\lambda) = \exp \left\{ \frac{i \lambda}{\pi} \int_{\mathbb{C}} \frac{\partial}{\partial \sigma} \tilde{f}(z)(\text{Tr}(G(z)) - \mathbb{E} \text{Tr} G(z)) d^2 z \right\}.
$$

(3.3)
Taking the derivative of the characteristic function given in (2.14), and applying (3.3), we get
\[
\phi'(\lambda) = \frac{i}{\pi} \int_O \frac{\partial}{\partial z'} \tilde{f}(z) \mathbb{E} \left[ e(\lambda)(\text{Tr}(G(z)) - \text{ETr}(G(z))) \right] dz. \tag{3.4}
\]

Following [33], we restrict the domain of the spectral parameter to \( \Omega_0 \), as the very local scales do not contribute to \( \phi(\lambda) \). Indeed, using that \( y \rightarrow \text{Im} m_N(z) y \) is increasing, we can extend the local law as follows:
\[
|\text{Tr}G(x + iy) - \text{ETr}G(x + iy)| \lesssim O \left( \frac{1}{|y|} \right), \tag{3.5}
\]
uniformly in \(|y| > 0 \) and \(|x| < M \); see (2.9). Together with (3.1), we have
\[
\text{Tr} f(X_N) - \text{ETr} f(X_N) = \frac{1}{\pi} \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z)(\text{Tr}(G(z)) - \text{ETr}(G(z))) dz^2 + O_\prec (N^{-\tau}). \tag{3.6}
\]
Using the same argument, since \(|e(\lambda)| = 1\), we have
\[
\phi'(\lambda) = \frac{i}{\pi} \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z) \mathbb{E} \left[ e(\lambda)(\text{Tr}(G(z)) - \text{ETr}(G(z))) \right] dz^2 + O_\prec (N^{-\tau}). \tag{3.7}
\]
Similarly, we restrict the integration domain of \( e(\lambda) \) in (3.3) to \( \Omega_0 \). Let
\[
e_0(\lambda) := \exp \left\{ \frac{i\lambda}{\pi} \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z)(\text{Tr}(G(z)) - \text{ETr}(G(z))) dz^2 \right\}. \tag{3.8}
\]
In addition, (3.6) implies that \(|e(\lambda) - e_0(\lambda)| = O_\prec (|\lambda| N^{-\tau})\). We also have \(|e_0(\lambda)| = 1\), using \(|e(\lambda)| = 1\) and \(\text{Tr}(\pi) = \text{Tr}(\tilde{G})\). If we further replace \(e(\lambda)\) by \(e_0(\lambda)\) in (3.7), then we get
\[
\phi'(\lambda) = \frac{i}{\pi} \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z) \mathbb{E} \left[ (e_0(\lambda)(\text{Tr}(G(z)) - \text{ETr}(G(z)) \right] + O_\prec (|\lambda| \log N N^{-\tau}). \tag{3.9}
\]

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 [33]. The proof is provided in Appendix B.

**Lemma 3.2.** Suppose \( H(z) \) is a holomorphic function on \( \Omega_0 \) and \(|H(z)| \leq \frac{K}{|z|^s}\) for some constants \(s, K \geq 0\), then there exists some constant \( C \) such that
\[
\left| \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z) H(z) dz^2 \right| \leq CK N^{r+s} \rho_0^{1-s}.
\]
Thus, in order to study \( \phi'(\lambda) \), it is sufficient to estimate \( \mathbb{E} \left[ (e_0(\lambda)(\text{Tr}(G(z)) - \text{ETr}(G(z)) \right\}

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 \( h \). Then for any fixed \( n \in \mathbb{N} \), we have
\[
\mathbb{E}h f(h) = \sum_{k=0}^{1} \frac{1}{k!} c_{k+1}(h) f^{(k)}(h) + R_{t+1},
\]
where the error term satisfies
\[
|R_{t+1}| \leq C_i |h|^{t+2} \sup_{|x| \leq M} |f^{(t+1)}(x)| + C_i |h|^{t+2} 1_{|h| > M} \| f^{(t+1)} \|_\infty,
\]
and \( M > 0 \) is an arbitrary fixed cutoff.

For reference, we refer e.g. to Lemma 3.1 in [27]. We give the proof the following lemma in Section 5.

**Lemma 3.4.** For any \( z \in \Omega_0 \cap D' \), see (2.9), we have
\[
\mathbb{E}[e_0(\lambda)(\text{Tr}(G(z) - \text{ETr}(G(z))) = \frac{i\lambda}{\pi} \int_{\Omega_0} \frac{\partial}{\partial z'} \tilde{f}(z') K(z, z') dz'^2 + \mathcal{E}(z),
\]
where \( I \) is given in (2.16) and
\[
K(z, z') = 2 \frac{\partial}{\partial z'} \left( \frac{1}{1 - I(z, z')} \frac{\partial}{\partial z} I(z, z') \right) + (m_2 - 2) \frac{\partial^2}{\partial z \partial z'} I(z, z') + \frac{W_4 - 3}{2} \frac{\partial^2}{\partial z \partial z'} I^2(z, z'),
\]
and $E(z)$ is analytic in $\Omega_0$ and satisfies
\[ E(z) = O_\prec \left( \frac{1 + |\lambda|^4}{\sqrt{\kappa}} \right) \left( \frac{1}{N(Imz)^2} + \frac{1}{\sqrt{N(Imz)^2}} + \frac{1}{N\eta_0 Imz} + \frac{1}{\sqrt{N\eta_0 Imz}} \right). \]

Admitting Lemma 3.4 and plugging in (3.9), we have
\[ \phi'(\lambda) = -\lambda e(\lambda) V(f) + O_\prec \left( |\lambda| \log NN^{-r} \right) + \tilde{E}, \]
where
\[ V(f) = \frac{1}{\pi i} \int_{\Omega_0} \frac{d}{dz} \tilde{f}(z) \frac{d}{dz'} \tilde{f}(z') K(z, z') d^2 z d^2 z', \quad \tilde{E} = \frac{i}{\pi} \int_{\Omega_0} \frac{d}{dz} \tilde{f}(z) \tilde{E}(z) d^2 z. \]
By the definition of $\kappa_0$ in (2.13), $\kappa \geq \kappa_0$. Moreover $|\text{Im} z| \geq N^{-r} \eta_0$, for $z \in \Omega_0$. Using Lemma 3.2, we hence obtain the estimate
\[ \tilde{E} = O_\prec \left( \frac{(1 + |\lambda|^4)N^\frac{3}{4}}{N\eta_0 \sqrt{\kappa_0 + \eta_0}} \right) + O_\prec \left( \frac{(1 + |\lambda|^4)N^\frac{3}{4}}{\sqrt{N\eta_0}} \right). \]
Assuming $V(f) \prec O(1)$, we can replace $e_0(\lambda)$ by $e(\lambda)$ with error $O_\prec \left( |\lambda|N^{-r} \right)$. Thus we have completed the proof of Proposition 2.8.

4. Properties of the free convolution

4.1. Properties of $m_{fc}$ and $\tilde{m}_{fc}$. In this subsection, we recall some properties of the Stieltjes transforms $m_{fc}$ and $\tilde{m}_{fc}$ of the free convolution measures. Let $\kappa = \kappa(E)$ be the distance from $E$ to the closest spectral edge, i.e.
\[ \kappa = \min \{|E - L_-|, |E - L_+|\}. \]
Similarly define $\tilde{\kappa} = \min \{|E - \tilde{L}_-|, |E - \tilde{L}_+|\}$. Define the spectral domain
\[ D := \{ z = E + i\eta : |E| < M, 0 < \eta \leq 3 \} \].

**Lemma 4.1. (Lemma 3.5, Lemma A.1 in [37])**

1. There exists $C > 1$ such that
\[ C^{-1} \sqrt{\kappa + \eta} \leq |\text{Im} m_{fc}(z)| \leq C \sqrt{\kappa + \eta}, \quad (4.1) \]
if $E \in [\tilde{L}_-, \tilde{L}_+]$. If $E \in [L_-, L_+]^c$, then
\[ C^{-1} \frac{\eta}{\sqrt{K + \eta}} \leq |\text{Im} \tilde{m}_{fc}(z)| \leq C \frac{\eta}{\sqrt{K + \eta}}. \quad (4.2) \]

2. (Stability bound) There exists $C > 1$, such that
\[ C^{-1} \leq |a - z - \tilde{m}_{fc}(z)| \leq C, \quad (4.3) \]
uniformly for $z \in D$ and $a \in \text{supp}(\mu_\alpha)$. 

3. There exist $k, K > 0$ such that
\[ k \sqrt{\kappa + \eta} \leq \left| 1 - \int_D \frac{1}{(x - z - m_{fc}(z))^2} d\mu_\alpha(x) \right| \leq K \sqrt{\kappa + \eta}. \quad (4.4) \]

4. There exist $C > 0$ and $c_0 > 0$ such that for all $z \in D$ satisfying $\kappa + \eta \leq c_0$,
\[ C^{-1} \leq \left| \int_D \frac{1}{(x - z - m_{fc}(z))^3} d\mu_\alpha(x) \right| \leq C; \quad (4.5) \]
moreover, there exists $C > 1$ such that for all $z \in D$,
\[ \left| \int_D \frac{1}{(x - z - m_{fc}(z))^3} d\mu_\alpha(x) \right| \leq C. \]

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

Lemma 4.2. (Lemma 3.6 in [37]) Under the Assumptions 2.2 and 2.3, for sufficiently large \( N \), statements 1-4 in Lemma 4.1 hold true with \( \tilde{m}_{fc}, \hat{\kappa}, \mu_{\alpha} \) and \( \tilde{L}_{\pm} \) replaced by \( m_{fc}, \kappa, \mu_{A} \) and \( L_{\pm} \) respectively. Moreover, the constants in these inequalities can be chosen uniformly in \( N \) for sufficient large \( N \). Furthermore, there exists \( c > 0 \) such that
\[
\max_{z \in D} |\tilde{m}_{fc}(z) - m_{fc}(z)| \leq N^{-\frac{\alpha}{2}}, \quad |\tilde{L}_{\pm} - L_{\pm}| \leq N^{-\alpha c}\]
for sufficient large \( N \).

Recall the function \( I(z_1, z_2) \) given in (2.16) and \( I_s(z) \) in (2.18). As analogues we also define \( \tilde{I} \) by replacing the measure \( \mu_{A} \) by its limiting measure \( \mu_{\alpha} \), i.e.
\[
\tilde{I}(z_1, z_2) := \int_{\mathbb{R}} \frac{1}{(x - z_1 - \tilde{m}_{fc}(z_1))(x - z_2 - \tilde{m}_{fc}(z_2))} \, d\mu_{\alpha}(x); \quad \tilde{I}_s(z) := \tilde{I}(z, z).
\]
By direct computation, one proves the following lemma. It holds for both \( I \) and \( \tilde{I} \).

