

# Anisotropic local laws for random matrices

Antti Knowles\*      Jun Yin†

August 5, 2016

We develop a new method for deriving local laws for a large class of random matrices. It is applicable to many matrix models built from sums and products of deterministic or independent random matrices. In particular, it may be used to obtain local laws for matrix ensembles that are *anisotropic* in the sense that their resolvents are well approximated by deterministic matrices that are not multiples of the identity. For definiteness, we present the method for sample covariance matrices of the form  $Q := TXX^*T^*$ , where  $T$  is deterministic and  $X$  is random with independent entries. We prove that with high probability the resolvent of  $Q$  is close to a deterministic matrix, with an optimal error bound and down to optimal spectral scales.

As an application, we prove the edge universality of  $Q$  by establishing the Tracy-Widom-Airy statistics of the eigenvalues of  $Q$  near the soft edges. This result applies in the single-cut and multi-cut cases. Further applications include the distribution of the eigenvectors and an analysis of the outliers and BBP-type phase transitions in finite-rank deformations; they will appear elsewhere.

We also apply our method to Wigner matrices whose entries have arbitrary expectation, i.e. we consider  $W + A$  where  $W$  is a Wigner matrix and  $A$  a Hermitian deterministic matrix. We prove the anisotropic local law for  $W + A$  and use it to establish edge universality.

## 1. Introduction

**1.1. Local laws.** The empirical eigenvalue density of a large random matrix typically converges to a deterministic asymptotic density. For Wigner matrices this law is the celebrated Wigner semicircle law [39] and for uncorrelated sample covariance matrices it is the Marchenko-Pastur law [29]. This convergence is best formulated using *Stieltjes transforms*. Let  $Q$  be an  $M \times M$  Hermitian random matrix, normalized so that its eigenvalues are typically of order one, and denote by  $R(z) := (Q - zI_M)^{-1}$  its resolvent. Here  $z = E + i\eta$  is a spectral parameter with positive imaginary part  $\eta$ . Then the Stieltjes transform of the empirical eigenvalue density is equal to  $M^{-1} \text{Tr } R(z)$ , and the convergence mentioned above may be written informally as

$$\frac{1}{M} \text{Tr } R(z) = \frac{1}{M} \sum_{i=1}^M R_{ii}(z) \approx m(z) \quad (1.1)$$

---

\*ETH Zürich, knowles@math.ethz.ch.

†University of Wisconsin, jyin@math.wisc.edu.

for large  $M$  and with high probability. Here  $m(z)$  is the Stieltjes transform of the asymptotic eigenvalue density, which we denote by  $\varrho$ . We call an estimate of the form (1.1) an *averaged law*.

As may be easily seen by taking the imaginary part of (1.1), control of the convergence of  $M^{-1} \text{Tr} R(z)$  yields control of an order  $\eta M$  eigenvalues around the point  $E$ . A *local law* is an estimate of the form (1.1) for all  $\eta \gg M^{-1}$ . Note that the approximation (1.1) cannot be correct at or below the scale  $\eta \asymp M^{-1}$ , at which the behaviour of the left-hand side of (1.1) is governed by the fluctuations of individual eigenvalues. Such local laws have become a cornerstone of random matrix theory, starting from the work [16] where a local law was first established for Wigner matrices. Well known corollaries of a local law include bounds on the eigenvalue counting function as well as eigenvalue rigidity. Moreover, local laws constitute the main tool needed to analyse (a) the distribution of eigenvalues (including the universality of the local spectral statistics), (b) eigenvector delocalization, (c) the distribution of eigenvectors, and (d) finite-rank deformations of  $Q$ .

In fact, for all of the applications (a)–(d), the averaged local law from (1.1) is not sufficient. One has to control not only the normalized trace of  $R(z)$  but the matrix  $R(z)$  itself, by showing that  $R(z)$  is close to some deterministic matrix depending on  $z$ , provided that  $\eta \gg M^{-1}$ . Such control was first obtained for Wigner matrices in [19], where the closeness was established in the sense of individual matrix entries:  $R_{ij}(z) \approx m(z)\delta_{ij}$ . We call such an estimate an *entrywise local law*. More generally, in [7, 25] this closeness was established in the sense of *generalized matrix entries*:

$$\langle \mathbf{v}, R(z) \mathbf{w} \rangle \approx m(z) \langle \mathbf{v}, \mathbf{w} \rangle, \quad \eta \gg M^{-1}, \quad |\mathbf{v}|, |\mathbf{w}| \leq 1. \quad (1.2)$$

Analogous results for uncorrelated sample covariance matrices were obtained in [7, 32]. The estimate (1.2) states that for large  $M$  the resolvent  $R(z)$  is approximately *isotropic* (i.e. proportional to the identity matrix), and we accordingly call an estimate of the form (1.2) an *isotropic local law*. We remark that the basis-independent control in (1.2) is crucial for many applications, including the distribution of eigenvectors and the study of finite-rank deformations of  $Q$ .

Unlike in the case of Wigner matrices and uncorrelated sample covariance matrices mentioned above, the resolvent  $R(z)$  is in general not close to a multiple of the identity matrix, but rather to some general deterministic matrix  $P(z)$ . In that case (1.2) is to be replaced with

$$\langle \mathbf{v}, R(z) \mathbf{w} \rangle \approx \langle \mathbf{v}, P(z) \mathbf{w} \rangle, \quad \eta \gg M^{-1}, \quad |\mathbf{v}|, |\mathbf{w}| \leq 1. \quad (1.3)$$

We call an estimate of the form (1.3) an *anisotropic local law*. The goal of this paper is to develop a method yielding (anisotropic) local laws for many matrix models built from sums and products of deterministic or independent random matrices. Applications include all the four (a)–(d) listed above, some of which we illustrate in this paper.

**1.2. Sample covariance matrices.** For definiteness, and motivated by applications to multivariate statistics, in this paper we focus mainly on sample covariance matrices. (In Section 12, we also explain how to apply our method to deformed Wigner matrices.) We consider sample covariance matrices of the form  $Q = N^{-1} A A^*$ , where  $A$  is an  $M \times N$  matrix. The columns of  $A$  represent  $N$  independent and identically distributed observations of some random  $M$ -dimensional vector  $\mathbf{a}$ . We shortly outline the statistical interpretation of  $Q$ , and refer e.g. to [8] for more details. A fundamental goal of multivariate statistics is to obtain information on the population covariance matrix  $\Sigma := \mathbb{E} \mathbf{a} \mathbf{a}^*$  from  $N$  independent samples  $A = [\mathbf{a}^{(1)} \cdots \mathbf{a}^{(N)}]$  of the population  $\mathbf{a}$ . Then  $Q$  corresponds to  $\Sigma$  with the expectation replaced by the empirical average over the

$N$  samples. If the entries  $\mathbf{a}$  do not have mean zero, then the population covariance matrix  $\Sigma$  reads  $\mathbb{E}(\mathbf{a} - \mathbb{E}\mathbf{a})(\mathbf{a} - \mathbb{E}\mathbf{a})^*$ , and the sample covariance matrix is accordingly obtained by subtracting the empirical average  $\frac{1}{N} \sum_{\nu=1}^N A_{i\nu}$  from each entry  $A_{i\mu}$ . We may therefore write the sample covariance matrix as  $\dot{Q} = (N-1)^{-1} A(I_N - \mathbf{e}\mathbf{e}^*)A^*$ , where we introduced the normalized vector  $\mathbf{e} := N^{-1/2}(1, 1, \dots, 1)^* \in \mathbb{R}^N$ . Since  $\dot{Q}$  is invariant under the deterministic shift  $A_{i\mu} \mapsto A_{i\mu} + f_i$ , we may without loss of generality assume that  $\mathbb{E}A_{i\mu} = 0$ . We shall always make this assumption.

For the sample vector  $\mathbf{a}$  we take a linear model  $\mathbf{a} = T\mathbf{b}$ , where  $T$  is a deterministic  $M \times \widehat{M}$  matrix and  $\mathbf{b}$  is a random  $\widehat{M}$ -dimensional vector with zero expectation. We assume that the entries of  $\mathbf{b}$  are independent. This model includes in particular the usual linear (or “signal + noise”) models from multivariate statistics; see [7, Section 1.2] for more details. Note that, in addition to the assumption  $\mathbb{E}b_i = 0$ , we may without loss of generality assume that  $\mathbb{E}|b_i|^2 = 1$  by absorbing the variance of  $b_i$  into the deterministic matrix  $T$ . Hence, without loss of generality, throughout the following we make the assumptions  $\mathbb{E}b_i = 0$  and  $\mathbb{E}|b_i|^2 = 1$ .

Letting  $X$  be an  $\widehat{M} \times N$  matrix of independent entries satisfying  $\mathbb{E}X_{i\mu} = 0$  and  $\mathbb{E}|X_{i\mu}|^2 = N^{-1}$ , we may therefore write the matrices  $Q$  and  $\dot{Q}$  as

$$Q = TXX^*T^*, \quad \dot{Q} = \frac{N}{N-1}TX(I_N - \mathbf{e}\mathbf{e}^*)X^*T^*. \quad (1.4)$$

It is easy to check that in both cases the associated population covariance matrix is  $\Sigma = \mathbb{E}Q = \mathbb{E}\dot{Q} = TT^*$ . The matrix  $Q$  was first studied in the seminal work of Marchenko and Pastur [29], where it was proved that the Stieltjes transform  $m$  of the asymptotic density  $\varrho$  may be characterized as the solution of an integral equation, (2.11) below, depending on the spectrum of  $\Sigma$ . In the language of free probability, the asymptotic density  $\varrho$  is the free multiplicative convolution of the famous Marchenko-Pastur law with the empirical eigenvalue density of  $\Sigma$ .

**1.3. Outline of results.** For simplicity, we focus on the matrix  $Q$ , bearing in mind that similar results also apply for  $\dot{Q}$  (see Section 11.2). We assume that the three matrix dimensions  $M, \widehat{M}, N$  are comparable, and that the entries of  $\sqrt{N}X$  possess a sufficient number of bounded moments. Moreover, we assume that  $\|\Sigma\|$  is bounded, and that the spectrum of  $\Sigma$  satisfies certain regularity conditions, given Definition 2.7 below, which essentially state that the connected components of the support of  $\varrho$  are separated by some positive constant, and that the density of  $\varrho$  has square root decay near its edges in  $(0, \infty)$ . Note that we do not assume that  $T$  is square, and in particular  $T$  may have many vanishing singular values; this allows us to cover e.g. the general linear model of multivariate statistics.

Our main result is the anisotropic local law for  $Q$ . Roughly, it states that (1.3) holds with

$$P(z) := -(z(I_M + m(z)\Sigma))^{-1},$$

where  $m(z)$  is the Stieltjes transform of the asymptotic density  $\varrho$ . In fact, we prove a more general anisotropic local law that is more useful in applications. Its formulation is most transparent under the additional assumption that  $T = T^* = \Sigma^{1/2}$ , although this assumption is a mere convenience and may be easily removed (see Section 11.1). We prove an anisotropic local law of the form

$$\langle \mathbf{v}, G(z)\mathbf{w} \rangle \approx \langle \mathbf{v}, \Pi(z)\mathbf{w} \rangle, \quad (1.5)$$

for  $\eta \gg M^{-1}$  and  $|\mathbf{v}|, |\mathbf{w}| \leq 1$ , where we defined

$$G(z) := \begin{pmatrix} -\Sigma^{-1} & X \\ X^* & -zI_N \end{pmatrix}^{-1}, \quad \Pi(z) := \begin{pmatrix} -\Sigma(I_M + m(z)\Sigma)^{-1} & 0 \\ 0 & m(z)I_N \end{pmatrix}.$$

A simple application of Schur's complement formula to (1.5) yields the anisotropic local law for the resolvent of  $Q = TXX^*T^*$  and a similar result for the resolvent of the companion matrix  $X^*T^*TX$ . The estimate (1.5) holds with high probability, and we give an explicit and optimal error bound. We remark that the anisotropic local law holds under very general assumptions on the distribution of  $X$ , the dimensions of  $X$  and  $T$ , and the spectrum of  $TT^*$ . In particular, we make no assumptions on the singular vectors of  $T$ . We remark that, previously, an anisotropic global law, valid for  $\eta \asymp 1$ , was derived in [22] for a different matrix model.