Lemma 4.3. For \( z_1 \neq z_2 \), we have
\[
I(z_1, z_2) = \frac{m_{fc}(z_1) - m_{fc}(z_2)}{z_1 + m_{fc}(z_1) - z_2 - m_{fc}(z_2)}; \quad I_s(z) = \frac{m'_{fc}(z)}{1 + m'_{fc}(z)}.
\]

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
\[
|I(z_1, z_2)| \leq C; \quad |I_s(z)| \leq 1; \quad C^{-1}\sqrt{\kappa + \eta} \leq |1 - I_s(z)| \leq C\sqrt{\kappa + \eta};
\]
\[
|m_{fc}(z)| \leq C; \quad |m'_{fc}(z)| \leq \frac{C}{\sqrt{\kappa + \eta}}; \quad |m''_{fc}(z)| \leq \frac{C}{(\sqrt{\kappa + \eta})^2},
\]
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 2.6, we introduce the anisotropic local law. Recall the control parameters \( \Psi \) and \( \Theta \) from (2.10).

Theorem 4.5. (Theorem 12.2, 12.4 in [32]; Theorem 2.1, 2.2 in [22]; 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
\[
\left| \langle v, G(z)w \rangle - \langle v, \tilde{G}(z)w \rangle \right| \leq \|v\|_2\|w\|_2 \Psi(z); \quad \left| N^{-1}\text{Tr}(B(G(z) - \tilde{G}(z))) \right| \leq \|B\|_{\text{op}} \Theta(z),
\]
uniformly in \( z \in \mathbb{C} \).}

Finally, we conclude this subsection by recalling some properties of stochastic domination. More details are found in Chapter 6.3 in [23].

Lemma 4.6. (Proposition 6.5 in [23])

(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 \( X \prec Y + N^{-\epsilon}X \) for some \( \epsilon > 0 \), then \( X \prec Y \);

(4) If \( X \prec Y, \; \mathbb{E}Y \geq N^{-c} \) and \( |X| \leq N^{c} \) almost surely with some fixed exponent \( c \), then we have \( \mathbb{E}X \prec \mathbb{E}Y \).

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
\[
g_i(z) := \frac{1}{a_i - z - m_{fc}(z)}, \quad z \in \mathbb{C} \setminus \text{supp}(\mu_{fc}).
\]
Before we proceed the proof of Lemma 3.4, we state a useful lemma.
Lemma 5.1. For any $i,j$, we have

\[ \frac{\partial e_0(\lambda)}{\partial H_{ij}} = -\frac{i(2 - \delta_{ij})\lambda}{\pi} e_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) \frac{d}{dz} G_{ji} d^2z. \] (5.1)

\[ \frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}} = \frac{i(2 - \delta_{ij})\lambda}{\pi} e_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) \frac{d}{dz} (g_i(z)g_j(z)) d^2z + O_\prec \left( \frac{(1 + |\lambda|^2)}{\sqrt{N \theta_0}} \right). \] (5.2)

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

\[ \left| \frac{\partial^k G_{ij}}{\partial H_{ij}^k} \right| < O(1); \quad \left| \frac{\partial^k e_0(\lambda)}{\partial^k H_{ij}} \right| < O((1 + |\lambda|)^k). \] (5.3)

The above lemma follows from the relation

\[ \frac{\partial G_{ij}}{\partial H_{ab}} = -\frac{G_{ia}G_{bj} + G_{ib}G_{aj}}{1 + \delta_{ab}}. \] (5.4)

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

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

Using the cumulant expansion Theorem 3.3, we obtain

\[ (z - a_i)E[e_0(\lambda)(G_{ii} - \mathbb{E}G_{ii})] = I_1 + I_2 + I_3 + O_\prec (N^{-\frac{2}{3}}(1 + |\lambda|^4)), \] (5.5)

where

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

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

\[ I_3 = \frac{1}{3N^2} \sum_{j=1}^{N} c_{ij}^{(4)} \left( \mathbb{E} \left[ \frac{\partial^3 e_0(\lambda)}{\partial^3 H_{ij}} G_{ji} \right] + 3\mathbb{E} \left[ \frac{\partial^2 e_0(\lambda)}{\partial^2 H_{ij}} \frac{\partial G_{ji}}{\partial H_{ij}} \right] + 3\mathbb{E} \left[ \frac{\partial e_0(\lambda)}{\partial H_{ij}} \frac{\partial^2 G_{ji}}{\partial^2 H_{ij}} \right] \right. \]

\[ \left. + \mathbb{E} \left[ \left( \frac{\partial^3 G_{ji}}{\partial^3 H_{ij}} - \mathbb{E} \left[ \frac{\partial^3 G_{ji}}{\partial^3 H_{ij}} \right] \right)e_0(\lambda) \right] \right). \]

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

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

The last error term of (5.5) is estimated by (5.3), (2.2) and Lemma 4.6. Note that for $z \in \Omega_0 \cap D'$, we have the deterministic bound $|G_{ij}| \leq ||G||_{op} \leq (\Im z)^{-1} = O(N^2)$. Combining with $|e_0(\lambda)| = 1$, we can use the fourth statement of Lemma 4.6. 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 $I_1, I_2, I_3$ respectively.
5.1. Estimate on $I_1$. Using (5.4), we have for each $i$,  
\[
I_1 = - \frac{1}{N} \mathbb{E}[e_0(\lambda)((G^2)_{ii} - \mathbb{E}(G^2)_{ii})] - \frac{1}{N} \mathbb{E}[e_0(\lambda)(\text{Tr}GG_{ii} - \text{ETr}GG_{ii})] \\
- \frac{m_2 - 2}{N} \mathbb{E}[e_0(\lambda)(G_{ii}G_{ii} - \mathbb{E}G_{ii}G_{ii})] + \frac{1}{N} \sum_{j=1}^{N} (1 + (m_2 - 1)\delta_{ij}) \mathbb{E} \left[ \frac{\partial e_0(\lambda)}{\partial H_{ij}} G_{ji} \right]
\]
\[
:= A_1(i) + A_2(i) + A_3(i) + A_4(i).
\]

Next, we consider the linear statistics of $A_1(i)$,  
\[
\sum_{i=1}^{N} g_i(z)A_1(i) = -\mathbb{E} \left[ e_0(\lambda)(1 - \mathbb{E}) \left( \frac{1}{N} \sum_{i=1}^{N} g_i(z) d i z G_{ii}(z) \right) \right]
\]
\[
= -\mathbb{E} \left[ e_0(\lambda)(1 - \mathbb{E}) \frac{d}{dz} \left( \frac{1}{N} \sum_{i=1}^{N} g_i(z) G_{ii} \right) \right] + \mathbb{E} \left[ e_0(\lambda)(1 - \mathbb{E}) \left( \frac{1}{N} \sum_{i=1}^{N} g_i(z) G_{ii} \right) \right].
\]  

This will be used later in Section 5.4. We first use the anisotropic local law to deal with the first term of (5.6). Let $B$ in Theorem 4.5 to be $B = \text{Diag}(g_i(z))$. By (4.3), we have \( \|B\|_{\text{op}} \leq C \). Though $B$ depends on $z$, since $g_i(z)$ is uniformly bounded and analytic in $\Omega_0 \cap D'$, then \( |g_i'(z)| \leq \frac{C}{\text{Im} z} = O(N^0) \), we can use a continuity argument to show that the anisotropic local law still holds, i.e.,  
\[
\left| \frac{1}{N} \sum_{i=1}^{N} g_i(z) G_{ii}(z) \right| - \frac{1}{N} \sum_{i=1}^{N} (g_i(z))^2 < \Theta(z),
\]
with $\Theta$ as in (2.10). Since $G_{ii}(z)$ and $g_i(z)$ are analytic in $\Omega_0 \cap D'$, the Cauchy integral formula yields  
\[
\mathbb{E} \left[ e_0(\lambda)(1 - \mathbb{E}) \frac{d}{dz} \left( \frac{1}{N} \sum_{i=1}^{N} g_i(z) G_{ii} \right) \right] < \Theta(z) \frac{1}{\text{Im} z}.
\]  

Similarly using the anisotropic local law, the second term of (5.6) is also bounded from above by $\frac{\Theta(z)}{\text{Im} z}$. Thus,  
\[
\sum_{i=1}^{N} g_i(z)A_1(i) \prec \Theta(z) \frac{1}{\text{Im} z} = N\Theta^2(z).
\]  

Next, we consider the second term $A_2(i)$. Using the local law, we have  
\[
A_2(i) = - \frac{1}{N} \mathbb{E}[e_0(\lambda)(\text{Tr}G(G_{ii} - \mathbb{E}G_{ii}) + \mathbb{E}G_{ii}(\text{Tr}G - \text{ETr}G) + \text{ETr}GEG_{ii} - \text{ETr}GG_{ii})]
\]
\[
= -m_{f_2}(z) \mathbb{E}[e_0(\lambda)(G_{ii} - \mathbb{E}G_{ii})] - \frac{1}{N} \mathbb{E}[e_0(\lambda)(\text{Tr}G - \text{ETr}G)] + O_\prec \left( \Theta(z)\Psi(z) \right),
\]
with $\Psi(z)$ as in (2.10). Here the first term of $A_2$ will be moved to the left side of the equation (5.5). Note that if we take the linear statistics of the error term as in (5.8), using the same argument in (5.7), we will get an error as $O_\prec \left( N\Theta^2(z) \right)$. In addition, the local law also implies that $A_3(i) = O_\prec \left( \frac{\Psi(z)}{\text{Im} z} \right)$.

Note that $A_4$ is a leading term of $I_1$. Using the local law, (5.1) and Lemma 5.3, we write  
\[
A_4(i) = A_{41}(i) + A_{42}(i) + O_\prec \left( (1 + |\lambda|)N^{-1}\Psi(z) \right),
\]
where  
\[
A_{41}(i) = \frac{1}{N} \sum_{j=1}^{N} \mathbb{E} \left[ \frac{\partial e_0(\lambda)}{\partial H_{ij}} (1 + \delta_{ij})G_{ji} \right], \quad \text{and} \quad A_{42}(i) = \frac{m_2 - 2}{N} \mathbb{E} \left[ \frac{\partial e_0(\lambda)}{\partial H_{ii}} g_i(z) \right].
\]  

We compute these two terms below in the Section 5.4.
5.2. Estimate on $I_2$. In this subsection, we will show that $I_2$ is negligible. Since the third cumulants are bounded, it is enough to study
\[
\frac{1}{N^2} \sum_{j=1}^{N} c_{ij}^{(3)} \left[ \mathbb{E} \left( \frac{\partial^2 \epsilon_0(\lambda)}{\partial^2 H_{ij}} G_{ij} \right) \right] + 2\mathbb{E} \left[ \frac{\partial \epsilon_0(\lambda)}{\partial H_{ij}} \frac{\partial G_{ij}}{\partial H_{ij}} \right] + \mathbb{E} \left[ \epsilon_0(\lambda) \left( \frac{\partial^2 G_{ij}}{\partial^2 H_{ij}} - \mathbb{E} \frac{\partial^2 G_{ij}}{\partial^2 H_{ij}} \right) \right]
:= B_1(i) + B_2(i) + B_3(i).
\]
First, we study the last term $B_3(i)$. Using (5.4) and the local law, we have, for some coefficients $K_1$ and $K_2$, that
\[
B_3(i) = \frac{1}{N^2} \sum_{j=1}^{N} c_{ij}^{(3)} \left[ \mathbb{E} \left( \epsilon_0(\lambda) \left( K_1 G_{ii} G_{jj} G_{ij} + K_2(G_{ij})^3 - K_1 \mathbb{E}[G_{ii}G_{jj}G_{ij}] - K_2 \mathbb{E}(G_{ij})^3 \right) \right) \right]
\]
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 $w_j = \frac{1}{\sqrt{N}} c_{ij}^{(3)} g_j(z)$. Note that $\|w\|_2$ is bounded because of the stability bound (4.3) and the moment condition (2.2). Though $w$ depends on $z$, we can use a continuity argument to show that
\[
\left| \frac{1}{\sqrt{N}} \sum_{j=1}^{N} c_{ij}^{(3)} g_j(z) G_{ij}(z) - \frac{1}{\sqrt{N}} c_{ij}^{(3)} (g_j(z))^2 \right| \prec \Psi(z).
\]  
Because of (4.3) and $\Psi(z) \geq C N^{-\frac{1}{2}}$, we have $\left| \frac{1}{\sqrt{N}} \sum_{j=1}^{N} c_{ij}^{(3)} g_j(z) G_{ij}(z) \right| \prec \Psi(z)$. Therefore, we obtain the upper bound
\[
B_3(i) = O_\prec \left( N^{-1} \Psi(z) \right) + O_\prec \left( N^{-\frac{1}{2}} \Psi^2(z) \right) = O_\prec \left( N^{-\frac{1}{2}} \Psi^2(z) \right).
\]
For the second term, by (5.4), (5.1), and the local law we have
\[
B_2(i) = -\frac{2}{N^2} \sum_{j=1}^{N} c_{ij}^{(3)} \mathbb{E} \left[ \frac{\partial \epsilon_0(\lambda)}{\partial H_{ij}} (G_{ij} G_{jj} + G_{ii} G_{jj}) \right]
\]
\[
= -\frac{2}{N^2} \sum_{j=1}^{N} c_{ij}^{(3)} \mathbb{E} \left[ \frac{\partial \epsilon_0(\lambda)}{\partial H_{ij}} g_i(z) g_j(z) \right] + O \left( \frac{(1 + |\lambda|) \Psi(z)}{N \sqrt{\eta_0}} \right)
\]
\[
= \frac{4\lambda}{\pi N^2} \mathbb{E} \left[ \epsilon_0(\lambda) \sum_{j=1}^{N} c_{ij}^{(3)} \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z') \frac{d^2 z'}{d^2 (G(z'))_{ij}} \right] + O \left( \frac{(1 + |\lambda|) \Psi(z)}{N \sqrt{\eta_0}} \right).
\]
By the same argument as in (5.10) and the Cauchy integral formula, we have
\[
\left| \frac{d}{dz} \frac{1}{\sqrt{\eta_0}} \left( \sum_{j=1}^{N} c_{ij}^{(3)} g_j(z) (G(z'))_{ij} \right) \right| \prec \frac{\Psi(z')}{\text{Im} z'}.
\]
Using the stability bound (4.3) and Lemma 3.2, we have
\[
|B_2(i)| \prec \frac{1 + |\lambda|}{N \sqrt{\eta_0}} \frac{(1 + |\lambda|) \Psi(z)}{N \sqrt{\eta_0}} = O_\prec \left( \frac{(1 + |\lambda|) \Psi(z)}{N \sqrt{\eta_0}} \right).
\]
Similarly, by plugging (5.2) in the expression of $B_1$, we have
\[
B_1(i) = \frac{2\lambda}{\pi N^2} \sum_{j=1}^{N} c_{ij}^{(3)} \mathbb{E} \left[ \epsilon_0(\lambda) \left( \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z') \frac{d^2 z'}{d^2 (g_i(z') g_j(z'))} \right) G_{ij} \right] + O \left( \frac{(1 + |\lambda|^2) \Psi(z)}{N \sqrt{\eta_0}} \right).
\]
Using the anisotropic local law, we have
\[
B_1(i) = O_\prec \left( (1 + |\lambda|^2) N^{-1} \Psi(z) \right) + O_\prec \left( \frac{(1 + |\lambda|^2) \Psi(z)}{N \sqrt{\eta_0}} \right) = O_\prec \left( \frac{(1 + |\lambda|^2) \Psi(z)}{N \sqrt{\eta_0}} \right).
\]
Therefore, using the stability bound (4.3), we have
\[
\left| \sum_{i=1}^{N} g_i(z)(B_1(i) + B_2(i) + B_3(i)) \right| = O_{\prec} \left( \sqrt{N}\Psi^2(z) \right) + O_{\prec} \left( \frac{(1 + |\lambda|^2)\Psi(z)}{\sqrt{N}} \right).
\]

5.3. Estimate on \( I_3 \). It is not hard to show that the diagonal terms for \( i = j \) are negligible. Thus we can just 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 (5.3), we have \( |D_1(i)| < (1 + |\lambda|^2)N^{-1}\Psi(z) \). Similarly, using (5.4), (5.1) and the local law, we have \( |D_3(i)| < \frac{(1 + |\lambda|^2)\Psi(z)}{N\sqrt{N}} \). For the last term \( D_4 \), using (5.4) and the local law, for some coefficient \( K_1, K_2, K_3 \),
\[
D_4(i) = \frac{1}{6N^2} \sum_{j=1}^{N} \mathbb{E} \left[ c_0(\lambda)(1 - \mathbb{E} \left( K_1 G_{ii} G_{jj}(G_{ij})^2 + K_2(G_{ii})^2(G_{jj})^2 + K_3(G_{ij})^4 \right) \right] < N^{-1}\Psi(z).
\]

Finally, we look at the leading term \( D_2(i) \). Using the local law and (5.3), we have
\[
D_2(i) = - \frac{W_4 - 3}{2N^2} \sum_{j=1}^{N} \mathbb{E} \left[ \frac{\partial^2 c_0(\lambda)}{\partial^2 H_{ij}} ((G_{jj})^2 + G_{ii} G_{jj}) \right] = - \frac{W_4 - 3}{2N^2} \sum_{j=1}^{N} \mathbb{E} \left[ \frac{\partial^2 c_0(\lambda)}{\partial^2 H_{ij}} g_i(z)g_j(z) \right] + O_{\prec} \left( (1 + |\lambda|^2)N^{-1}\Psi(z) \right).
\]

5.4. Adding up the contributions to (5.5). Summing up the contributions from the previous subections, we write (5.5) as
\[
(z - a_i + m_{fc}) \mathbb{E}[c_0(\lambda)(G_{ii} - \mathbb{E} G_{ii})] = - \frac{1}{N} \sum_{i=1}^{N} g_i(z) \mathbb{E}[c_0(\lambda)(\text{Tr} G - \mathbb{E} \text{Tr} G)] + A_{41}(i) + A_{42}(i) + D_2(i) + \epsilon(i),
\]
where \( D_2 \) is given in (5.11) and \( A_{41}, A_{42} \) in (5.9), 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 \( z - a_i + m_{fc} \) to get
\[
\mathbb{E}[c_0(\lambda)(G_{ii} - \mathbb{E} G_{ii})] = \frac{1}{N} \sum_{i=1}^{N} g_i(z) \mathbb{E}[c_0(\lambda)(\text{Tr} G - \mathbb{E} \text{Tr} G)] + g_i(z) (A_{41}(i) + A_{42}(i) + D_2(i) + \epsilon(i)).
\]

Summing over \( i \) and rearranging, we find
\[
(1 - I_z(z)) \mathbb{E}[c_0(\lambda)(\text{Tr} G - \mathbb{E} \text{Tr} G)] = \sum_{i=1}^{N} g_i(z) (A_{41}(i) + A_{42}(i) + D_2(i)) + \mathcal{E}_1,
\]
where \( \mathcal{E}_1 \) is the linear statistics of \( \epsilon(i) \). By the argument in Section 5.1-5.3, we get
\[
\mathcal{E}_1 = O_{\prec} \left( (1 + |\lambda|^4)N\Theta^2(z) \right) + O_{\prec} \left( (1 + |\lambda|^4)\sqrt{N}\Psi^2(z) \right) + O_{\prec} \left( \frac{(1 + |\lambda|^4)\Psi(z)}{\sqrt{N}} \right).
\]

Next, we study the leading terms of the right side of (5.12). Plugging (5.1) in (5.9), we have
\[
\sum_{i=1}^{N} \frac{A_{41}(i)}{z - a_i + m_{fc}} = - \frac{2i\lambda}{\pi N} \sum_{i=1}^{N} g_i(z) \mathbb{E} \left[ c_0(\lambda) \int_{\mathbb{R}_n} \frac{\partial}{\partial z} \tilde{f}(z) \frac{\partial}{\partial z'} (G(z')G(z))_{ii} d^2z' \right].
\]
By the resolvent identity,
\[
G(z)G(z') = \frac{G(z) - G(z')}{z - z'}, \quad z \neq z',
\]
we can write
\[
F(z, z') := \frac{1}{N} \sum_{i=1}^{N} g_i(z)(G(z')G(z))_{ii} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z' - z} g_i(z)(G_{ii}(z') - G_{ii}(z)).
\]
We separate into two cases:

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

\[
F(z, z') - \frac{1}{z'-z} \frac{1}{N} \sum_{i=1}^{N} g_i(z)(g_i(z') - g_i(z)) = \frac{1}{N} \sum_{i=1}^{N} g_i(z)(G_{ii}(z') - g_i(z'))
\]

\[
+ \frac{1}{|z'-z|} \frac{1}{N} \sum_{i=1}^{N} g_i(z)(G_{ii}(z) - g_i(z)) = O_\prec \left( \frac{\Theta(z) + \Theta(z')}{|\text{Im}z|} \right).
\]

**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 \( |\text{Im}z - \text{Im}z'| \geq \frac{1}{2} |\text{Im}z| \), then we can use the same argument as in Case 1. Thus it is sufficient to study when \( |\text{Im}z - \text{Im}z'| \leq \frac{1}{2} |\text{Im}z| \), which means \( \frac{1}{2} |\text{Im}z| \leq |\text{Im}z| \leq 2 |\text{Im}z'| \). Note that

\[
F(z, z') - \frac{1}{z'-z} \frac{1}{N} \sum_{i=1}^{N} g_i(z)(g_i(z') - g_i(z)) \leq \left| \frac{1}{N} \sum_{i=1}^{N} (g_i(z) - g_i(z'))(G_{ii}(z') - g_i(z')) \right| \frac{1}{z-z'}
\]

\[
+ \frac{1}{N} \sum_{i=1}^{N} g_i(z')(G_{ii}(z') - g_i(z')) \leq \frac{1}{N} \sum_{i=1}^{N} g_i(z)(G_{ii}(z) - g_i(z)) \leq \frac{C}{\text{Im}z}.
\]

For the first term on the right side, by direct computation, we get

\[
\left| \frac{g_i(z) - g_i(z')}{z-z'} \right| \leq |g_i(z)||g_i(z')| \left( 1 + \left| \frac{m_{f_c}(z) - m_{f_c}(z')}{z-z'} \right| \right).
\]

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

\[
\left| \frac{m_{f_c}(z) - m_{f_c}(z')}{z-z'} \right| \leq \sup_{\omega \in L(z, z')} |m_{f_c}(\omega)| \leq \frac{C}{|\text{Im}z|}.
\]

Combining with (4.3), we have

\[
|g_i(z) - g_i(z')| \leq \frac{C'}{|\text{Im}z|}.
\]

Using the anisotropic local law and the same argument as in (5.7), we obtain that the first term is bounded as \( O_\prec \left( \frac{\Theta(z')}{|\text{Im}z|} \right) \).