As an application of the anisotropic local law, we prove the edge universality of the eigenvalues near the soft spectral edges, whereby the joint distribution of the eigenvalues is asymptotically governed by the Tracy-Widom-Airy statistics of random matrix theory. More precisely, we prove that the asymptotic distribution of the eigenvalues near the soft edges depends only on the nonzero spectrum of  $TT^*$ . This may be regarded as a universality result in both the distribution of the entries of  $X$  and the (left and right) singular vectors of  $T$ , including their dimensions. We then conclude that the Tracy-Widom-Airy statistics hold near the soft edges by noting that they have been previously established [11, 21, 26, 30] for Gaussian  $X$  and diagonal  $T$ .

We comment briefly on the history of edge universality for sample covariance matrices of the form  $Q$ . The Tracy-Widom-Airy statistics were first established near the rightmost spectral edge in the case of complex Gaussian  $X$  in [11, 30]. In [5], this result was extended to general complex  $X$  under the assumption that  $\Sigma$  is diagonal, i.e. the population vector  $\mathbf{a}$  is uncorrelated. Very recently, in [21] the Tracy-Widom-Airy statistics were established near all soft edges in the case of complex Gaussian  $X$ . Moreover, in [26], building on a new comparison method developed in [27], the Tracy-Widom-Airy statistics near the rightmost edge were established also in the case of real Gaussian  $X$ , or general real  $X$  and diagonal  $\Sigma$ . We remark that all proofs from [11, 21, 30] crucially rely on the integrable structure of  $Q$  in the complex Gaussian case (the Harish-Chandra-Itzykson-Zuber formula and the determinantal form of the eigenvalue process); this structure is not available in the real case, and the method of [26] is different from that of [11, 21, 30].

We also prove the rigidity of eigenvalues, as well as the complete delocalization of the eigenvectors with respect to an arbitrary deterministic basis. Further applications of the anisotropic local law, such as the distribution of the eigenvectors and an analysis of the outliers and BBP-type phase transitions in finite-rank deformations, will appear elsewhere.

Finally, we also apply our method to *deformed Wigner matrices* of the form  $W + A$ , where  $W$  is a Wigner matrix and  $A$  a bounded Hermitian matrix. This model describes Wigner matrices whose entries may have arbitrary expectations. As for  $Q$ , we establish the Tracy-Widom-Airy statistics near the spectral edges of  $W + A$ . More precisely, we prove that the asymptotic distribution of the eigenvalues near the edges depends only on the asymptotic spectrum of  $A$ , which may be regarded as a universality result in the distribution of  $W$  and the eigenvectors of  $A$ . We then conclude that the Tracy-Widom-Airy statistics hold near the edges by noting that they have been previously established [27] for diagonal  $A$ .

**1.4. Overview of the proof.** We conclude this section by outlining some ideas of the proof of the anisotropic local law. Roughly, the proof proceeds in three steps: (A) the entrywise local law for Gaussian  $X$  and diagonal  $\Sigma$ , (B) the anisotropic local law for Gaussian  $X$  and general  $\Sigma$ , and (C) the anisotropic local law for general  $X$  and general  $\Sigma$ . Steps (A) and (B) may be performed by adapting the methods of [7], and we do not comment on them any further.

The main argument, and the bulk of the proof, is Step (C). Its core is a *self-consistent comparison method*, which yields the anisotropic local law for general  $X$  assuming it has been proved for

Gaussian  $X$ . Up to now, all interpolation or Lindeberg replacement arguments in random matrix theory have crucially relied on a local law as input. Since the local law is exactly what we are trying to prove, a different approach is clearly needed – one that does not need a local law to work. Our method yields a new way of deriving local laws for general random matrices expressed as polynomials of deterministic and random matrices whose entries are independent. It relies on two key novel ideas: (a) continuous Green function comparison argument where the errors are controlled self-consistently, and (b) a bootstrapping of the Green function comparison on the spectral scale  $\eta$ .

In the remainder of this subsection we give a few details of our method, and in particular explain the ideas (a) and (b) in more detail. We construct a continuous family  $(X^\theta)_{\theta \in [0,1]}$  of matrices, whereby  $X^0$  is a Gaussian ensemble and  $X^1$  is the ensemble in which we are interested. Then we operate in the  $(\theta, \eta)$ -plane and perform simultaneously a continuous interpolation in  $\theta$  and a discrete bootstrapping in  $\eta$ . (See Figure 1.1 below.)

Interpolation methods have been extensively used in random matrix theory, in the form of both discrete Lindeberg-type replacement schemes [10, 19, 20, 36] and continuous interpolations [26–28]. Moreover, Dyson Brownian motion, as used e.g. in [17, 18, 26, 27], may be regarded as a form of continuous interpolation for the case that  $X^0$  is Gaussian. The usual choice of interpolation, as used in [26, 28] and also resulting from Dyson Brownian motion, is  $X^\theta = \sqrt{1-\theta} X^0 + \sqrt{\theta} X^1$ . In this paper, we instead interpolate using i.i.d. Bernoulli random variables:

$$X_\alpha^\theta := \chi_\alpha^\theta X_\alpha^1 + (1 - \chi_\alpha^\theta) X_\alpha^0, \quad (1.6)$$

where  $\alpha = (i, \mu)$  indexes the matrix entries and  $(\chi_\alpha^\theta)$  is a family of i.i.d. Bernoulli- $\theta$  random variables. The choice (1.6) is convenient for our purposes, since the expectation of a function of  $X^\theta = (X_\alpha^\theta)$  is particularly easy to differentiate in  $\theta$ , resulting in explicit formulas that are the starting point of our analysis. Indeed, as explained in Section 7 (see Remark 7.18), for any smooth enough function  $F$  and  $\ell \in \mathbb{N}$  we have an identity of the form

$$\frac{d}{d\theta} \mathbb{E} F(X^\theta) = \sum_{n=1}^{\ell} \sum_{\alpha} K_{n,\alpha}^\theta \mathbb{E} \left( \frac{\partial}{\partial X_\alpha^\theta} \right)^n F(X^\theta) + \mathcal{E}_\ell, \quad (1.7)$$

where  $K_{n,\alpha}^\theta$  is defined through the formal power series

$$\sum_{n \geq 1} K_{n,\alpha}^\theta t^n := \frac{\mathbb{E} e^{tX_\alpha^1} - \mathbb{E} e^{tX_\alpha^0}}{\mathbb{E} e^{tX_\alpha^\theta}}, \quad (1.8)$$

and the error term  $\mathcal{E}_\ell$  is negligible for large enough  $\ell$ . In particular,  $K_{n,\alpha}^\theta$  depends only on the first  $n$  moments of  $X_\alpha^0$  and  $X_\alpha^1$ , and vanishes if the first  $n$  moments of  $X_\alpha^0$  and  $X_\alpha^1$  match. We remark that the interpolation method from (1.6)–(1.8) can be used to simplify several previous comparison arguments in random matrix theory, especially the proofs of edge universality in [13, 20].

Next, we explain the basic strategy behind the self-consistent comparison method mentioned in (a) above. We choose a function

$$F_{\mathbf{v}}^p(X, z) := |\langle \mathbf{v}, (G(z) - \Pi(z)) \mathbf{v} \rangle|^p$$

that controls the error in the anisotropic law. Here  $\mathbf{v}$  is a deterministic vector on the unit sphere  $\mathbb{S}$  and  $p$  a fixed integer. Our goal is to prove that  $|\langle \mathbf{v}, (G(z) - \Pi(z)) \mathbf{v} \rangle| \leq N^{o(1)} \Psi(z)$  with high

probability, where  $\Psi$  is an explicit error parameter defined in (3.8) below. By Markov's inequality, it suffices to prove  $\mathbb{E}F_{\mathbf{v}}^p(X^1, z) \leq N^{o(1)}\Psi(z)^p$  for large but fixed  $p$ . Since we know (by assumption on the Gaussian case) that  $\mathbb{E}F_{\mathbf{v}}^p(X^0, z) \leq N^{o(1)}\Psi(z)^p$ , it suffices to estimate the derivative as  $\frac{d}{d\theta}\mathbb{E}F_{\mathbf{v}}^p(X^\theta, z) \leq N^{o(1)}\Psi(z)^p$ . In fact, by Grönwall's inequality, it suffices to establish the weaker *self-consistent* estimate

$$\frac{d}{d\theta}\mathbb{E}F_{\mathbf{v}}^p(X^\theta, z) \leq N^{o(1)}\Psi(z)^p + C \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E}F_{\mathbf{w}}^p(X^\theta, z). \quad (1.9)$$

Note that (1.9) is self-consistent in the sense that the right-hand side depends on the quantities to be estimated. Most of the work is therefore to estimate the right-hand side of (1.7) with  $F := F_{\mathbf{v}}^p$  by the right-hand side of (1.9). The derivatives on the right-hand side of (1.7) are polynomials of generalized matrix entries of  $G$ . Their structure has to be carefully tracked, which we do in Sections 7 and 8 by encoding them using an appropriately constructed family of words.

It turns out, however, that in general these polynomials involve factors that cannot be estimated by either term on the right-hand side of (1.9). To handle these bad terms, as mentioned in (b) above, we make use of a *bootstrapping in the spectral scale*  $\eta$ . The basic idea is to perform the interpolation from  $\theta = 0$  to  $\theta = 1$  on a spectral scale  $\tilde{\eta}$ , and to use the obtained anisotropic local law at scale  $\tilde{\eta}$  to deduce rough a priori bounds on the generalized entries of  $G$  on the smaller spectral scale  $\eta$ . The bootstrapping is started at  $\eta = 1$ , where the trivial a priori bound  $\eta^{-1} = 1$  on the entries of  $G$  is sufficient. The bootstrapping proceeds in multiplicative increments of  $N^{-\delta}$ , where  $\delta > 0$  is a fixed small exponent. Hence, the bootstrapping iteration needs to be performed only an order  $O(1)$  times. (We remark that this bootstrapping is different from the stochastic continuity arguments commonly used to establish local laws for Wigner matrices [14–16], which propagate much more precise estimates but have to be iterated an order  $N^{O(1)}$  times.)

At each step of the bootstrapping, we use the anisotropic local law on the scale  $\tilde{\eta} = N^\delta \eta$  and simple monotonicity properties of  $G$  to obtain a priori bounds on  $|G_{\mathbf{v}\mathbf{w}}|$  and  $\text{Im } G_{\mathbf{w}\mathbf{w}}$  on the scale  $\eta$ . Roughly, the smallness that allows us to propagate the bootstrapping arises from a certain structure of summation indices in the polynomials, which yields factors of  $\text{Im } G_{\mathbf{w}\mathbf{w}}$ . It is important to ensure that a sufficiently large number of such factors is generated, and that the a priori bound on them is close to optimal. In contrast, the a priori bound that one can use for general entries  $|G_{\mathbf{v}\mathbf{w}}|$  is very rough. Figure 1.1 gives a schematic representation of the self-consistent comparison method, which combines a continuous interpolation in  $\theta$  and a discrete bootstrapping in  $\eta$ . We refer to Sections 7.1 and 8.1 for a more detailed outline of the proof.

The next section is devoted to the definition of the model and the statement of results. We give an outline of the structure of the paper in Section 3.5 below.

**Conventions.** The fundamental large parameter is  $N$ . All quantities that are not explicitly constant may depend on  $N$ ; we almost always omit the argument  $N$  from our notation.