For the second term, we write it as \( h(z) - h(z') \frac{1}{z-z'} \), where

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

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

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

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

Therefore, in both cases, we have

\[
F(z, z') = \frac{1}{z'-z} \left( \frac{1}{N} \sum_{i=1}^{N} g_i(z)g_i(z') - \frac{1}{N} \sum_{i=1}^{N} g_i^2(z) \right) + O_\prec \left( \frac{\Theta(z)}{|\text{Im}z|} \right) + O_\prec \left( \frac{\Theta(z')}{|\text{Im}z|} \right).
\]

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

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

Then by using Lemma 3.2, we have

\[
\sum_{i=1}^{N} \frac{A_{ii}(z)}{z - a_i + m_{f_c}} = \frac{2i\lambda}{\pi} \mathbb{E}[\epsilon_0(\lambda)] \int_{\partial\Omega} \frac{\partial}{\partial z} \tilde{f}(z') \frac{\partial}{\partial z'} \left( \frac{1}{1 - I(z, z')} \frac{\partial I(z, z')}{\partial z} (1 - I_s(z)) \right) d^2z'.
\]
Using Lemma 4.4 and the stability bound (4.3), we have

\[ +O_{\prec}\left(\frac{\theta(z)}{\text{Im} z}\right) + O_{\prec}\left(\frac{1}{N\theta_0\text{Im} z}\right) = O_{\prec}\left(\frac{\theta(z)}{\eta_0}\right). \]

Similarly, plugging (5.1) in (5.9), we have

\[ \sum_{i=1}^{N} \frac{A_{42}(i)}{z - a_i + m_{fc}} = \frac{(m_2 - 2)i\lambda}{\pi} E[e_0(\lambda)] \int_{\Omega_0} \frac{\partial}{\partial z'} \tilde{f}(z') \frac{\partial}{\partial z'} \left( \frac{\partial I(z, z')}{\partial z}(1 - I_\circ(z)) \right) d^2z' + O_{\prec}\left(\frac{1}{\sqrt{N\eta_0}}\right). \]

Finally, plugging (5.2) in the leading term of (5.11) we have

\[ \sum_{i=1}^{N} \frac{D_2(i)}{z - a_i + m_{fc}} \frac{(W_4 - 3)i\lambda}{\pi} E[e_0(\lambda)] \int_{\Omega_0} \frac{\partial}{\partial z'} \tilde{f}(z') \frac{\partial}{\partial z'} \left( \frac{\partial I(z, z')}{\partial z}(1 - I_\circ(z))I(z, z') \right) + O_{\prec}\left(\frac{1 + |\lambda|^2}{\sqrt{N\eta_0}}\right). \]

Therefore, we have

\[ (1 - I_\circ(z))E[e_0(\lambda)(\text{Tr} G - \text{ETr} G)] = (1 - I_\circ(z))\frac{i\lambda}{\pi} E[e_0(\lambda)] \int_{\Omega_0} \frac{\partial}{\partial z'} \tilde{f}(z')K(z, z')d^2z' + \varepsilon_2, \]

where

\[ K(z, z') = \frac{\partial}{\partial z'} \left( 2\frac{\partial I(z, z')}{\partial z} - \frac{1}{1 - I(z, z')} \right) + (m_2 - 2)\frac{\partial I(z, z')}{\partial z} + (W_4 - 3)I(z, z')\frac{\partial I(z, z')}{\partial z}, \]

and

\[ \varepsilon_2 = (1 + |\lambda|^2)\left[ O_{\prec}\left( N\Theta^2(z) \right) + O_{\prec}\left( (\sqrt{N}\Psi^2(z) \right) + O_{\prec}\left( \frac{\Psi(z)}{\eta_0} \right) + O_{\prec}\left( \frac{\Theta(z)}{\eta_0} \right) + O_{\prec}\left( \frac{1}{\sqrt{N\eta_0}} \right) \right]. \]

Dividing both sides by \(1 - I_\circ(z)\), recalling from Lemma 4.4 that \(\frac{\tau}{1 - I_\circ(z)} \sim \frac{1}{\sqrt{N\eta_0\text{Im} z}}\), and using (4.1), (4.2), we have completed the proof of Lemma 3.4. \(\square\)

### 6. Proof of Theorem 2.11 and Theorem 2.12

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 (2.17) with \(f\) given in (2.12).

**Lemma 6.1.** Under the assumptions and notations of Theorem 2.9, we have

\[ V(f) = \frac{1}{2\beta \pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2)^2} dx_1 dx_2 + O(N^{-2}) + O_{\prec}(\eta_0). \]

Assuming that we have proved the above lemma, \(V(f)\) converges to some positive constant since \(g \in C^2_c(\mathbb{R})\). Theorem 2.11 is a direct result of Proposition 2.8 after integrating \(\phi'(\lambda)\) and using the Levy continuity theorem.

**Proof of Lemma 6.1.** Recall that

\[ V(f) = \frac{1}{\pi^2} \int_{\Omega_0} \int_{\eta_0} \frac{\partial}{\partial z_1} \tilde{f}(z_1) \frac{\partial}{\partial z_2} \tilde{f}(z_2)(K_1 + K_2 + K_3)d^2z_1d^2z_2, \]

where

\[ K_1 = \left( m_2 - \frac{2}{\beta} \right) \frac{\partial^2}{\partial z_1 \partial z_2} I; \quad K_2 = \left( W_4 - 1 - \frac{2}{\beta} \right) \left( I - \frac{\partial^2}{\partial z_1 \partial z_2} I + \frac{\partial}{\partial z_1} I \frac{\partial}{\partial z_2} I \right); \]

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

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

\[ \frac{\partial}{\partial z_1} I(z_1, z_2) = \frac{1}{N} \sum_{i=1}^{N} \frac{1 + m_{fc}(z_1)}{a_i - z_1 - m_{fc}(z_1)(a_i - z_2 - m_{fc}(z_2))} = O_{\prec}\left( \frac{1}{\sqrt{\kappa_1 + \eta_1}} \right); \]

\[ \frac{\partial}{\partial z_2} I(z_1, z_2) = \frac{1}{N} \sum_{i=1}^{N} \frac{1 + m_{fc}(z_2)}{a_i - z_1 - m_{fc}(z_1)(a_i - z_2 - m_{fc}(z_2))} = O_{\prec}\left( \frac{1}{\sqrt{\kappa_2 + \eta_2}} \right); \]

and

\[ \frac{\partial^2}{\partial z_1 \partial z_2} I(z_1, z_2) = \frac{1 + m_{fc}(z_1)}{N} \sum_{i=1}^{N} \frac{1}{(a_i - z_1 - m_{fc}(z_1)(a_i - z_2 - m_{fc}(z_2)))^2} = O_{\prec}\left( \frac{1}{\sqrt{\kappa_1 + \eta_1}} \right). \]
\[
\frac{\partial^2}{\partial z_1 \partial z_2} I(z_1, z_2) = \frac{1}{N} \sum_{i=1}^{N} \frac{(1 + m'_{fc}(z_1))(1 + m'_{fc}(z_2))}{(a_i - z_1 - m_{fc}(z_1))^2(a_i - z_2 - m_{fc}(z_2))^2} = O_{\infty}\left(\frac{1}{\sqrt{(\kappa_1 + \eta_1)(\kappa_2 + \eta_2)}}\right). (6.5)
\]

In addition, recalling (4.8), for \(z_1 \neq z_2\), we have
\[
\frac{1}{I(z_1, z_2)} = 1 + \frac{m_{fc}(z_1) - m_{fc}(z_2)}{z_1 - z_2}. (6.6)
\]

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

If \(z_1, z_2\) belong to different half planes, using Lemma 4.4, then we have
\[
\left|\frac{1}{I(z_1, z_2)}\right| \leq 1 + \left|\frac{m_{fc}(z_1) - m_{fc}(z_2)}{z_1 - z_2}\right| \leq \frac{C}{|z_1 - z_2|} \leq \frac{C}{\eta_1 + |\eta_2|}. (6.8)
\]

Now, we are ready to compute \(V(f)\). Since \(\frac{\partial}{\partial z} K_i(z, z') = \frac{\partial}{\partial z} K_i(z, z') = 0\), \((i = 1, 2, 3)\), and by Stokes’ formula, we have
\[
V(f) = -\frac{1}{4\pi^2} \int_{\Gamma_1} \int_{\Gamma_2} \left(f(x_1) + iy_1 f'(x_1)\right) \left(f(x_2) + iy_2 f'(x_2)\right)(K_1 + K_2 + K_3) dz_1 dz_2 := V_1 + V_2 + V_3,
\]
where \(\Gamma_1 = \{x_1 + iy_1 : |y_1| = N^{-\tau}\eta_0\}\) and \(\Gamma_2 = \{x_2 + iy_2 : |y_2| = \frac{1}{2} N^{-\tau}\eta_0\}\). We choose the orientation of both contours to be counterclockwise. The parts on the upper half plane are denoted as \(\Gamma^+\), while the parts on the lower half plane are \(\Gamma^-\).

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 \(|V_1 + V_2| = O_{\infty}(\eta_0^3)\). It is sufficient to estimate \(K_3\). 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,
\[
\left(\int_{\Gamma^+} \int_{\Gamma^+} \int_{\Gamma^+} \int_{\Gamma^+} \right) \hat{f}(z_1) \hat{f}(z_2) K_3(z_1, z_2) dz_1 dz_2 = O_{\infty}(\eta_0^3).
\]

**Case 2:** Consider \(z_1, z_2\) are in different half planes. For notational simplicity, we define \(m_1 = m_{fc}(z_1)\) and \(m_2 = m_{fc}(z_2)\). Differentiating \(I\) given in (4.8), we have
\[
\frac{\partial}{\partial z_1} I = \frac{(z_1 - z_2)(m_1 - m_1 + m_2)}{(z_1 + m_1 - z_2 - m_2)^2}; \quad \frac{\partial}{\partial z_2} I = \frac{(z_2 - z_1)(m_2 - m_1 + m_2)}{(z_1 + m_1 - z_2 - m_2)^2}. (6.9)
\]

Using (6.8) (6.6), (6.3)-(6.5) and Lemma 4.4, we have
\[
K_3 = \frac{2}{\beta} \left(\frac{1}{z_1 - z_2}\right)^2 \left((z_1 - z_2)(m_1 - m_1 + m_2)((z_2 - z_1) m_2 + m_1 - m_2) + O_{\infty}(\eta_0^{-1})\right).
\]

Note that if \(z \in \mathbb{C}^+\) and in the bulk, then there exists \(k, K > 0\) such that \(k \leq \text{Im} m_{fc}(z) \leq K\). 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
\[
K_3 = \frac{2}{\beta} \left(\frac{m_1 - m_2}{z_1 - z_2}\right)^2 O_{\infty}(\eta_0^{-1}) + O(1) = \frac{2}{\beta} \left(\frac{1}{z_1 - z_2}\right)^2 + O_{\infty}(\eta_0^{-1}).
\]

Therefore, by symmetry and (3.1) we have
\[
V_3 = \frac{1}{2\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{g(x_1) + iN^{-\tau}\eta_0 f'(x_1)\right)(g(x_2) + iN^{-\tau}\eta_0 f'(x_2)) (x_1 - x_2 + 2iN^{-\tau}\eta_0)^2 dx_1 dx_2 + O_{\infty}(\eta_0).
\]

Changing the variable
\[
x_1' = \frac{x_1 - E_0}{\eta_0}; \quad x_2' = \frac{x_2 - E_0}{\eta_0}, (6.10)
\]
we have
\[
V(f) = -\frac{1}{2\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{g(x_1') + iN^{-\tau}g'(x_1')}\right)(g(x_2) + iN^{-\tau}g'(x_2)) (x_1 - x_2 + 2iN^{-\tau}) dx_1 dx_2 + O_{\infty}(\eta_0)
\]
Using (3.1), one can show by direct computation, we have
\[ \int_R \int_R (g(x_1) - g(x_2))^2 dx_1 dx_2 = \int_R \int_R \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2 + 2iN^{-\gamma})^2} dx_1 dx_2 + O(1) \]
Note that
\[ \int_R \int_R (g(x_1) - g(x_2))^2 dx_1 dx_2 = \left( \int_{|x_1 - x_2| \geq N^{-\gamma}} + \int_{|x_1 - x_2| \leq N^{-\gamma}} \right) \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2 + 2iN^{-\gamma})^2} dx_1 dx_2. \]
Since the integrand is uniformly bounded, the second integral is then \( O(N^{-\gamma}) \). For the first integral, we have
\[ \int_{|x_1 - x_2| \geq N^{-\gamma}} \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2 + 2iN^{-\gamma})^2} dx_1 dx_2 = \int_{|x_1 - x_2| \geq N^{-\gamma}} \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2)^2(1 + iO(N^{-\gamma}))^2} dx_1 dx_2 \]
Thus we complete the proof.

6.2. Near the edge. Theorem 2.12 is a result of the following lemma and Proposition 2.8.

**Lemma 6.2.** Under the assumptions and notations of Theorem 2.9, we have
\[ \lim_{N \to \infty} V(f_N) = \left. \frac{1}{2\beta \pi} \int_R \int_R \frac{(g(-x^2) - g(-y^2))^2}{x - y} \right|^{x=0}_{y=0}. \]

**Proof of Lemma 6.2.** Similarly as in the bulk, using Stokes’ formula, we have
\[ V(f) = -\frac{1}{4\pi^2} \int_{\Gamma_1} \int_{\Gamma_2} \tilde{f}(z_1) \tilde{f}(z_2) (K_1 + K_2 + K_3) dz_1 dz_2 := V_1 + V_2 + V_3, \]
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
\[ |K_1 + K_2| = O \left( \frac{1}{\sqrt{\Im(z_1) \Im(z_2)}} \right) = O(N^{-\gamma} \eta_0^{-1}). \]
Using (3.1), one can show \( |V_1 + V_2| = O_\sim(\eta_0) \). Thus it is sufficient to study the integral involved \( K_3 \).

Using (6.9), (6.6), and
\[ \frac{\partial^2}{\partial z \partial z'} I(z, z') = \frac{\partial^2}{\partial z' \partial z} I(z, z') = \frac{(m'_1 + m'_2 + 2m'_1 m'_2)(z_1 - z_2) - (m'_1 + m'_2 + 2)(m_1 - m_2)}{(z_1 + m_1 - z_2 - m_2)^3}, \]
by direct computation, we have
\[ K_3 = \frac{2}{\beta} \left( \frac{(1 + m'_1)(1 + m'_2)}{(z_1 + m_1 - z_2 - m_2)^2} - \frac{1}{(z_1 + m_1 - z_2 - m_2)^2} \right). \]

For the second integrand, using similar argument as in the previous subsection, we have
\[ \int_{\Gamma_1} \int_{\Gamma_2} \tilde{f}(z_1) \tilde{f}(z_2) dz_1 dz_2 = -\frac{1}{2} \int_{\Gamma_1} \int_{\Gamma_2} (\tilde{f}(z_1) - \tilde{f}(z_2))^2 / (z_1 - z_2)^2 dz_1 dz_2 \]
\[ = -\frac{1}{2} \int_R \int_R \frac{(g(x_1) - g(x_2) + iN^{-\gamma}(g'(x_1) - g'(x_2)))^2}{(x_1 - x_2 + \frac{i}{2}N^{-\gamma})^2} dx_1 dx_2 = -\frac{1}{2} \int_R \int_R \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2)^2} dx_1 dx_2 + O(N^{-\gamma}). \]
We treat the integrals along \( \Gamma_1, \Gamma_2 \) similarly, due to the opposite integral direction, we have
\[ \int_{\Gamma_1} \int_{\Gamma_2} \tilde{f}(z_1) \tilde{f}(z_2) dz_1 dz_2 = \frac{1}{2} \int_R \int_R \frac{(g(x_1) - g(x_2))^2}{(x_1 - x_2)^2} dx_1 dx_2 + O(N^{-\gamma}). \]
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,
\[ V_3(f) = -\frac{1}{2\beta \pi^2} \left( \int_{\Gamma_1} \int_{\Gamma_2} + \int_{\Gamma_1} \int_{\Gamma_2} + \int_{\Gamma_1} \int_{\Gamma_2} + \int_{\Gamma_1} \int_{\Gamma_2} \right) \tilde{f}(z_1) \tilde{f}(z_2) \frac{(1 + m'_1)(1 + m'_2)}{(z_1 + m_1 - z_2 - m_2)^2} dz_1 dz_2 \]
\[ : V_3^{++} + V_3^{-} + V_3^{+} + V_3^{-}. \]

Let \( \zeta = z + m_{fc}(z) \) and \( \zeta_{\pm} = L_{\pm} + m_{fc}(L_{\pm}) \in \mathbb{R} \). Define \( F(\zeta) := \zeta - \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(\zeta - \zeta_1_{\pm})^2} \) so that (2.6) is equivalent to \( z = F(\zeta) \). Assumptions 2.2 and 2.3 imply that \( \zeta \) is analytic in a neighborhood of \( \zeta_{++} \), where \( \zeta_{++} \) is the smallest interval that contains the support of \( \mu_A \); see (4.3). Hence, \( F(\zeta) \) is analytic in a neighborhood of \( \zeta_{++} \), where we write

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

By (2.7), \( F'(\zeta_{++}) = 1 - \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(\zeta - \zeta_1_{++})^2} = 0 \). Moreover, \( F''(\zeta_{++}) = -\frac{2}{N} \sum_{i=1}^{N} \frac{1}{(\zeta - \zeta_1_{++})^3} \), and by (4.5) it is bounded uniformly from below. In general, we have \( |F^{(k)}(\zeta_{++})| = \frac{\sqrt{1}\sum_{i=1}^{N} \frac{1}{(\zeta - \zeta_1_{++})^{k+1}}| = O(1) \) of (4.4). Inverting \( F(\zeta) = z \) in the neighborhood of \( \zeta_{++} \), we have the expansion

\[ \zeta = z + m_{fc}(z) = \zeta_{++} + c_{\pm} \sqrt{z - L_{\pm}} \left( 1 + A_{\pm}(\sqrt{z - L_{\pm}}) \right), \]

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

\[ 1 + m_{fc}'(z) = \frac{c_{\pm}}{2 \sqrt{z - L_{\pm}}} + d_+ + O(\sqrt{z - L_{\pm}}), \]

where \( d_+ \) is some number which depends on \( N \) but is uniformly bounded. Let \( z = L_{\pm} + \eta_0(x + iN^{-\tau}) \). Then

\[ z + m_{fc}(z) = \zeta_{++} + c_+ \sqrt{\eta_0(x + iN^{-\tau})} + O(\eta_0); \quad 1 + m_{fc}'(z) = \frac{c_+}{2 \sqrt{\eta_0(x + iN^{-\tau})}} + d_+ + O(\sqrt{\eta_0}). \]