We use  $C$  to denote a generic large positive constant, which may depend on some fixed parameters and whose value may change from one expression to the next. Similarly, we use  $c$  to denote a generic small positive constant. If a constant  $C$  depends on some additional quantity  $\alpha$ , we indicate this by writing  $C_\alpha$ . For two positive quantities  $A_N$  and  $B_N$  depending on  $N$  we use the notation  $A_N \asymp B_N$  to mean  $C^{-1}A_N \leq B_N \leq CA_N$  for some positive constant  $C$ . For  $a < b$  we set  $\llbracket a, b \rrbracket := [a, b] \cap \mathbb{Z}$ . We use the notation  $\mathbf{v} = (\mathbf{v}(i))_{i=1}^n$  for vectors in  $\mathbb{C}^n$ , and denote by  $|\cdot| = \|\cdot\|_2$  the Euclidean norm of vectors. We use  $\|\cdot\|$  to denote the Euclidean operator norm of a matrix. We often write the  $k \times k$  identity matrix  $I_k$  simply as 1 when there is no risk of confusion.

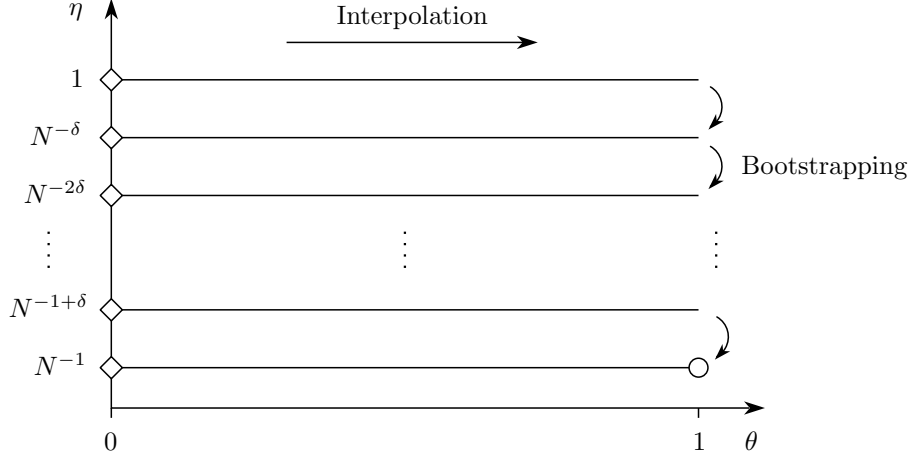


FIGURE 1.1. A schematic overview of the self-consistent comparison method. The horizontal axis depicts the interpolation parameter  $\theta \in [0, 1]$ , with  $\theta = 0$  corresponding to the Gaussian case and  $\theta = 1$  to the general case. The vertical axis depicts the spectral scale  $\eta$ , with  $\eta = 1$  corresponding to the macroscopic scale and  $\eta = N^{-1}$  the microscopic scale. The self-consistent comparison argument takes place in the  $(\theta, \eta)$ -plane, and follows the horizontal lines from left to right and top to bottom, i.e.  $(0, 1) \rightarrow (1, 1)$ ,  $(0, N^{-\delta}) \rightarrow (1, N^{-\delta})$ ,  $(0, N^{-2\delta}) \rightarrow (1, N^{-2\delta})$ ,  $\dots$ . The diamonds on the vertical axis indicate that the anisotropic local law is known at  $(\theta, \eta)$ . The circle indicates the goal  $(\theta, \eta) = (1, N^{-1})$ .

We use  $\tau > 0$  in various assumptions to denote a positive constant that may be chosen arbitrarily small. A smaller value of  $\tau$  corresponds to a weaker assumption. All constants may  $C, c$  depend on  $\tau$ , and we neither indicate nor track this dependence.

## 2. Model

In this section we define our model, list our key assumptions, and explain the basic structure of the asymptotic eigenvalue density  $\varrho$ .

**2.1. Definition of the model.** We consider the  $M \times M$  matrix  $Q := TXX^*T^*$ , where  $T$  is a deterministic  $M \times \widehat{M}$  matrix and  $X$  a random  $\widehat{M} \times N$  matrix. We regard  $N$  as the fundamental parameter and  $M \equiv M_N$  and  $\widehat{M} \equiv \widehat{M}_N$  as depending on  $N$ . Here, and throughout the following, we omit the index  $N$  from our notation, bearing in mind that all quantities that are not explicitly constant (such as the constant  $\tau$ ) may depend on  $N$ . For simplicity, we always make the assumption that

$$M \asymp \widehat{M} \asymp N. \quad (2.1)$$

(This assumption may be relaxed to  $\log M \asymp \log \widehat{M} \asymp \log N$  with some extra work; see [8]. We do not pursue this direction here.) We introduce the dimensional ratio

$$\phi := \frac{M}{N}. \quad (2.2)$$



Note that (2.1) implies

$$\tau \leq \phi \leq \tau^{-1} \quad (2.3)$$

provided  $\tau > 0$  is chosen small enough.

We assume that the entries  $X_{i\mu}$  of the  $\widehat{M} \times N$  matrix  $X$  are independent (but not necessarily identically distributed) real-valued random variables satisfying

$$\mathbb{E}X_{i\mu} = 0, \quad \mathbb{E}|X_{i\mu}|^2 = \frac{1}{N} \quad (2.4)$$

for all  $i$  and  $\mu$ . In addition, we assume that, for all  $p \in \mathbb{N}$ , the random variables  $\sqrt{N}X_{i\mu}$  have a uniformly bounded  $p$ -th moment. In other words, we assume that for all  $p \in \mathbb{N}$  there is a constant  $C_p$  such that

$$\mathbb{E}|\sqrt{N}X_{i\mu}|^p \leq C_p \quad (2.5)$$

for all  $i$  and  $\mu$ . The assumption that (2.5) hold for all  $p \in \mathbb{N}$  may be easily relaxed. For instance, it is easy to check that our results and their proofs remain valid, after minor adjustments, if we only require that (2.5) holds for all  $p \leq C$  for some large enough constant  $C$ . We do not pursue such generalizations.

For definiteness, in this paper we focus mostly on the real symmetric case ( $\beta = 1$ ), where all matrix entries are real. We remark, however, that all of our results and proofs also hold, after minor changes, in the complex Hermitian case ( $\beta = 2$ ), where  $X_{i\nu} \in \mathbb{C}$  and in addition to (2.4) we have  $\mathbb{E}X_{i\mu}^2 = 0$ .

The population covariance matrix is defined as

$$\Sigma := \mathbb{E}Q = TT^*.$$

We denote the eigenvalues of  $\Sigma$  by

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_M \geq 0.$$

Let

$$\pi := \frac{1}{M} \sum_{i=1}^M \delta_{\sigma_i} \quad (2.6)$$

denote the empirical spectral density of  $\Sigma$ . We suppose that

$$\sigma_1 \leq \tau^{-1} \quad (2.7)$$

and that

$$\pi([0, \tau]) \leq 1 - \tau. \quad (2.8)$$

This latter assumption means that the spectrum of  $\Sigma$  cannot be concentrated at zero.

Sometimes it will be convenient to make the following stronger assumption on  $T$ :

$$\widehat{M} = M \text{ and } T = T^* = \Sigma^{1/2} > 0. \quad (2.9)$$

The assumption (2.9) will frequently simplify the presentation and the proofs. Thanks to our general assumptions (2.7) and (2.8), it will always be relatively easy to remove (2.9). In particular, we emphasize that the assumption  $\Sigma > 0$  is purely qualitative in nature, and is made in order to simplify expressions involving the inverse of  $\Sigma$ . The case of  $\Sigma \geq 0$  may always be easily obtained by considering  $\Sigma + \varepsilon I_M$  and then taking  $\varepsilon \downarrow 0$  at fixed  $N$ . We refer to Section 11.1 below for the details on how to remove the assumption (2.9).

To avoid repetition, we summarize our basic assumptions for future reference.



ASSUMPTION 2.1. We suppose that (2.1), (2.4), (2.5), (2.7), and (2.8) hold.

**2.2. Asymptotic eigenvalue density.** Next, we define the asymptotic eigenvalue density of  $X^*T^*TX$ ,  $\varrho$ , via its Stieltjes transform,  $m$ . Let  $\mathbb{C}_+$  denote the complex upper-half plane.

LEMMA 2.2. Let  $\pi$  be a compactly supported probability measure on  $\mathbb{R}$ , and let  $\phi > 0$ . Then for each  $z \in \mathbb{C}_+$  there is a unique  $m \equiv m(z) \in \mathbb{C}_+$  satisfying

$$\frac{1}{m} = -z + \phi \int \frac{x}{1 + mx} \pi(dx). \quad (2.10)$$

Moreover,  $m(z)$  is the Stieltjes transform of a probability measure  $\varrho$  with bounded support in  $[0, \infty)$ .

PROOF. This is a well-known result, which may be proved using a fixed point argument in  $\mathbb{C}_+$ ; the function  $m(z)$  is the Stieltjes transform of the multiplicative free convolution of  $\pi$  and the Marchenko-Pastur law with dimensional ratio (2.2). See e.g. [34, Section 5] and [3] for more details.  $\square$

DEFINITION 2.3 (ASYMPTOTIC DENSITY). We define the deterministic function  $m \equiv m_{\Sigma, N} : \mathbb{C}_+ \rightarrow \mathbb{C}_+$  as the unique solution  $m(z)$  of (2.10) with  $\phi$  defined in (2.2) and  $\pi$  defined in (2.6). We denote by  $\varrho \equiv \varrho_{\Sigma, N}$  the probability measure associated with  $m$ , and call it the asymptotic eigenvalue density.

The rest of this subsection is devoted to a discussion of the basic properties of the asymptotic eigenvalue density  $\varrho$ . Much this discussion is well known; see e.g. [1, 21, 35]. The reader interested only in the principal components of  $Q$ , i.e. its top eigenvalues, can skip this subsection and proceed directly to the results in Theorem 3.14 (with  $k = 1$ ) and Corollary 3.19 (i).

Let  $n := |\text{supp } \pi \setminus \{0\}|$  be the number of distinct nonzero eigenvalues of  $\Sigma$ , and write

$$\text{supp } \pi \setminus \{0\} = \{s_1 > s_2 > \cdots > s_n\}.$$

Let  $z \in \mathbb{C}_+$ . Then  $m \equiv m(z)$  may also be characterized as the unique solution of the equation

$$z = f(m), \quad \text{Im } m > 0, \quad (2.11)$$

where we defined

$$f(x) := -\frac{1}{x} + \phi \sum_{i=1}^n \frac{\pi(\{s_i\})}{x + s_i^{-1}}. \quad (2.12)$$

In Figures 2.1 and 2.2, we illustrate the graph of  $f$  for the cases  $\phi < 1$  and  $\phi > 1$  respectively. In Figure 2.3 we plot the density of  $\varrho$  for the examples from Figures 2.1 and 2.2. The behaviour of  $\varrho$  may be entirely understood by an elementary analysis of  $f$ .

It is convenient to extend the domain of  $f$  to the real projective line  $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \cong S^1$ . Clearly,  $f$  is smooth on the  $n + 1$  open intervals of  $\overline{\mathbb{R}}$  defined through

$$I_1 := (-s_1^{-1}, 0), \quad I_i := (-s_i^{-1}, -s_{i-1}^{-1}) \quad (i = 2, \dots, n), \quad I_0 := \overline{\mathbb{R}} \setminus \bigcup_{i=1}^n \overline{I_i}.$$

Next, we introduce the multiset  $\mathcal{C} \subset \overline{\mathbb{R}}$  of critical points of  $f$ , using the conventions that a nondegenerate critical point is counted once and a degenerate critical point twice, and if  $\phi = 1$  then  $\infty$  is a nondegenerate critical point. The following three elementary lemmas are proved in Appendix A.