Therefore, after changing the variable as in (6.10), we have

\[ V_3^{++} = -\frac{1}{8\beta \pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\tilde{g}(x_1)\tilde{g}(x_2)}{\sqrt{x_1 + iN^{-\tau}}} + O(\sqrt{\eta_0}) \left( \frac{1}{\sqrt{x_2 + \frac{1}{2} N^{-\tau}} + O(\sqrt{\eta_0})} \right)^2 dx_1 dx_2 \]

where \( \tilde{g}(x) = g(x) + iN^{-\tau}g'(x) \). The last step follows from the fact that \( |\sqrt{x_1 + iN^{-\tau}} - \sqrt{x_2 + \frac{1}{2} N^{-\tau}} + O(\sqrt{\eta_0})| \geq N^{-\sigma} \), when \( x_1, x_2 \) belong to some compact set. Let \( \gamma_1^+ := \{ x_1 + i N^{-\tau} : x_1 \in \mathbb{R} \} \) and \( \gamma_2^+ := \{ x_2 + \frac{1}{2} N^{-\tau} : x_2 \in \mathbb{R} \} \). Then we obtain

\[ V_3^{++} = -\frac{1}{8\beta \pi^2} \int_{\gamma_1^+} \int_{\gamma_2^+} \frac{\tilde{g}(z_1)\tilde{g}(z_2)}{\sqrt{z_1 + iN^{-\tau}} + O(\sqrt{\eta_0})} d\gamma_1 d\gamma_2 + O(\sqrt{\eta_0}), \]

where \( \tilde{g}(x + iy) = g(x) + ig'(x)\chi(y) \). Since \( \gamma_1^+ \) and \( \gamma_2^+ \) are disjoint, for any fixed \( z_2 \),

\[ \int_{\gamma_1^+} \int_{\gamma_2^+} \frac{1}{\sqrt{z_1 + iN^{-\tau}} - \sqrt{z_2 + \frac{1}{2} N^{-\tau}}} d\gamma_1 d\gamma_2 = 0, \]

and thus

\[ V_3^{++} = \frac{1}{16\beta \pi^2} \int_{\gamma_1^+} \int_{\gamma_2^+} \frac{(\tilde{g}(z_1) - \tilde{g}(z_2))^2}{\sqrt{z_1 + iN^{-\tau}} - \sqrt{z_2 + \frac{1}{2} N^{-\tau}}} d\gamma_1 d\gamma_2 + O(\sqrt{\eta_0}). \]
Therefore, we get
\[
\lim_{N \to \infty} V_{3}^{++} = \frac{1}{16\beta \pi^2} \lim_{N \to \infty} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(g(x_1) - g(x_2) + iN^{-\tau}g'(x_1) - \frac{i}{2}N^{-\tau}g'(x_2))^2}{\sqrt{x_1 + iN^{-\tau}} \sqrt{x_2 + \frac{i}{2}N^{-\tau}} - \sqrt{x_2 + \frac{i}{2}N^{-\tau}}} \, dx_1 \, dx_2.
\]
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
\[
\sqrt{x_1 + iN^{-\tau}} - \sqrt{x_2 + \frac{i}{2}N^{-\tau}} \geq C |\sqrt{x_1} - \sqrt{x_2}|.
\]
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.

Case 1: If \( x_1, x_2 \) have the same sign, then
\[
h(x_1, x_2) = \frac{(g(x_1) - g(x_2))^2 + (g'(x_1) - g'(x_2))^2}{\sqrt{|x_1| \sqrt{|x_2|} \sqrt{|x_1| - \sqrt{x_2}^2}}}
\]
\[
\leq \frac{1}{8M} \frac{|g'|^2_{\infty} + |g''|^2_{\infty}}{\sqrt{|x_1| \sqrt{|x_2|} \sqrt{|x_1| - \sqrt{x_2}^2}}).
\]

Case 2: If \( x_1 \) and \( x_2 \) are of opposite signs, using \(|x_1 - x_2| = (\sqrt{|x_1| - i \sqrt{x_2}})(\sqrt{|x_1| + i \sqrt{x_2}})\), we have
\[
h(x_1, x_2) = \frac{(g(x_1) - g(x_2))^2 + (g'(x_1) - g'(x_2))^2}{\sqrt{|x_1| \sqrt{|x_2|} \sqrt{|x_1| - i \sqrt{x_2}}}^2}
\]
\[
\leq \frac{1}{8M} \frac{|g'|^2_{\infty} + |g''|^2_{\infty}}{\sqrt{|x_1| \sqrt{|x_2|} \sqrt{|x_1| - i \sqrt{x_2}}}^2}.
\]

If \( x_1 \notin [-2M, 2M] \), then \( x_2 \in [-M, M] \) otherwise \( h(x_1, x_2) = 0 \). So for \((x_1, x_2) \in [-2M, 2M]^c \times [-2M, 2M] \),
\[
h(x_1, x_2) \leq \frac{4|g|^2_{\infty} + 4|g'|^2_{\infty}}{\sqrt{|x_1| \sqrt{|x_2|} \sqrt{|x_1| - \sqrt{x_2}^2}^2}} \leq \frac{4|g|^2_{\infty} + 4|g'|^2_{\infty}}{\sqrt{|x_1| \sqrt{|x_2|} \sqrt{|x_1| - \sqrt{x_2}^2}}} = \frac{C}{|x_1|^{3/2}|x_2|^{1/2}}.
\]
Therefore, \( h(x_1, x_2) \) is integrable. Thus by dominated convergence,
\[
\lim_{N \to \infty} V_{3}^{++} = \frac{1}{16\beta \pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(g(x_1) - g(x_2))^2}{\sqrt{x_1 + i0 \sqrt{x_2} + i0(\sqrt{x_1} + i0 - \sqrt{x_2} + i0)^2}} \, dx_1 \, dx_2
\]
\[
= \frac{1}{4\beta \pi^2} \int_{\phi(R+i0)} \int_{\phi(R+i0)} \frac{(g(w_1^2) - g(w_2^2))^2}{(w_1 - w_2)^2} \, dw_1 \, dw_2,
\]
where we change the variable \( \phi : z \to \sqrt{z} \) (with branch cut such that \( \phi : \mathbb{C}^+ \to \mathbb{C}^+ \)).

Similarly, we have
\[
\lim_{N \to \infty} V_{3}^{--} = \frac{1}{4\beta \pi^2} \int_{\phi(R-i0)} \int_{\phi(R-i0)} \frac{(g(w_1^2) - g(w_2^2))^2}{(w_1 - w_2)^2} \, dw_1 \, dw_2;
\]
\[
\lim_{N \to \infty} V_{3}^{--} = \frac{1}{4\beta \pi^2} \int_{\phi(R+i0)} \int_{\phi(R+i0)} \frac{(g(w_1^2) - g(w_2^2))^2}{(w_1 - w_2)^2} \, dw_1 \, dw_2;
\]
\[
\lim_{N \to \infty} V_{3}^{+-} = \frac{1}{4\beta \pi^2} \int_{\phi(R-i0)} \int_{\phi(R+i0)} \frac{(g(w_1^2) - g(w_2^2))^2}{(w_1 - w_2)^2} \, dw_1 \, dw_2.
\]
The contours are shown in Figure 1. Note that the horizontal parts of the blue and the red lines of
We treat the expectation similarly using the cumulant expansion and (5.4):

\begin{equation*}
\text{Proof of Proposition 2.10.} \text{ We treat the expectation similarly using the cumulant expansion and (5.4):}
\end{equation*}

\begin{align*}
&\lim_{N \to \infty} V_3 = \frac{1}{4\beta \pi^2} \int_{-i\infty}^{i\infty} \int_{-i\infty}^{i\infty} \frac{(g(w_1^2) - g(w_2^2))^2}{(w_1 - w_2)^2} \, dw_1 \, dw_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 dx_1 \, dx_2.
\end{align*}

This concludes the proof of Theorem 2.12. \qed

7. PROOF OF PROPOSITION 2.10 AND COMPUTATION OF THE BIAS

In this section, we first prove Proposition 2.10, using the same technique as in Proposition 2.8. After this, we compute the bias on mesoscopic scales inside the bulk and at the edges.

**Proof of Proposition 2.10.** We treat the expectation similarly using the cumulant expansion and (5.4):

\begin{equation*}
(z - a_i) \mathbb{E} G_{ii} = \mathbb{E}(H G)_{ii} - 1 = \frac{1}{N^2} \mathbb{E} \sum_{j=1}^{N} c_{ij}^{(2)} \frac{\partial G_{jj}}{\partial H_{ij}} - 1 + \frac{1}{2N^2} \sum_{j=1}^{N} c_{ij}^{(3)} \mathbb{E} \frac{\partial^2 G_{ii}}{\partial H_{ij}^2} + \frac{1}{3N^2} \sum_{j=1}^{N} c_{ij}^{(4)} \mathbb{E} \frac{\partial^3 G_{jj}}{\partial H_{ij}^3} + O_N(N^{-\frac{2}{3}})
\end{equation*}

\begin{align*}
&= -\frac{1}{N} \sum_{j=1}^{N} \mathbb{E} G_{ii} G_{jj} - \frac{1}{N} \mathbb{E}(G^2)_{ii} - \frac{m_2 - 2}{N} \mathbb{E}(G_{ii})^2 - 1 + \frac{1}{2N^2} \sum_{j=1}^{N} c_{ij}^{(3)} (6 G_{ii} G_{jj} G_{jj} + 2 G_{ij}^3)
\end{align*}

\begin{align*}
&+ \frac{1}{6N^2} \sum_{j=1}^{N} (W_4 - 3) (-36 G_{ii} G_{jj} G_{jj}^2 - 6 G_{ii}^2 G_{jj} - 6 G_{ij}^4) + O_N(N^{-\frac{2}{3}}).
\end{align*}

Combining with the local law, we have

\begin{equation*}
(z - a_i) \mathbb{E} G_{ii} = -\frac{1}{N} \mathbb{E} G_{ii} \text{Tr} G - \frac{1}{N} \frac{d}{dx} a_i - z - m_{fc} - \frac{m_2 - 2}{N} \frac{1}{(a_i - z - m_{fc})^2} - 1
\end{equation*}

\begin{align*}
&+ \frac{3}{N^2} \sum_{j=1}^{N} \frac{c_{ij}^{(3)} G_{ij}}{N} (a_i - z - m_{fc})(a_j - z - m_{fc}) G_{ij} - \frac{1}{N^2} \sum_{j=1}^{N} (W_4 - 3) (a_i - z - m_{fc})^2 (a_j - z - m_{fc})^2 + O_N(N^{-\frac{1}{2}})(\Psi).
\end{align*}

Using the anisotropic local law and the argument as in (5.10), one can show that the second term of the last line of above equation is $O_N(N^{-1} \Psi)$. Therefore, we have

\begin{equation*}
(z - a_i) \mathbb{E} G_{ii} = -\frac{1}{N} \mathbb{E} \left( G_{ii} - \frac{1}{a_i - z - m_{fc}} \right) \text{Tr} G - \frac{1}{N} \frac{1}{a_i - z - m_{fc}} \frac{1}{E} \text{Tr} G - 1 - \frac{1}{N} \frac{1 + m_{fc}(z)}{(a_i - z - m_{fc})^2}
\end{equation*}
and thus
\[(z - a_i + m_{fc}) \left( \mathbb{E}G_{ii} - \frac{1}{a_i - z - m_{fc}} \right) = -\frac{1}{N} a_i - z - m_{fc} \left( \mathbb{ETr} G - N m_{fc} \right) \]
\[-\frac{1}{N} \frac{1 + m'_{fc}(z)}{(a_i - z - m_{fc})^2} - \frac{m_2 - 2}{N} \frac{1}{(a_i - z - m_{fc})^2} = -\frac{1}{N} I_s(z) \left( \frac{W_4 - 3}{(a_i - z - m_{fc})^2} + O_{\asymp} \left( \frac{\Psi}{N \eta} \right) \right).\]

Dividing both sides by \( a_i - z - m_{fc} \), we obtain
\[
(1 - I_s(z)) \mathbb{E}(\text{Tr} G - N m_{fc}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1 + m'_{fc}(z)}{(a_i - z - m_{fc})^2} + \frac{m_2 - 2}{N} \sum_{i=1}^{N} \frac{1}{(a_i - z - m_{fc})^3}
+ \frac{W_4 - 3}{N} \sum_{i=1}^{N} \frac{I_s(z)}{(a_i - z - m_{fc})^3} + O_{\asymp} \left( \frac{\Psi}{\eta} \right).
\]

Dividing both sides by \( 1 - I_s(z) \) and using the relation \( 1 - I_s(z) = \frac{1}{1 + m'_{fc}(z)} \sim \sqrt{\kappa + \eta} \), we obtain
\[
\mathbb{E}(\text{Tr} G - N m_{fc}) = \frac{1}{1 - I_s(z)} \frac{dI_s(z)}{dz} + \frac{m_2 - 2}{2} \frac{dI_s(z)}{dz} + \frac{W_4 - 3}{2} I_s(z) \frac{dI_s(z)}{dz} + O_{\asymp} \left( \frac{1}{\eta \sqrt{\kappa + \eta}} \right).
\]

Plugging into (3.6) (here we replace \( \mathbb{E} \mu_N \) by \( \mu_{fc} \)), using Lemma 3.2 and Stokes’ formula, we have
\[
\text{ETrf}(X_N) - N \int_{\mathbb{R}} f(x) \rho_{fc}(x) dx = \frac{1}{4 \pi I} \int_{\partial \Omega_0} (\tilde{f}(z) b(z) dz + O_{\asymp} \left( \frac{1}{\sqrt{N \eta \Phi + \eta \Phi}} \right) + O_{\asymp} (N^{-\tau}),
\]
where
\[
b(z) = \frac{1}{1 - I_s(z)} \frac{dI_s(z)}{dz} + (m_2 - 2) \frac{dI_s(z)}{dz} + (W_4 - 3) I_s(z) \frac{dI_s(z)}{dz}.
\]

Using the relation \( I_s = \frac{m'_{fc}}{1 + m'_{fc}} \), it coincides with the expectation that obtained in the global CLT given in Theorem 2.7. \( \square \)

Next, we explicitly compute the bias in the bulk and at the edges, for the scaled test function in (2.12).

### 7.1. Bias in the mesoscopic bulk

Note that
\[
\frac{dI_s}{dz} = \frac{2}{N} \sum_{i=1}^{N} \frac{1 + m'_{fc}}{(a_i - z - m_{fc})^3} = O \left( \frac{1}{\sqrt{\kappa + \eta}} \right); \quad 1 - I_s(z) \sim \frac{1}{\sqrt{\kappa + \eta}}; \quad | I_s(z) | = O(1). \tag{7.2}
\]

If \( \kappa \geq \kappa_0 > c > 0 \), then \( | b(z) | = O(1) \). In combination with (3.1), we have
\[
\text{ETrf}(X_N) - N \int_{\mathbb{R}} f(x) \rho_{fc}(x) dx = \frac{1}{4 \pi I} \int_{\partial \Omega_0} (f(x) + ig'f(x)) b(z) dz = O_{\asymp} (\eta_0) + O_{\asymp} (N^{-\tau}),
\]

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

### 7.2. Bias at the mesoscopic edge

Similarly, using (7.2) and (3.1), the last two terms of \( b(z) \) will contribute \( O_{\asymp} (\sqrt{\eta_0}) \). We have
\[
\text{ETrf}(X_N) - N \int_{\mathbb{R}} f(x) \rho_{fc}(x) dx = \frac{1}{4 \pi I} \int_{\partial \Omega_0} \tilde{f}(z) \frac{m''_{fc}}{1 + m'_{fc}} dz + O_{\asymp} \left( N^{-\tau} + \frac{1}{\sqrt{N \eta_0 \Phi + \eta \Phi}} + \sqrt{\eta_0} \right).
\]

Using (6.11), we obtain the following expansions:
\[1 + m'_{fc}(z) = \frac{c_+}{2 \sqrt{z - L_+}} + O(1), \quad m''_{fc}(z) = -\frac{c_+}{4 \sqrt{(z - L_+)^3}} + O \left( \frac{1}{\sqrt{|z - L_+|}} \right),\]

and then
\[
\frac{m''_{fc}}{1 + m'_{fc}} = -\frac{1}{2 (z - L_+)} + O \left( \frac{1}{\sqrt{|z - L_+|}} \right).
\]
Changing variables and using (3.1), we have

$$
\mathbb{E}T(r f(x) - N \int f(x) \rho_f(x) dx) = - \frac{1}{8\pi i} \int_{\mathbb{R}} \left( g(x) + iN^{-\tau} g'(x) \right) \frac{1}{x + iN^{-\tau}} dx
$$

$$
+ \frac{1}{8\pi i} \int_{\mathbb{R}} \left( g(x) - iN^{-\tau} g'(x) \right) \frac{1}{x - iN^{-\tau}} dx + O_{<} \left( N^{-\tau} + \frac{1}{\sqrt{N\eta_0\sqrt{\kappa_0} + \eta_0}} \right)
$$

$$
= - \frac{1}{8\pi i} \int_{\mathbb{R}} \frac{g(x)}{x + iN^{-\tau}} dx + \frac{1}{8\pi i} \int_{\mathbb{R}} \frac{g(x)}{x - iN^{-\tau}} dx + O_{<} \left( N^{-\tau} + \frac{1}{\sqrt{N\eta_0\sqrt{\kappa_0} + \eta_0}} \right).
$$

Using the Sokhotski-Plemelj lemma, we have

$$
\mathbb{E}T(r f(x) - N \int f(x) \rho_f(x) dx) = \frac{g(0)}{4} + O_{<} \left( N^{-\tau} + \frac{1}{\sqrt{N\eta_0\sqrt{\kappa_0} + \eta_0}} \right),
$$

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

**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

$$
c_{p,q} := (-1)^{p+q} \left( \frac{\partial^{p+q}}{\partial s^p \partial \bar{s}^q} \log \mathbb{E}e^{sh + u\bar{s}} \right).
$$

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

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

where the error term satisfies

$$
|R_{l+1}| \leq C_l \mathbb{E}|h|^{l+2} \max_{p+q=l+1} \left\{ \sup_{||z|| \leq M} |f^{(p,q)}(z, \bar{z})| \right\} + C_l \mathbb{E} \left[ |h|^{l+2} \right] \max_{p+q=l+1} \|f^{(p,q)}(z, \bar{z})\|_{\infty},
$$

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

Instead of (5.4), we have

$$
\frac{\partial G_{ij}}{\partial H_{ab}} = -G_{ia} G_{bj},
$$

(A.1)

from which we obtain the analogue of Lemma 5.1.

The assumption $\mathbb{E}H_{ij}^2 = 0$ implies that $c^{(1,1)}_{ij} = 1$, $c^{(2,2)}_{ij} = W_4 - 2$ for $i \neq j$. 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

$$
(z - a_i) \mathbb{E}e_0(\lambda)(G_{ii} - \mathbb{E}G_{ii}) = \frac{1}{N} \sum_{j=1}^{N} c^{(1,1)}_{ij} \mathbb{E} \left[ \frac{\partial}{\partial H_{ji}} (e_0(\lambda)(G_{ji} - \mathbb{E}G_{ji})) \right]
$$

$$
+ \frac{1}{2N^2} \sum_{j=1}^{N} c^{(2,2)}_{ij} \mathbb{E} \left[ \frac{\partial^2}{\partial^2 H_{ji} \partial H_{ij}} (e_0(\lambda)(G_{ji} - \mathbb{E}G_{ji})) \right] + \cdots
$$

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

$$
+ \frac{1}{N^2} \sum_{j=1}^{N} (W_4 - 2) \mathbb{E} \left[ \frac{\partial e_0(\lambda)}{\partial H_{ji} \partial H_{ij} \partial H_{ji}} \right] + \cdots.
$$
Thus Proposition 2.8 holds with modified variance, i.e. \( m_2 - 2 \) be replaced by \( m_2 - 1 \), \( W_4 - 2 \) be replaced by \( W_4 - 2 \), and the coefficient of the remaining term be 1 instead of 2. Similarly, as for the expectation,

\[
(z - a_i)E_{ii} = E(WG)_{ii} - 1 = \frac{1}{N} \sum_{j=1}^{N} c_{ij}^{(1,1)} E \frac{\partial G_{ii}}{\partial H_{jj}} - 1 + \frac{1}{2N^2} \sum_{j=1}^{N} c_{ij}^{(2,2)} E \frac{\partial^3 G_{ii}}{\partial^2 H_{jj} \partial H_{ij}} + \cdots .
\]

Thus the first term of \( b(z) \) given in (7.1) 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.** Since \( H(z) \) is holomorphic on \( \Omega_0 \), \( \frac{\partial}{\partial z} \tilde{f}(z)H(z) = \frac{\partial}{\partial z} (\tilde{f}(z)H(z)) \). Using Stokes’ formula, we have

\[
\int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z)H(z) d^2 z = - \frac{i}{2} \int_{\partial \Omega_0} \tilde{f}(z)H(z) dz.
\]

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

\[
\Gamma_0 := \{ x + iy : x \in \text{supp}(f), |y| = N^{-\gamma}\eta_0 \} .
\]

Using (3.1) we have

\[
\left| \int_{\Omega_0} \frac{\partial}{\partial z} \tilde{f}(z)H(z) d^2 z \right| \leq CK \int_{\Gamma_0} (|y|^{-\gamma}|f(x)| + |y|^{-\gamma}|f'(x)|) dz \leq CKN^{-\gamma}\eta_0^{1-\gamma}.
\]

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

\[
m_{fc}(z_1) - m_{fc}(z_2) = \frac{1}{N} \sum_{i=1}^{N} \left( a_i - z_1 - m_{fc}(z_1) - a_i - z_2 - m_{fc}(z_2) \right)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \left( \frac{z_1 + m_{fc}(z_1) - z_2 - m_{fc}(z_2)}{(a_i - z_1 - m_{fc}(z_1))(a_i - z_2 - m_{fc}(z_2))} \right) .
\]

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

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1 + m_{fc}'(z)}{(a_i - z - m_{fc}(z))^2} = m_{fc}'(z) . \tag{B.1}
\]

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

**Proof of Lemma 4.4.** Note that

\[
|I_s(z)| \leq \frac{1}{N} \sum_{i=1}^{N} \frac{1}{|a_i - z - m_{fc}(z)|^2} = \frac{\text{Im} m_{fc}(z)}{\text{Im} m_{fc}(z) + \eta} < 1.
\]