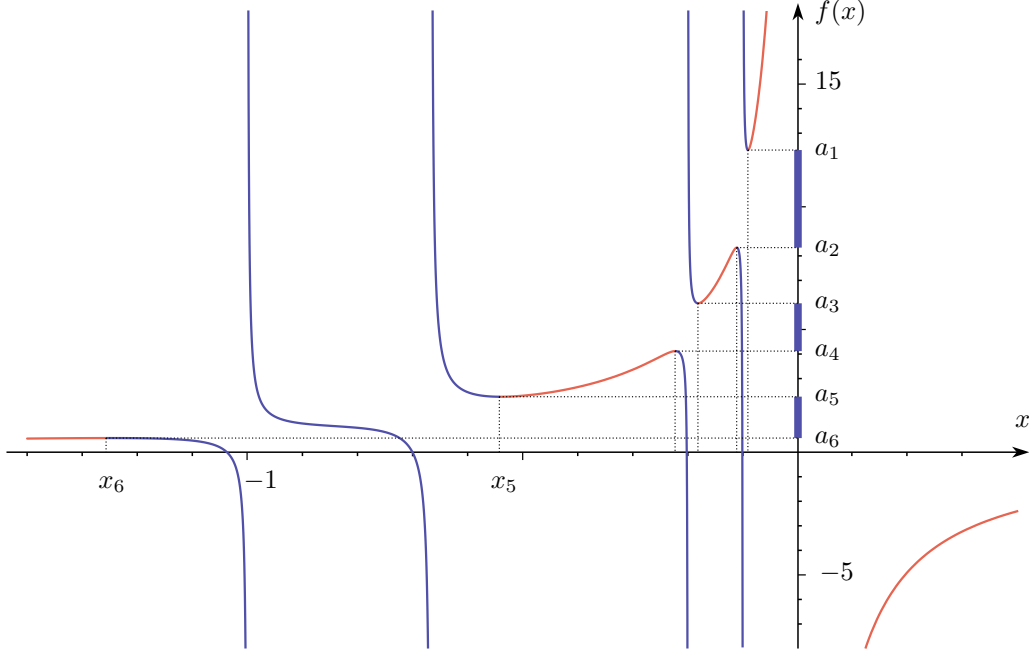


FIGURE 2.1. The function  $f(x)$  for  $\phi\pi = 0.01\delta_{10} + 0.01\delta_5 + 0.05\delta_{1.5} + 0.03\delta_1$ . Here,  $\phi = 0.1$  and we have  $p = 3$  connected components. The vertical asymptotes are located at  $-\sigma^{-1}$  for  $\sigma \in \text{supp } \pi$ . The support of  $\varrho$  is indicated with thick blue lines on the vertical axis. The inverse of  $m|_{\mathbb{R} \setminus \text{supp } \varrho}$  is drawn in red.

LEMMA 2.4 (CRITICAL POINTS). *We have  $|\mathcal{C} \cap I_0| = |\mathcal{C} \cap I_1| = 1$  and  $|\mathcal{C} \cap I_i| \in \{0, 2\}$  for  $i = 2, \dots, n$ .*

We deduce from Lemma 2.4 that  $|\mathcal{C}| = 2p$  is even. We denote by  $x_1 \geq x_2 \geq \dots \geq x_{2p-1}$  the  $2p - 1$  critical points in  $I_1 \cup \dots \cup I_n$ , and by  $x_{2p}$  the unique critical point in  $I_0$ . For  $k = 1, \dots, 2p$  we define the critical values  $a_k := f(x_k)$ .

LEMMA 2.5 (ORDERING OF THE CRITICAL VALUES). *We have  $a_1 \geq a_2 \geq \dots \geq a_{2p}$ . Moreover, for  $k = 1, \dots, 2p$  we have  $a_k = f(x_k)$  and  $x_k = m(a_k)$ , where  $m$  is defined on  $\mathbb{R}$  by continuous extension from  $\mathbb{C}_+$ , and we use the convention  $m(0) := \infty$  for  $\phi = 1$ . Finally, under the assumptions (2.3) and (2.7) there exists a constant  $C$  (depending on  $\tau$ ) such that  $a_k \in [0, C]$  for  $k = 1, \dots, 2p$ .*

The following result gives the basic structure of  $\varrho$ .

LEMMA 2.6 (STRUCTURE OF  $\varrho$ ). *We have*

$$\text{supp } \varrho \cap (0, \infty) = \left( \bigcup_{k=1}^p [a_{2k}, a_{2k-1}] \right) \cap (0, \infty). \quad (2.13)$$

Our assumptions on  $\pi$  (i.e. on the spectrum of  $\Sigma$ ) take the form of the following regularity conditions.

DEFINITION 2.7 (REGULARITY). *Fix  $\tau > 0$ .*

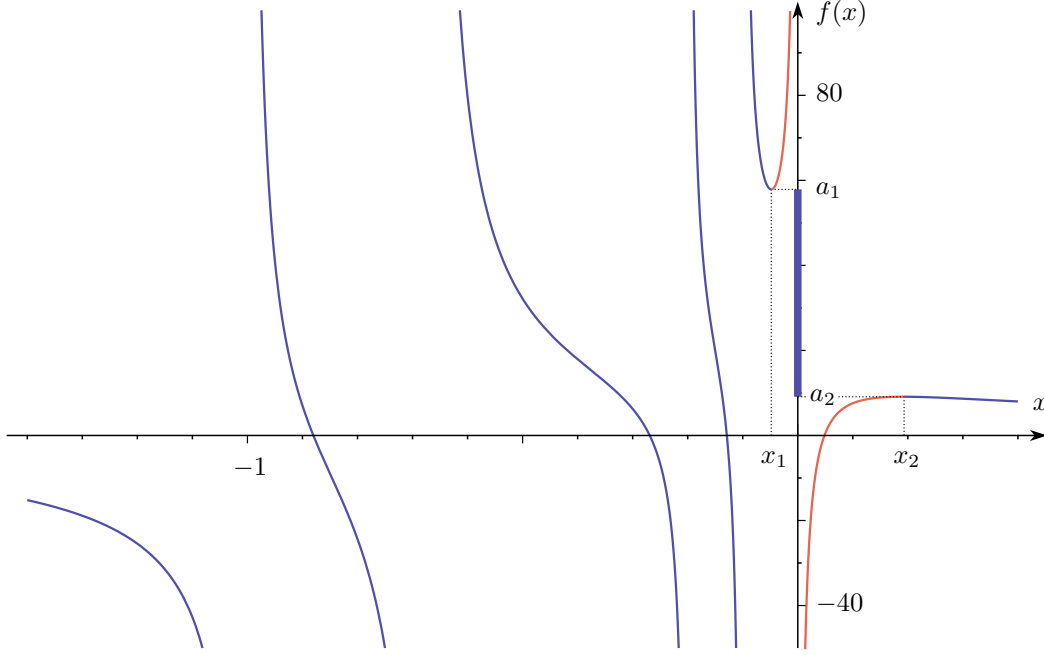


FIGURE 2.2. The function  $f(x)$  for  $\phi\pi = \delta_{10} + \delta_5 + 5\delta_{1.5} + 3\delta_1$ . Here,  $\phi = 10$  and we have  $p = 1$  connected component. The vertical asymptotes are located at  $-\sigma^{-1}$  for  $\sigma \in \text{supp } \pi$ . The support of  $\varrho$  is indicated with a thick blue line. The inverse of  $m|_{\mathbb{R} \setminus \text{supp } \varrho}$  is drawn in red.

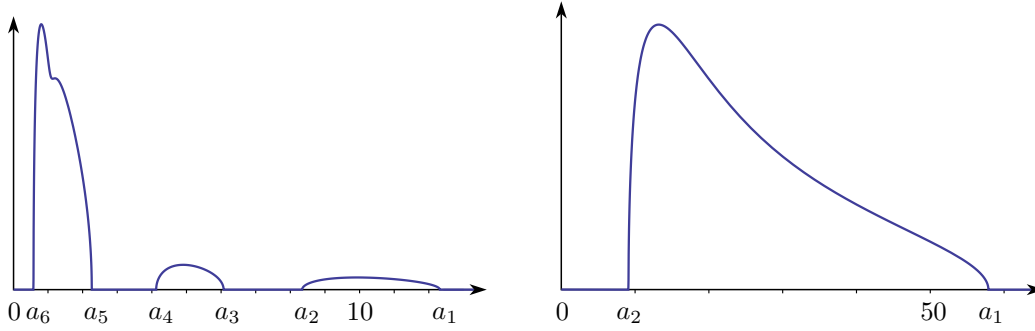


FIGURE 2.3. The densities of  $\varrho$  for the examples from Figures 2.1 (left) and 2.2 (right).

(i) We say that the edge  $k = 1, \dots, 2p$  is regular if

$$a_k \geq \tau, \quad \min_{l \neq k} |a_k - a_l| \geq \tau, \quad \min_i |x_k + s_i^{-1}| \geq \tau. \quad (2.14)$$

(ii) We say that the bulk component  $k = 1, \dots, p$  is regular if for any fixed  $\tau' > 0$  there exists a constant  $c \equiv c_{\tau, \tau'} > 0$  such that the density<sup>1</sup> of  $\varrho$  in  $[a_{2k} + \tau', a_{2k-1} - \tau']$  is bounded from below by  $c$ .

<sup>1</sup>This means the density of the absolutely continuous part of  $\varrho$ . In fact, as explained after Lemma 4.10 below,  $\varrho$  is always absolutely continuous on  $\mathbb{R} \setminus \{0\}$ .

The edge regularity condition from Definition 2.7 (i) has previously appeared (in a slightly different form) in several works on sample covariance matrices. For the rightmost edge  $k = 1$ , it was introduced in [11] and was subsequently used in the works [5, 26, 30] on the distribution of eigenvalues near the top edge  $a_1$ . For general  $k$ , it was introduced in [21]. The second condition of (2.14) states that the gap in the spectrum of  $\varrho$  adjacent to the edge  $a_k$  does not close for large  $N$ ; the third condition of (2.14) ensures a regular square-root behaviour of the spectral density near  $a_k$ , and in particular rules out outliers.

The bulk regularity condition from Definition 2.7 (ii) imposes a lower bound on the density of eigenvalues away from the edges. Without it, one can have points in the interior of  $\text{supp } \varrho$  with an arbitrarily small density, and our results can be shown to be false. Such points may be easily constructed by taking  $\pi$  with a degenerate critical point  $x$ , which corresponds to two components of  $\varrho$  touching at  $a = f(x)$ . Then it is not hard to check that one can construct an arbitrarily small perturbation  $\tilde{\pi}$  of  $\pi$  such that the point  $a$  is well-separated from the edges of the support of the associated density  $\tilde{\varrho}$ , and the density  $\tilde{\varrho}$  at  $a$  is an arbitrarily small positive number. The condition from Definition 2.7 (ii) is stable under perturbation of  $\pi$ ; see Remark A.7 below.

We conclude this subsection with a couple of examples verifying the regularity conditions of Definition 2.7.

**EXAMPLE 2.8** (Bounded number of distinct eigenvalues). We suppose that  $n$  is fixed, and that  $s_1, \dots, s_n$  and  $\phi\pi(\{s_1\}), \dots, \phi\pi(\{s_n\})$  all converge in  $(0, \infty)$  as  $N \rightarrow \infty$ . We suppose that all critical points of  $\lim_N f$  are nondegenerate, and that  $\lim_N a_i > \lim_N a_{i+1}$  for  $i = 1, \dots, 2p$ . Then it is easy to check that, for small enough  $\tau$ , all edges  $k = 1, \dots, 2p$  and bulk components  $k = 1, \dots, p$  are regular in the sense of Definition 2.7.

**EXAMPLE 2.9** (Continuous limit). We suppose that  $\phi$  converges in  $(0, \infty) \setminus \{1\}$ . Moreover, we suppose that  $\pi$  is supported in some interval  $[a, b] \subset (0, \infty)$ , and that  $\pi$  converges in distribution to some measure  $\pi_\infty$  that is absolutely continuous and whose density satisfies  $\tau \leq d\pi_\infty(E)/dE \leq \tau^{-1}$  for  $E \in [a, b]$ . It is easy to check that in this case  $p = 1$ , and that the two edges and the single bulk component are regular in the sense of Definition 2.7.