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

\[
|m_{fc}'(z)| \leq \frac{1}{|1 - I_s(z)|} \sim \frac{1}{\sqrt{\kappa + \eta}} . \tag{B.2}
\]

Differentiating (B.1) again, we obtain that

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

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 (5.4), we have
\[ \frac{\partial c_0(\lambda)}{\partial H_{ij}} = \frac{i\lambda}{\pi} \epsilon_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) \left( \sum_{l=1}^{N} \frac{\partial G_{il}}{\partial H_{lj}} \right) d^2z = -\frac{i(2 - \delta_{ij})\lambda}{\pi} \epsilon_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z)(G^2)_{ji} d^2z. \]
Note that $(G^2)_{ji} = \frac{1}{2} G_{ij}$. Since $G_{ij}$ is analytic in $D^*$, using the Cauchy integral formula and the local law, we have that for $i \neq j$, $(G^2)_{ji} \sim \frac{\Psi(z)}{1 + |z|^2}$. Combining with Lemma 3.2, we obtain that, for $i \neq j$,
\[ \left| \frac{\partial c_0(\lambda)}{\partial H_{ij}} \right| = O_{\prec} \left( \frac{1 + |\lambda|}{\sqrt{N\eta_0}} \right). \]
Similarly, if $i = j$, we have
\[ \frac{\partial c_0(\lambda)}{\partial H_{ii}} = -\frac{i\lambda}{\pi} \epsilon_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) \frac{1}{a_i - z - m_{f_i}(z)} d^2z + O_{\prec} \left( \frac{1 + |\lambda|}{\sqrt{N\eta_0}} \right). \]
Furthermore, we compute that
\[ \frac{\partial^2 c_0(\lambda)}{\partial^2 H_{ij}} = -\frac{\lambda^2(2 - \delta_{ij})}{\pi^2} \epsilon_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) (2(G^2)_{ji}G_{ij} + (1 - \delta_{ij})(G^2)_{ii}G_{jj} + (1 - \delta_{ij})(G^2)_{jj}G_{ii}) d^2z. \]
For $i \neq j$, combining the local law and Lemma 3.2, we have
\[ \frac{\partial^2 c_0(\lambda)}{\partial^2 H_{ij}} = \frac{2\lambda^2}{\pi} \epsilon_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) \left( G_{ji}G_{ij} + G_{ii}G_{jj} + G_{jj}G_{ii} \right) d^2z + O_{\prec} \left( \frac{(1 + |\lambda|)^2}{N\eta_0} \right) \]
\[ = \frac{2\lambda^2}{\pi} \epsilon_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) \left( \frac{1}{a_i - z - m_{f_i}} \right) d^2z + O_{\prec} \left( \frac{(1 + |\lambda|)^2}{\sqrt{N\eta_0}} \right). \]
Similarly, for $i = j$, we have
\[ \frac{\partial^2 c_0(\lambda)}{\partial^2 H_{ii}} = \frac{i\lambda}{\pi} \epsilon_0(\lambda) \int_{\Omega_0} \frac{\partial}{\partial z} f(z) \left( \frac{1}{a_i - z - m_{f_i}} \right) d^2z + O_{\prec} \left( \frac{(1 + |\lambda|)^2}{\sqrt{N\eta_0}} \right). \]
In general, using the local law, (5.4) and Lemma 3.2, we complete the proof of (5.3).

\[ \square \]

REFERENCES

[1] A. Adhikari, J. Huang, Dyson Brownian Motion for General $\beta$ and Potential at the Edge, Preprint, arXiv:1810.08308, (2018).
[2] N. I. Akhiezer, The classical moment problem: and some related questions in analysis, Hafner Publishing Co., New York, 1965.
[3] J. Alt, L. Erdős, T. Krüger and D. Schröder, Correlated Random Matrices: Band Rigidity and Edge Universality, Preprint, arXiv:1804.07744, (2018).
[4] Z. D. Bai, J. F. Yao, On the convergence of the spectral empirical process of Wigner matrices, Bernoulli 11(6), 1059-1092 (2005).
[5] E. L. Basor and H. Widom, Determinants of Airy Operators and Applications to Random Matrices, J. Stat. Phys. 96, 1-20 (1999).
[6] F. Bekerman and A. Lodhia, Mesoscopic central limit theorem for general $\beta$-ensembles, Ann. Inst. H. Poincare Probab. Statist. 54, 1917-1938 (2018).
[7] P. Biane, On the Free Convolution with a Semi-circular Distribution, Jiduana Univ. Math. J. 46, 705-718 (1997).
[8] P. Biane, Processes with free increments, Math. Z. 227(1), 143-174 (1998).
[9] A. Boutet de Monvel, A. Khorunzhy, Asymptotic distribution of smoothed eigenvalue density. I. Gaussian random matrices, Random Oper. and Stoch. Equ. 7(1), pp 1-22 (1999).
[10] A. Boutet de Monvel, A. Khorunzhy, Asymptotic distribution of smoothed eigenvalue density. II. Wigner random matrices, Random Oper. and Stoch. Equ. 7(2), pp 149-168 (1999).
[11] J. Breuer and M. Duits, Universality of mesoscopic fluctuations for orthogonal polynomial ensembles, Comm. Math. Phys. 342 (2), 491-531 (2016).
[12] M. Capitaine, C. Donati-Martin, D. Feral, and M. Fevrier, Free Convolution with a Semicircular Distribution and Eigenvalues of spiked deformations of Wigner Matrices, Electron. J. Probab. 16(64), 1750-1792 (2011).
[16] S. Dallaporta, M. Fevrier, Fluctuations Of Linear Spectral Statistics Of Deformed Wigner Matrices. Preprint, arXiv:1903.11324.
[17] G. Cipolloni, L. Erdős, T. Krüger, D. Schröder. Cusp universality for random matrices II: the real symmetric case. Preprint arXiv:1811.04055 (2018).
[18] M. Duits, K. Johansson. On mesoscopic equilibrium for linear statistics in Dyson’s Brownian Motion, Mem. Amer. Math. Soc. 255 (1222), (2018).
[19] L. Erdős and A. Knowles. The Altshuler-Shklovskii formulas for random band matrices I: the unimodular case, Comm. Math. Phys. 335, 1365-1416 (2015).
[20] L. Erdős and A. Knowles. The Altshuler-Shklovskii formulas for random band matrices II: the general case, Ann. H. Poincaré 16, 709-799 (2015).
[21] L. Erdős, A. Knowles, H.-T. Yau. Averaging fluctuations in resolvents of random band matrices, Ann. Henri Poincaré 14, 1837-1926 (2013).
[22] L. Erdős, T. Krüger and D. Schröder. Random Matrices with Slow Correlation Decay, Forum of Mathematics, Sigma 7(8), (2019).
[23] L. Erdős and H.-T. Yau. A dynamical approach to random matrix theory, volume 28 of Courant Lecture Notes. American Mathematical Soc., 2017.
[24] L. Erdős, T. Krüger, D. Schröder. Cusp universality for random matrices I: Local law and the complex Hermitian case. Preprint arXiv:1809.03971 (2018).
[25] Y. He. Bulk eigenvalue fluctuations of sparse random matrices, Preprint, arXiv:1904.07140 (2019).
[26] Y. He, A. Knowles. Mesoscopic Eigenvalue Density Correlations of Wigner Matrices. Preprint arXiv: 1808.09436.
[27] Y. He, A. Knowles. Mesoscopic eigenvalue statistics of Wigner matrices, Ann. Appl. Probab. 27(3), 1510-1550 (2017).
[28] J. Huang, B. Landon. Rigidity and a mesoscopic central limit theorem for Dyson Brownian motion for general beta and potentials, Prob. Theory and Related Fields 175(1-2), 209-253 (2019).
[29] H. C. Ji, J. O. Lee. Gaussian fluctuations for linear spectral statistics of deformed Wigner matrices, Preprint, arXiv:1712.00931.
[30] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices, Duke Math. J. 91, 151-204 (1998).
[31] K. Johansson. From Gumbel to Tracy-Widom, Prob. Theory and Related Fields 138(1-2), 75-112 (1998).
[32] A. Knowles, J. Yin, Anisotropic local laws for random matrices, Prob. Theory and Related Fields 169(1-2), 257-352 (2017).
[33] B. Landon, P. Sosoe, Applications of mesoscopic CLTs in random matrix theory, Preprint, arXiv:1811.05915 .
[34] B. Landon, P. Sosoe and H.-T. Yau, Fixed energy universality of Dyson Brownian motion, Advances in Math, 346, pp. 1137-1332. 2019.
[35] J. O. Lee, K. Schnelli, Local deformed semicircle law and complete delocalization for Wigner matrices with random potential, J. Math. Phys 54, 103504 (2013).
[36] J. O. Lee, K. Schnelli, Extremal eigenvalues and eigenvectors of deformed Wigner matrices, Probab. Theory Related Fields 164(1), 165-241 (2016).
[37] J. O. Lee, K. Schnelli, B. Stetler and H.-T. Yau, Bulk universality for deformed Wigner matrices, Ann. Probab. 44(3), 2349-2425 (2016).
[38] A. Lodhia, N. J. Simm. Mesoscopic linear statistics of Wigner matrices, Preprint, arXiv:1503.03533.
[39] A. Lytova, L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries, Ann. Prob. 37(5), 1778-1840 (2009).
[40] C. Min, Y. Chen. Linear Statistics of Random Matrix Ensembles at the Spectrum Edge Associated with the Airy Kernel, Preprint, arXiv:1806.11297v2.
[41] J. A. Mingo, R. Speicher. Second order freeness and fluctuations of random matrices: I. Gaussian and Wishart matrices and cyclic Fock spaces, J. Func. Anal. 235(1), 226-270 (2006).
[42] L. Pastur. The spectrum of random matrices, Teor. Math. Phys 10, 64-74 (1972).
[43] M. Shcherbina. Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices, Z. Mat. Fiz. Anal. Geom. 7(2), 176-192 (2011).
[44] T. Shcherbina, On universality of local bulk regime for the deformed Gaussian unitary ensemble, Math. Phys. Anal. Geom. 5, 396-433 (2009).
[45] T. Shcherbina, On universality of local edge regime for the deformed Gaussian unitary ensemble, J. Stat. Phys. 143, 455-481 (2011).
[46] P. Sosoe, P. Wong. Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices. Advances in Mathematics 249, 37-87 (2013)
[47] D. Voiculescu, Addition of certain non-commuting random variables, J. Funct. Anal. 66 (3), 323-346 (1986).
[48] D. Voiculescu, The analogues of entropy and of Fisher’s information theory in free probability theory, I, Comm. Math. Phys. 155, 71-92 (1993).
[49] E. P. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, Ann. of Math. 62 (2), 548-564 (1952).