### 3. Results

In this section we state our main results for the sample covariance matrix  $Q$  defined in Section 2. To ease readability, we state the anisotropic local law under three different sets of increasingly weak assumptions. In Section 3.2, we assume that all edges and bulk components are regular in the sense of Definition 2.7. In Section 3.3, we assume only regularity of a single edge or bulk component, and state the anisotropic local law in the vicinity of the corresponding edge or bulk component. As an application, we prove the edge universality near a regular edge. Finally, in Section 3.4 we give a general anisotropic law, where the concrete regularity assumptions from Definition 2.7 are replaced with a more general but abstract stability condition. The reader interested only in the principal components of  $Q$  can proceed directly to the results in Theorem 3.14 (with  $k = 1$ ) and Corollary 3.19 (i).

**3.1. Basic definitions.** In this preliminary subsection we introduce some basic notations and definitions. Our main results take the form of *local laws*, which relate the resolvents

$$R_N(z) := (X^* T^* T X - z)^{-1} \quad \text{and} \quad R_M(z) := (T X X^* T^* - z)^{-1} \quad (3.1)$$

to the Stieltjes transform  $m$  of the asymptotic density  $\varrho$ . These local laws may be formulated in a simple, unified fashion under the assumption (2.9) using an  $(M + N) \times (M + N)$  block matrix, which is a linear function of  $X$ .

DEFINITION 3.1 (INDEX SETS). *We introduce the index sets*

$$\mathcal{I}_M := \llbracket 1, M \rrbracket, \quad \mathcal{I}_N := \llbracket M + 1, M + N \rrbracket, \quad \mathcal{I} := \mathcal{I}_M \cup \mathcal{I}_N = \llbracket 1, M + N \rrbracket.$$

*We consistently use the letters  $i, j \in \mathcal{I}_M$ ,  $\mu, \nu \in \mathcal{I}_N$ , and  $s, t \in \mathcal{I}$ . We label the indices of the matrices according to*

$$X = (X_{i\mu} : i \in \mathcal{I}_M, \mu \in \mathcal{I}_N), \quad \Sigma = (\Sigma_{ij} : i, j \in \mathcal{I}_M).$$

DEFINITION 3.2 (LINEARIZING BLOCK MATRIX). *Under the condition (2.9) and for  $z \in \mathbb{C}_+$  we define the  $\mathcal{I} \times \mathcal{I}$  matrix*

$$G(z) \equiv G^\Sigma(X, z) := \begin{pmatrix} -\Sigma^{-1} & X \\ X^* & -z \end{pmatrix}^{-1}. \quad (3.2)$$

The motivation behind this definition is that, assuming (2.9), a control of  $G$  immediately yields control of the resolvents  $R_N$  and  $R_M$  via the identities

$$G_{ij} = (z\Sigma^{1/2}R_M\Sigma^{1/2})_{ij} \quad (3.3)$$

for  $i, j \in \mathcal{I}_M$  and

$$G_{\mu\nu} = (R_N)_{\mu\nu} \quad (3.4)$$

for  $\mu, \nu \in \mathcal{I}_N$ . Both of these identities may be easily checked using Schur's complement formula.

Next, we introduce a deterministic matrix  $\Pi$ , which we shall prove is close to  $G$  with high probability (in the sense of Definition 3.4 (iii) below).

DEFINITION 3.3 (DETERMINISTIC EQUIVALENT OF  $G$ ). *For  $z \in \mathbb{C}_+$  we define the  $\mathcal{I} \times \mathcal{I}$  deterministic matrix*

$$\Pi(z) \equiv \Pi^\Sigma(z) := \begin{pmatrix} -\Sigma(1 + m(z)\Sigma)^{-1} & 0 \\ 0 & m(z) \end{pmatrix}. \quad (3.5)$$

We also extend  $\Sigma$  to an  $\mathcal{I} \times \mathcal{I}$  matrix

$$\underline{\Sigma} := \begin{pmatrix} \Sigma & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.6)$$

We consistently use the notation  $z = E + i\eta$  for the spectral parameter  $z$ . Throughout the following we regard the quantities  $E(z)$  and  $\eta(z)$  as functions of  $z$  and usually omit the argument unless it is needed to avoid confusion. The spectral parameter  $z$  will always lie in the fundamental domain

$$\mathbf{D} \equiv \mathbf{D}(\tau, N) := \{z \in \mathbb{C}_+ : |z| \geq \tau, |E| \leq \tau^{-1}, N^{-1+\tau} \leq \eta \leq \tau^{-1}\}, \quad (3.7)$$

where  $\tau > 0$  is a fixed parameter.

The following notion of a high-probability bound was introduced in [12], and has been subsequently used in a number of works on random matrix theory. It provides a simple way of systematizing and making precise statements of the form “ $\xi$  is bounded with high probability by  $\zeta$  up to small powers of  $N$ ”.

DEFINITION 3.4 (STOCHASTIC DOMINATION). (i) Let

$$\xi = (\xi^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}), \quad \zeta = (\zeta^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)})$$

be two families of nonnegative random variables, where  $U^{(N)}$  is a possibly  $N$ -dependent parameter set.

We say that  $\xi$  is stochastically dominated by  $\zeta$ , uniformly in  $u$ , if for all (small)  $\varepsilon > 0$  and (large)  $D > 0$  we have

$$\sup_{u \in U^{(N)}} \mathbb{P} \left[ \xi^{(N)}(u) > N^\varepsilon \zeta^{(N)}(u) \right] \leq N^{-D}$$

for large enough  $N \geq N_0(\varepsilon, D)$ . Throughout this paper the stochastic domination will always be uniform in all parameters (such as matrix indices, deterministic vectors, and spectral parameters  $z \in \mathbf{D}$ ) that are not explicitly fixed. Note that  $N_0(\varepsilon, D)$  may depend on quantities that are explicitly constant, such as  $\tau$  and  $C_p$  from (2.5).

- (ii) If  $\xi$  is stochastically dominated by  $\zeta$ , uniformly in  $u$ , we use the notation  $\xi \prec \zeta$ . Moreover, if for some complex family  $\xi$  we have  $|\xi| \prec \zeta$  we also write  $\xi = O_{\prec}(\zeta)$ .
- (iii) We extend the definition of  $O_{\prec}(\cdot)$  to matrices in the weak operator sense as follows. Let  $A$  be a family of complex square random matrices and  $\zeta$  a family of nonnegative random variables. Then we use  $A = O_{\prec}(\zeta)$  to mean  $|\langle \mathbf{v}, A \mathbf{w} \rangle| \prec \zeta |\mathbf{v}| |\mathbf{w}|$  uniformly for all deterministic vectors  $\mathbf{v}$  and  $\mathbf{w}$ .

REMARK 3.5. Because of (2.1), all (or some) factors of  $N$  in Definition 3.4 could be replaced with  $M$  or  $\widehat{M}$  without changing the definition of stochastic domination.

**3.2. The full spectrum.** As explained at the beginning of this section, we first state the anisotropic local law for all arguments  $z$  under the assumption that all edges and bulk components are regular in the sense of Definition 2.7. In the next subsection, we relax these assumptions by restricting the domain of  $z$  to the vicinity of an edge or bulk component, and only requiring the regularity of the corresponding edge or bulk component.

Our main result is the following anisotropic local law. We introduce the fundamental control parameter

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m(z)}{N\eta}} + \frac{1}{N\eta} \tag{3.8}$$

and the Stieltjes transform of the empirical eigenvalue density of  $X^* T^* T X$ ,

$$m_N(z) := \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} (R_N)_{\mu\mu}(z). \tag{3.9}$$

THEOREM 3.6 (ANISOTROPIC LOCAL LAW). Fix  $\tau > 0$ . Suppose that (2.9) and Assumption 2.1 hold. Suppose moreover that every edge  $k = 1, \dots, 2p$  satisfying  $a_k \geq \tau$  and every bulk component  $k = 1, \dots, p$  is regular in the sense of Definition 2.7. Then

$$\underline{\Sigma}^{-1} (G(z) - \Pi(z)) \underline{\Sigma}^{-1} = O_{\prec}(\Psi(z)) \tag{3.10}$$

and

$$m_N(z) - m(z) = O_{\prec}\left(\frac{1}{N\eta}\right) \quad (3.11)$$

uniformly in  $z \in \mathbf{D}$ .

(Note that the presence of the factors  $\underline{\Sigma}$  in (3.10) strengthens the result; by (2.7), they can be trivially dropped to obtain a weaker estimate. They ensure that the control of the error is stronger in directions where the covariance  $\Sigma$  is small.)

Outside the support of the asymptotic spectrum, one has stronger control all the way down to the real axis.

**THEOREM 3.7 (ANISOTROPIC LOCAL LAW OUTSIDE OF THE SPECTRUM).** *Fix  $\tau > 0$ . Suppose that (2.9) and Assumption 2.1 hold. Suppose moreover that every edge  $k = 1, \dots, 2p$  satisfying  $a_k \geq \tau$  is regular in the sense of Definition 2.7 (i). Then*

$$\underline{\Sigma}^{-1}(G(z) - \Pi(z))\underline{\Sigma}^{-1} = O_{\prec}\left(\sqrt{\frac{\operatorname{Im} m(z)}{N\eta}}\right) \quad (3.12)$$

uniformly in  $z \in [\tau, \tau^{-1}] \times (0, \tau^{-1}]$  satisfying  $\operatorname{dist}(E, \operatorname{supp} \varrho) \geq N^{-2/3+\tau}$ .

An explicit expression for the error term in (3.12) may be obtained from (A.7) below.

**REMARK 3.8.** Theorem 3.7 can be used to obtain a complete picture of the *outlier eigenvalues* of  $Q$  in the case where a bounded number of eigenvalues of  $\Sigma$  are changed to some arbitrary values (in particular possibly violating the regularity assumption from Definition 2.7 (i)). We note that the outliers may also lie between bulk components of  $\varrho$ . The analysis is similar to the one performed in [8] for the case  $\Sigma = I_M$ ; we omit the details.

In the remainder of this subsection, we state several corollaries of Theorem 3.6 where the assumption (2.9) is removed. From Theorem 3.6 it is not hard to deduce the following result on the resolvents  $R_N$  and  $R_M$ , defined in (3.1).

**COROLLARY 3.9 (ANISOTROPIC LOCAL LAWS FOR  $X^*T^*TX$  AND  $TXX^*T^*$ ).** *Fix  $\tau > 0$ . Suppose that Assumption 2.1 holds. Suppose moreover that every edge  $k = 1, \dots, 2p$  satisfying  $a_k \geq \tau$  and every bulk component  $k = 1, \dots, p$  is regular in the sense of Definition 2.7. Then*

$$R_N - m(z) = O_{\prec}(\Psi(z)) \quad (3.13)$$

and

$$\Sigma^{-1/2}\left(R_M - \frac{-1}{z(1 + m(z)\Sigma)}\right)\Sigma^{-1/2} = O_{\prec}(\Psi(z)) \quad (3.14)$$

uniformly in  $z \in \mathbf{D}$ . Moreover, (3.11) holds uniformly in  $z \in \mathbf{D}$ .

**REMARK 3.10.** Theorem 3.7 has an analogous corollary for  $R_N$  and  $R_M$  for  $z$  satisfying  $\operatorname{dist}(E, \operatorname{supp} \varrho) \geq N^{-2/3+\tau}$ , whereby the right-hand sides of (3.13) and (3.14) are replaced with the right-hand side of (3.12); we omit the precise statement.



REMARK 3.11. The following consistency check may be applied to the deterministic matrices on the left-hand sides of (3.13) and (3.14). If, in the identity  $\frac{1}{M} \text{Tr } R_M = \frac{1}{M} \text{Tr } R_N + \frac{\phi-1}{\phi} \frac{1}{z}$ , we replace  $\text{Tr } R_M$  and  $\text{Tr } R_N$  with the corresponding deterministic matrices from the left-hand sides of (3.13) and (3.14), we recover (2.11).

We conclude this subsection with another consequence of Theorem 3.6 – eigenvalue rigidity. We denote by

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{M \wedge N}$$

the nontrivial eigenvalues of  $Q$ , and by  $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_{M \wedge N}$  the *classical eigenvalue locations* defined through

$$N \int_{\gamma_i}^{\infty} d\varrho = i - 1/2. \quad (3.15)$$

If  $p \geq 1$ , it is convenient to relabel  $\lambda_i$  and  $\gamma_i$  separately for each bulk component  $k = 1, \dots, p$ . To that end, for  $k = 1, \dots, p$  we define the *classical number of eigenvalues in the  $k$ -th bulk component* through

$$N_k := N \int_{a_{2k}}^{a_{2k-1}} d\varrho. \quad (3.16)$$

In Lemma A.1 below we prove that  $N_k \in \mathbb{N}$  for  $k = 1, \dots, p$ . For  $k = 1, \dots, p$  and  $i = 1, \dots, N_k$  we introduce the relabellings

$$\lambda_{k,i} := \lambda_{i + \sum_{l < k} N_l}, \quad \gamma_{k,i} := \gamma_{i + \sum_{l < k} N_l} \in (a_{2k}, a_{2k-1}).$$

Note that we may also characterize  $\gamma_{k,i}$  through  $N \int_{\gamma_{k,i}}^{a_{2k-1}} d\varrho = i - 1/2$ .

THEOREM 3.12 (EIGENVALUE RIGIDITY). *Fix  $\tau > 0$ . Suppose that Assumption 2.1 holds. Suppose moreover that every edge  $k = 1, \dots, 2p$  satisfying  $a_k \geq \tau$  and every bulk component  $k = 1, \dots, p$  is regular in the sense of Definition 2.7. Then we have for all  $k = 1, \dots, p$  and  $i = 1, \dots, N_k$  satisfying  $\gamma_{k,i} \geq \tau$  that*

$$|\lambda_{k,i} - \gamma_{k,i}| \prec (i \wedge (N_k + 1 - i))^{-1/3} N^{-2/3}. \quad (3.17)$$

REMARK 3.13. As in [7, Theorem 2.8], Corollary 3.9 and Theorem 3.12 imply the complete delocalization, with respect to an arbitrary deterministic basis, of the eigenvectors of  $TX X^* T^*$  and  $X^* T^* T X$  associated with eigenvalues  $\lambda_i$  satisfying  $\gamma_i \geq \tau$ .

Note that Theorem 3.12 in particular implies an exact separation of the eigenvalues into connected components, whereby the number of eigenvalues in the  $k$ -th connected component is with high probability equal to the deterministic number  $N_k$ . This phenomenon of exact separation was first established in [1, 2].

**3.3. Individual spectral regions and edge universality.** In this subsection we elaborate on the results of Subsection 3.2 by only requiring the regularity of the edge or bulk component near which the spectral parameter  $z$  lies. To that end, for fixed  $\tau, \tau' > 0$  we define the subdomains

$$\begin{aligned} \mathbf{D}_k^e &\equiv \mathbf{D}_k^e(\tau, \tau', N) := \{z \in \mathbf{D}(\tau, N) : E \in [a_k - \tau', a_k + \tau']\} & (k = 1, \dots, 2p), \\ \mathbf{D}_k^b &\equiv \mathbf{D}_k^b(\tau, \tau', N) := \{z \in \mathbf{D}(\tau, N) : E \in [a_{2k} + \tau', a_{2k-1} - \tau']\} & (k = 1, \dots, p), \\ \mathbf{D}^o &\equiv \mathbf{D}^o(\tau, \tau', N) := \{z \in \mathbf{D}(\tau, N) : \text{dist}(E, \text{supp } \varrho) \geq \tau'\}. \end{aligned}$$

The superscripts  $e$ ,  $b$ , and  $o$  stand for “edge”, “bulk”, and “outside”, respectively.

Roughly, the results from Section 3.2 hold in each one of these smaller domains (instead of the full domain  $\mathbf{D}$ ) provided that just the associated edge (for  $z \in \mathbf{D}_k^e$ ) or bulk component (for  $z \in \mathbf{D}_k^b$ ) is regular. Results outside of the spectrum (for  $z \in \mathbf{D}^o$ ) do not need any regularity assumption.

**THEOREM 3.14 (REGULAR EDGE).** *Fix  $\tau > 0$ . Suppose that Assumption 2.1 holds. Suppose that the edge  $k = 1, \dots, 2p$  is regular in the sense of Definition 2.7 (i). Then there exists a constant  $\tau' > 0$ , depending only on  $\tau$ , such that the following holds.*

- (i) *Under the assumption (2.9) the estimates (3.10) and (3.11) hold uniformly in  $z \in \mathbf{D}_k^e$ .*
- (ii) *The estimates (3.13) and (3.14) hold uniformly in  $z \in \mathbf{D}_k^e$ .*
- (iii) *Let  $\hat{k} := \lfloor (k+1)/2 \rfloor$  be the bulk component to which the edge  $k$  belongs. Then for all  $i = 1, \dots, N_{\hat{k}}$  satisfying  $\gamma_{\hat{k},i} \in [a_k - \tau', a_k + \tau']$  we have*

$$|\lambda_{\hat{k},i} - \gamma_{\hat{k},i}| \prec (i \wedge (N_{\hat{k}} + 1 - i))^{-1/3} N^{-2/3}.$$

**THEOREM 3.15 (REGULAR BULK).** *Fix  $\tau, \tau' > 0$ . Suppose that Assumption 2.1 holds. Suppose that the bulk component  $k = 1, \dots, 2p$  is regular in the sense of Definition 2.7 (ii).*

- (i) *Under the assumption (2.9) the estimates (3.10) and (3.11) hold uniformly in  $z \in \mathbf{D}_k^b$ .*
- (ii) *The estimates (3.13) and (3.14) hold uniformly in  $z \in \mathbf{D}_k^b$ .*
- (iii) *Suppose that at least one of the two edges  $2k$  and  $2k-1$  is regular in the sense of Definition 2.7 (i). Then for all  $i = 1, \dots, N_k$  satisfying  $\gamma_{k,i} \in [a_{2k} + \tau', a_{2k-1} - \tau']$  we have  $|\lambda_{k,i} - \gamma_{k,i}| \prec N^{-1}$ .*

**THEOREM 3.16 (OUTSIDE OF THE SPECTRUM).** *Fix  $\tau, \tau' > 0$ . Suppose that Assumption 2.1 holds.*

- (i) *Under the assumption (2.9) the estimates (3.10) and (3.11) hold uniformly in  $z \in \mathbf{D}^o$ .*
- (ii) *The estimates (3.13) and (3.14) hold uniformly in  $z \in \mathbf{D}^o$ .*

**REMARK 3.17.** As in Section 3.2, if  $\text{dist}(E, \text{supp } \varrho) \geq N^{-2/3+\tau}$  then the error parameters  $\Psi(z)$  on the right-hand sides of (3.10), (3.13), and (3.14) in (i) and (ii) of Theorems 3.14 and 3.16 can be replaced with the smaller quantity  $\sqrt{\text{Im } m(z)/(N\eta)}$  and the lower bound  $\eta \geq N^{-1+\tau}$  relaxed to  $\eta > 0$ . (See Theorem 3.7.) We omit the detailed statement.

Finally, as a concrete application of the anisotropic local law, we establish the edge universality of  $Q$ . For a regular edge  $k = 1, \dots, 2p$  we define  $\varpi_k := (|f''(x_k)|/2)^{1/3}$ , which has the interpretation of the curvature of the density of  $\varrho$  near the edge  $k$ . In Lemma A.3 below we prove that if  $k$  is a regular edge then  $\varpi_k \asymp 1$ . For any fixed  $l \in \mathbb{N}$  and bulk component  $k = 1, \dots, p$  we define the random vectors

$$\begin{aligned} \mathbf{q}_{2k-1,l} &:= \frac{N^{2/3}}{\varpi_{2k-1}} (\lambda_{k,1} - a_{2k-1}, \dots, \lambda_{k,l} - a_{2k-1}), \\ \mathbf{q}_{2k,l} &:= -\frac{N^{2/3}}{\varpi_{2k}} (\lambda_{k,N_k} - a_{2k}, \dots, \lambda_{k,N_k-l+1} - a_{2k}). \end{aligned}$$

The interpretation of  $\mathbf{q}_k$  is the appropriately centred and normalized family of  $l$  extreme eigenvalues of  $Q$  near the edge  $k$ .

**THEOREM 3.18 (EDGE UNIVERSALITY).** *Fix  $\tau > 0$  and  $l \in \mathbb{N}$ . Suppose that Assumption 2.1 holds. Suppose that the edge  $k = 1, \dots, 2p$  is regular in the sense of Definition 2.7 (i). Then the distribution of  $\mathbf{q}_{k,l}$  depends asymptotically only on  $\pi$ . More precisely, for any fixed continuous bounded function  $h \in C_b(\mathbb{R}^l)$  there exists  $b(h, \pi) \equiv b_N(h, \pi)$ , depending only on  $\pi$ , such that  $\lim_{N \rightarrow \infty} (\mathbb{E}h(\mathbf{q}_{k,l}) - b(h, \pi)) = 0$ .*

Theorem 3.18 says in particular that the asymptotic distribution of  $\mathbf{q}_{k,l}$  does not depend on the distribution of the entries of  $X$ , on the left and right singular vectors of  $T$ , or on the dimensions of  $T$ . In particular, we may choose the entries of  $X$  to be Gaussian and  $T = T^*$  to be diagonal. We remark that Theorem 3.18 also holds for the joint distribution at several regular edges, i.e. for the random vector  $(\mathbf{q}_{k,l})_{k \in K}$ , where each  $k \in K$  is a regular edge.

Combining Theorem 3.18 with the works [11, 21, 26, 30] on the Gaussian case, we deduce convergence of  $\mathbf{q}_{k,l}$  to the *Tracy-Widom-Airy random vector*  $\mathbf{q}_l^\beta$ , where  $\beta = 1, 2$  is the customary symmetry index of random matrix theory, equal to 1 in the real symmetric case and 2 in the complex Hermitian case (see Section 2.1). Explicitly,  $\mathbf{q}_l^\beta$  may be expressed as the limit in distribution of the appropriately centred and scaled top  $l$  eigenvalues of a GOE/GUE matrix:

$$N^{2/3}(\mu_1^\beta - 2, \dots, \mu_l^\beta - 2) \xrightarrow{d} \mathbf{q}_l^\beta \quad (N \rightarrow \infty),$$

where  $\mu_1^\beta \geq \mu_2^\beta \geq \dots \geq \mu_N^\beta$  denote the eigenvalues of GOE ( $\beta = 1$ ) or GUE ( $\beta = 2$ ). For instance,  $\mathbf{q}_1^\beta \in \mathbb{R}$  is a random variable with Tracy-Widom- $\beta$  distribution [37, 38].

**COROLLARY 3.19 (TRACY-WIDOM-AIRY STATISTICS NEAR A REGULAR EDGE).** *Fix  $\tau > 0$  and  $l \in \mathbb{N}$ . Suppose that Assumption 2.1 holds.*

- (i) *Suppose that the rightmost edge  $k = 1$  is regular in the sense of Definition 2.7 (i). Then, for real symmetric  $Q$  ( $\beta = 1$ ) or complex Hermitian  $Q$  ( $\beta = 2$ ), the random vector  $\mathbf{q}_{1,l}$  converges in distribution to the Tracy-Widom-Airy random vector  $\mathbf{q}_l^\beta$ .*
- (ii) *Suppose that the edge  $k = 1, \dots, 2p$  is regular in the sense of Definition 2.7 (i). Then, for complex Hermitian  $Q$  ( $\beta = 2$ ), the random vector  $\mathbf{q}_{k,l}$  converges in distribution to the Tracy-Widom-Airy random vector  $\mathbf{q}_l^2$ .*

Like Theorem 3.18, Corollary 3.19 (ii) also holds for the joint distribution at several regular edges. In particular, for the case  $\beta = 2$  and under the assumption that the top and bottom edges of  $\varrho$  are regular, we obtain the universality of the condition number of  $Q$ .

Corollary 3.19 (i) for  $\beta = 1, 2$  was previously established in [26] under the assumption that  $\Sigma$  is diagonal, corresponding to uncorrelated population entries. Before that, Corollary 3.19 (i) for  $\beta = 2$  and diagonal  $\Sigma$  was established in [5], following the results of [11, 30] in the complex Gaussian case. Corollary 3.19 (ii) was recently established for Gaussian  $X$  in [21].

**3.4. The general anisotropic local law.** In this subsection we conclude the statement of our results with a general anisotropic local law, which takes the form of a black box yielding the anisotropic local law for  $Q$  assuming it has been established for a much simpler matrix. This latter result may be proved independently. In particular, this black box formulation may be used to establish the anisotropic local law in cases where the regularity assumptions from Definition 2.7 fail; we do not pursue such generalizations here. Aside from its great generality, this black box

formulation also makes precise the three Steps (A) – (C) mentioned in the introduction, which constitute the basic strategy of our proof.

We begin by introducing some basic terminology.

**DEFINITION 3.20 (LOCAL LAWS).** *We call a subset  $\mathbf{S} \equiv \mathbf{S}(N) \subset \mathbf{D}(\tau, N)$  a spectral domain if for each  $z \in \mathbf{S}$  we have  $\{w \in \mathbf{D} : \operatorname{Re} w = \operatorname{Re} z, \operatorname{Im} w \geq \operatorname{Im} z\} \subset \mathbf{S}$ .*

*Let  $\mathbf{S} \subset \mathbf{D}$  be a spectral domain.*

(i) *We say that the entrywise local law holds with parameters  $(X, \Sigma, \mathbf{S})$  if*

$$\left( \underline{\Sigma}^{-1} (G(z) - \Pi(z)) \underline{\Sigma}^{-1} \right)_{st} = O_{\prec}(\Psi(z)) \quad (3.18)$$

*uniformly in  $z \in \mathbf{S}$  and  $s, t \in \mathcal{I}$ .*

(ii) *We say that the anisotropic local law holds with parameters  $(X, \Sigma, \mathbf{S})$  if*

$$\underline{\Sigma}^{-1} (G(z) - \Pi(z)) \underline{\Sigma}^{-1} = O_{\prec}(\Psi(z)) \quad (3.19)$$

*uniformly in  $z \in \mathbf{S}$ .*

(iii) *We say that the averaged local law holds with parameters  $(X, \Sigma, \mathbf{S})$  if*

$$m_N(z) - m(z) = O_{\prec} \left( \frac{1}{N\eta} \right)$$

*uniformly in  $z \in \mathbf{S}$ .*

The main conclusion of this paper is that the anisotropic local law holds for general  $X$  and  $T$  provided that the entrywise local law holds for Gaussian  $X$  and diagonal  $T$ . This latter case may be established independently, as we illustrate in Section 5 and Appendix A.

Aside from Assumption 2.1, the only assumption that we shall need is

$$|1 + m(z)\sigma_i| \geq \tau \quad \text{for all } z \in \mathbf{S} \text{ and } i \in \mathcal{I}_M. \quad (3.20)$$

This assumption holds for instance under the regularity assumptions of Definition 2.7 (see Lemmas A.4, A.6, and A.8 below). Clearly, we always have  $|1 + m(z)\sigma_i| > 0$  (see (2.11)), and (3.20) is a uniform version of this bound. Generally, the assumption (3.20) is necessary to guarantee that the generalized matrix entries of  $(Q - z)^{-1}$  (or, alternatively, of  $G(z)$ ) remain bounded. Indeed, in Corollary 3.9 we saw that the generalized entries of  $(Q - z)^{-1}$  are close to those of  $-z^{-1}(1 + m(z)\Sigma)^{-1}$ .

**THEOREM 3.21 (GENERAL LOCAL LAW).** *Fix  $\tau > 0$ . Suppose that  $X$  and  $\Sigma$  satisfy (2.9) and Assumption 2.1. Let  $X^{\text{Gauss}}$  be a Gaussian matrix satisfying (2.4). Let  $\mathbf{S} \subset \mathbf{D}$  be a spectral domain, and suppose that (3.20) holds. Define the diagonalization of  $\Sigma$  through*

$$D \equiv D(\Sigma) := \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_M). \quad (3.21)$$

(i) *If the entrywise local law holds with parameters  $(X^{\text{Gauss}}, D, \mathbf{S})$ , then the anisotropic local law holds with parameters  $(X, \Sigma, \mathbf{S})$ .*

(ii) If the entrywise local law and the averaged local law hold with parameters  $(X^{\text{Gauss}}, D, \mathbf{S})$ , then the averaged local law holds with parameters  $(X, \Sigma, \mathbf{S})$ .

The next theorem shows that the hypotheses in (i) and (ii) of Theorem 3.21 may be verified under a *stability condition* on the spectrum of  $\Sigma$ , made precise in Definition 5.4 below.

**THEOREM 3.22 (GENERAL CONDITIONS FOR LOCAL LAW WITH DIAGONAL  $\Sigma$ ).** Fix  $\tau > 0$ . Let  $\mathbf{S} \subset \mathbf{D}$  be a spectral domain. Suppose that  $\Sigma$  is diagonal and that (2.9), (3.20), and Assumption 2.1 hold. Moreover, suppose that the equation (2.11) is stable on  $\mathbf{S}$  in the sense of Definition 5.4 below. Then the entrywise local law holds with parameters  $(X, \Sigma, \mathbf{S})$ , and the averaged local law holds with parameters  $(X, \Sigma, \mathbf{S})$ .

Thus, the Steps (A) – (C) outlined in the introduction may be expressed as follows. Let  $\mathbf{S}$  be a spectral domain (e.g.  $\mathbf{S} = \mathbf{D}_k^e$  for some edge  $k$ ). First, we prove that (3.20) holds and that (2.11) is stable on  $\mathbf{S}$  in the sense of Definition 5.4 below. Then we use Theorem 3.22 to obtain the entrywise and averaged local laws with parameters  $(X^{\text{Gauss}}, D(\Sigma), \mathbf{S})$ . We then feed this result into Theorem 3.21 to obtain the anisotropic and averaged local laws with parameters  $(X, \Sigma, \mathbf{S})$ .

**3.5. Outline of the paper.** The bulk of this paper is devoted to the proof of the anisotropic local law, which is the content of Sections 4–8. For clarity of presentation, we first give all proofs under the assumption (2.9), and subsequently explain how to remove it. In Section 4 we collect the basic tools that we shall use throughout the proofs; they consist of basic identities and estimates for the matrix  $G$ . In Section 5 we perform Step (A) of the proof, by proving the entrywise local law under the assumption that  $\Sigma$  is diagonal (Theorem 3.22). In Section 6 we perform Step (B) of the proof, by proving the anisotropic local law for general  $\Sigma$  and Gaussian  $X$  (Theorem 3.21 (i) for Gaussian  $X$ ). The main step, Step (C), of the proof is the content of Sections 7 and 8. In Section 7 we explain the main ideas of the self-consistent comparison method and complete the proof under the additional assumption that  $\mathbb{E}X_{i\mu}^3 = 0$ . In Section 8 we give the additional arguments needed to complete the proof for general  $X$ . Together, these two sections complete the proof of Theorem 3.21 (i) for general  $X$ .

Having completed the proof of the anisotropic local law, we prove the averaged local law (Theorem 3.21 (ii)) in Section 9. This will conclude the proof of Theorems 3.21 and 3.22. At the end of Section 9, we explain how to deduce Theorem 3.6, Corollary 3.9, and Theorems 3.14 (i)–(ii), 3.15 (i)–(ii), and 3.16.

In Section 10 we prove the rigidity of the eigenvalues (Theorems 3.12, 3.14 (iii), and 3.15 (iii)) and the universality of their joint distribution near the edges (Theorem 3.18). Next, in Section 11 we explain how to remove the assumption (2.9) and how to extend all of our results from the matrix  $Q$  to the matrix  $\dot{Q}$ .

In Section 12, as a further illustration of the self-consistent comparison method, we present and prove analogous results for deformed Wigner matrices.

The first part of Appendix A is devoted to the proof the basic properties of the asymptotic eigenvalue density  $\varrho$  stated in Section 2.2. In the second part of Appendix A, we establish the two key assumptions of Theorem 3.22 – (3.20) and the stability of (2.11) – on each subdomain  $\mathbf{D}_k^e$ ,  $\mathbf{D}_k^b$ , and  $\mathbf{D}^o$  separately, under the regularity assumptions from Definition 2.7.

## 4. Basic tools

The rest of this paper is devoted to the proofs. In this preliminary section we collect various identities from linear algebra and simple estimates that we shall use throughout the paper.

We always use the following convention for matrix multiplication.

DEFINITION 4.1 (MATRIX MULTIPLICATION). *We use matrices of the form  $A = (A_{st} : s \in l(A), t \in r(A))$ , whose entries are indexed by arbitrary finite subsets of  $l(A), r(A) \subset \mathbb{N}$ . Matrix multiplication  $AB$  is defined for  $s \in l(A)$  and  $t \in r(B)$  by*

$$(AB)_{st} := \sum_{q \in r(A) \cap l(B)} A_{sq} B_{qt}.$$

DEFINITION 4.2. *Suppose (2.9). Define the  $\mathcal{I} \times \mathcal{I}$  matrices*

$$G(z) := H(z)^{-1}, \quad H(z) := \begin{pmatrix} -\Sigma^{-1} & X \\ X^* & -z \end{pmatrix},$$

*as well as the  $\mathcal{I}_M \times \mathcal{I}_M$  matrix*

$$G_M(z) := (-\Sigma^{-1} + z^{-1}XX^*)^{-1} = z\Sigma^{1/2}(\Sigma^{1/2}XX^*\Sigma^{1/2} - z)^{-1}\Sigma^{1/2} \quad (4.1)$$

*and the  $\mathcal{I}_N \times \mathcal{I}_N$  matrix*

$$G_N(z) := (X^*\Sigma X - z)^{-1}. \quad (4.2)$$

*Throughout the following we frequently omit the argument  $z$  from our notation.*

Since  $H(z)$  and  $G(z)$  are only defined under the assumption (2.9), we shall always tacitly assume (2.9) whenever we use them. Note that under the assumption (2.9) we have  $R_N = G_N$ .

DEFINITION 4.3 (MINORS). *For  $S \subset \mathcal{I}$  we define the minor  $H^{(S)} := (H_{st} : s, t \in \mathcal{I} \setminus S)$ . We also write  $G^{(S)} := (H^{(S)})^{-1}$ . The matrices  $G_N^{(S)}$  and  $G_M^{(S)}$  are defined similarly. We abbreviate  $(\{s\}) \equiv (s)$  and  $(\{s, t\}) \equiv (st)$ .*

LEMMA 4.4 (RESOLVENT IDENTITIES). (i) *We have*

$$G = \begin{pmatrix} \Sigma X G_N X^* \Sigma - \Sigma & \Sigma X G_N \\ G_N X^* \Sigma & G_N \end{pmatrix} = \begin{pmatrix} G_M & z^{-1} G_M X \\ z^{-1} X^* G_M & z^{-2} X^* G_M X - z^{-1} \end{pmatrix}. \quad (4.3)$$

(ii) *For  $\mu \in \mathcal{I}_N$  we have*

$$\frac{1}{G_{\mu\mu}} = -z - (X^* G^{(\mu)} X)_{\mu\mu}, \quad (4.4)$$

*and for  $\mu \neq \nu \in \mathcal{I}_N$*

$$G_{\mu\nu} = -G_{\mu\mu} (X^* G^{(\mu)})_{\mu\nu} = -G_{\nu\nu} (G^{(\nu)} X)_{\mu\nu} = G_{\mu\mu} G_{\nu\nu}^{(\mu)} (X^* G^{(\mu\nu)} X)_{\mu\nu}. \quad (4.5)$$

(iii) Suppose that  $\Sigma$  is diagonal. Then for  $i \in \mathcal{I}_M$  we have

$$\frac{1}{G_{ii}} = -\frac{1}{\sigma_i} - (XG^{(i)}X^*)_{ii}, \quad (4.6)$$

and for  $i \neq j \in \mathcal{I}_M$

$$G_{ij} = -G_{ii}(XG^{(i)})_{ij} = -G_{jj}(G^{(j)}X^*)_{ij} = G_{ii}G_{jj}^{(i)}(XG^{(ij)}X^*)_{ij}. \quad (4.7)$$

(iv) For  $i \in \mathcal{I}_M$  and  $\mu \in \mathcal{I}_N$  we have

$$G_{i\mu} = -G_{\mu\mu}(G^{(\mu)}X)_{i\mu}, \quad G_{\mu i} = -G_{\mu\mu}(X^*G^{(\mu)})_{\mu i}. \quad (4.8)$$

In addition, if  $\Sigma$  is diagonal, we have

$$G_{i\mu} = -G_{ii}(XG^{(i)})_{i\mu} = G_{ii}G_{\mu\mu}^{(i)}(-X_{i\mu} + (XG^{(i\mu)}X)_{i\mu}), \quad (4.9)$$

$$G_{\mu i} = -G_{ii}(G^{(i)}X)_{\mu i} = G_{\mu\mu}G_{ii}^{(\mu)}(-X_{\mu i}^* + (X^*G^{(\mu i)}X^*)_{\mu i}). \quad (4.10)$$

(v) For  $r \in \mathcal{I}$  and  $s, t \in \mathcal{I} \setminus \{k\}$  we have

$$G_{st}^{(r)} = G_{st} - \frac{G_{sr}G_{rt}}{G_{rr}} \quad (4.11)$$

(vi) All of the identities from (i)–(v) hold for  $G^{(S)}$  instead of  $G$  if  $S \subset \mathcal{I}_N$  or  $S \subset \mathcal{I}$  and  $\Sigma$  is diagonal.

PROOF. The identities (4.3), (4.4), and (4.6) follow from Schur's complement formula. The remaining identities follow easily from resolvent identities that have been previously derived in [13, 19]; they are summarized e.g. in [14, Lemma 4.5].  $\square$

Next, we introduce the spectral decomposition of  $G$ . We use the notation

$$\Sigma^{1/2}X = \sum_{k=1}^{M \wedge N} \sqrt{\lambda_k} \xi_k \zeta_k^*$$

for the singular value decomposition of  $\Sigma^{1/2}X$ , where

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{M \wedge N} \geq 0 = \lambda_{M \wedge N + 1} = \dots = \lambda_{N \vee M},$$

and  $\{\xi_k\}_{k=1}^M$  and  $\{\zeta_k\}_{k=1}^N$  are orthonormal bases of  $\mathbb{R}^{\mathcal{I}_M}$  and  $\mathbb{R}^{\mathcal{I}_N}$  respectively. Then for  $\mu, \nu \in \mathcal{I}_N$  we have

$$G_{\mu\nu} = \sum_{k=1}^N \frac{\zeta_k(\mu) \overline{\zeta_k(\nu)}}{\lambda_k - z}, \quad (4.12)$$

and for  $i, j \in \mathcal{I}_M$

$$G_{ij} = z \sum_{k=1}^M \frac{(\Sigma^{1/2}\xi_k)(i) \overline{(\Sigma^{1/2}\xi_k)(j)}}{\lambda_k - z} = -\Sigma_{ij} + \sum_{k=1}^M \frac{\lambda_k (\Sigma^{1/2}\xi_k)(i) \overline{(\Sigma^{1/2}\xi_k)(j)}}{\lambda_k - z}. \quad (4.13)$$



Moreover, for  $i \in \mathcal{I}_M$  and  $\mu \in \mathcal{I}_N$  we have

$$G_{i\mu} = \sum_{k=1}^{M \wedge N} \frac{\sqrt{\lambda_k} (\Sigma^{1/2} \boldsymbol{\xi}_k)(i) \overline{\zeta_k(\mu)}}{\lambda_k - z}, \quad G_{\mu i} = \sum_{k=1}^{M \wedge N} \frac{\sqrt{\lambda_k} \zeta_k(\mu) \overline{(\Sigma^{1/2} \boldsymbol{\xi}_k)(i)}}{\lambda_k - z}. \quad (4.14)$$

Summarizing, defining

$$\mathbf{u}_k := \begin{pmatrix} \mathbf{1}(k \leq M) \sqrt{\lambda_k} \boldsymbol{\xi}_k \\ \mathbf{1}(k \leq N) \zeta_k \end{pmatrix} \in \mathbb{R}^{\mathcal{I}},$$

we have

$$G = -\underline{\Sigma} + \underline{\Sigma}^{1/2} \sum_{k=1}^{N \vee M} \frac{\mathbf{u}_k \mathbf{u}_k^*}{\lambda_k - z} \underline{\Sigma}^{1/2}. \quad (4.15)$$

DEFINITION 4.5 (GENERALIZED ENTRIES). For  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{\mathcal{I}}$ ,  $s \in \mathcal{I}$ , and an  $\mathcal{I} \times \mathcal{I}$  matrix  $A$ , we abbreviate

$$A_{\mathbf{v}\mathbf{w}} := \langle \mathbf{v}, A\mathbf{w} \rangle, \quad A_{\mathbf{v}s} := \langle \mathbf{v}, A\mathbf{e}_s \rangle, \quad A_{s\mathbf{v}} := \langle \mathbf{e}_s, A\mathbf{v} \rangle,$$

where  $\mathbf{e}_s$  denotes the standard unit vector in the coordinate direction  $s$ .

We sometimes identify vectors  $\mathbf{v} \in \mathbb{R}^{\mathcal{I}_M}$  and  $\mathbf{w} \in \mathbb{R}^{\mathcal{I}_N}$  with their natural embeddings  $\begin{pmatrix} \mathbf{v} \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} 0 \\ \mathbf{w} \end{pmatrix}$  in  $\mathbb{R}^{\mathcal{I}}$ . The following result is our fundamental tool for estimating entries of  $G$ .

LEMMA 4.6. Fix  $\tau > 0$ . Then the following estimates hold for any  $z \in \mathbf{D}$ . We have

$$\|\underline{\Sigma}^{-1/2} G \underline{\Sigma}^{-1/2}\| \leq C\eta^{-1}, \quad \|\underline{\Sigma}^{-1/2} \partial_z G \underline{\Sigma}^{-1/2}\| \leq C\eta^{-2}. \quad (4.16)$$

Furthermore, let  $\mathbf{v} \in \mathbb{R}^{\mathcal{I}_M}$  and  $\mathbf{w} \in \mathbb{R}^{\mathcal{I}_N}$ . Then we have the bounds

$$\sum_{\mu \in \mathcal{I}_N} |G_{\mathbf{w}\mu}|^2 = \frac{\operatorname{Im} G_{\mathbf{w}\mathbf{w}}}{\eta}, \quad (4.17)$$

$$\sum_{i \in \mathcal{I}_M} |G_{\mathbf{v}i}|^2 \leq \frac{C\|X^*X\|}{\eta} \operatorname{Im} G_{\mathbf{v}\mathbf{v}} + 2(\Sigma^2)_{\mathbf{v}\mathbf{v}}, \quad (4.18)$$

$$\sum_{i \in \mathcal{I}_M} |G_{\mathbf{w}i}|^2 \leq C\|X^*X\| \sum_{\mu \in \mathcal{I}_N} |G_{\mathbf{w}\mu}|^2, \quad (4.19)$$

$$\sum_{\mu \in \mathcal{I}_N} |G_{\mathbf{v}\mu}|^2 \leq C\|X^*X\| \sum_{i \in \mathcal{I}_M} |G_{\mathbf{v}i}|^2. \quad (4.20)$$

Finally, the estimates (4.16)–(4.20) remain true for  $G^{(S)}$  instead of  $G$  if  $S \subset \mathcal{I}_N$  or  $S \subset \mathcal{I}$  and  $\Sigma$  is diagonal.

PROOF. The estimates (4.16) follow from (4.15), using  $\|A\| = \sup\{|\langle \mathbf{x}, A\mathbf{y} \rangle| : |\mathbf{x}|, |\mathbf{y}| \leq 1\}$ , (2.7), and  $|\lambda_k|/|\lambda_k - z| \leq C\eta^{-1}$  which follows from the bound  $|z| \asymp 1$ . Moreover, (4.17) easily follows from (4.2), (4.3), and the spectral decomposition (4.12).

In order to prove (4.18), we use (4.3) to write

$$\begin{aligned} \sum_{i \in \mathcal{I}_M} |G_{\mathbf{v}i}|^2 &= \sum_{i \in \mathcal{I}_M} |(\Sigma X G X^* \Sigma)_{\mathbf{v}i} - \Sigma_{\mathbf{v}i}|^2 \leq 2(\Sigma X G X^* \Sigma^2 X G^* X^* \Sigma)_{\mathbf{v}\mathbf{v}} + 2(\Sigma^2)_{\mathbf{v}\mathbf{v}} \\ &\leq C\|X^*X\|(\Sigma X G_N G_N^* X^* \Sigma)_{\mathbf{v}\mathbf{v}} + 2(\Sigma^2)_{\mathbf{v}\mathbf{v}} = \frac{C\|X^*X\|}{\eta} \operatorname{Im}(\Sigma X G X^* \Sigma)_{\mathbf{v}\mathbf{v}} + 2(\Sigma^2)_{\mathbf{v}\mathbf{v}} \\ &= \frac{C\|X^*X\|}{\eta} \operatorname{Im} G_{\mathbf{v}\mathbf{v}} + 2(\Sigma^2)_{\mathbf{v}\mathbf{v}}, \end{aligned}$$

which is the claim.

In order to prove (4.19), we use (4.3) and (4.12) to get

$$\sum_{i \in \mathcal{I}_M} |G_{\mathbf{w}i}|^2 = (GX^* \Sigma^2 XG^*)_{\mathbf{w}\mathbf{w}} \leq C(GX^* XG^*)_{\mathbf{w}\mathbf{w}} \leq C\|X^* X\|(G_N G_N^*)_{\mathbf{w}\mathbf{w}}.$$

The estimate (4.20) is proved similarly:

$$\sum_{\mu \in \mathcal{I}_N} |G_{\mathbf{v}\mu}|^2 = |z|^{-2} (GXX^* G^*)_{\mathbf{v}\mathbf{v}} \leq C\|X^* X\|(G_M G_M^*)_{\mathbf{v}\mathbf{v}}.$$

Finally, the same estimates for  $G^{(S)}$  instead of  $G$  follow using a trivial modification of the above argument.  $\square$

DEFINITION 4.7. *We say that an event  $\Xi$  holds with high probability if  $1 - \mathbf{1}(\Xi) \prec 0$ .*

The following result may be used to estimate the factors  $\|X^* X\|$  in Lemma 4.6 with high probability. It follows from [7, Theorem 2.10].

LEMMA 4.8. *Under the assumptions (2.1), (2.4), and (2.5), there exists a constant  $C > 0$  such that  $\|X^* X\| \leq C$  with high probability.*

Using Lemma 4.8, we observe that we may improve (4.16) provided we settle for a high-probability instead of a deterministic statement.

LEMMA 4.9. *We have the bounds*

$$\|\underline{\Sigma}^{-1} (G + \underline{\Sigma}) \underline{\Sigma}^{-1}\| \leq C\|X^* X\|\eta^{-1}, \quad \|\underline{\Sigma}^{-1} \partial_z G \underline{\Sigma}^{-1}\| \leq C\|X^* X\|\eta^{-2}$$

for all  $z \in \mathbf{D}$ .

PROOF. The claim is an easy consequence of the first identity of (4.3) combined with (4.2).  $\square$

We conclude this section with the following basic properties of  $m$ , which can be proved as in [5] and the references therein.

LEMMA 4.10 (GENERAL PROPERTIES OF  $m$ ). *Fix  $\tau > 0$  and suppose that (2.3), (2.7), and (2.8) hold. Then there exists a constant  $C > 0$  such that*

$$C^{-1} \leq |m(z)| \leq C \tag{4.21}$$

and

$$\operatorname{Im} m(z) \geq C^{-1} \eta \tag{4.22}$$

for all  $z \in \mathbb{C}_+$  satisfying  $\tau \leq |z| \leq \tau^{-1}$ .

In particular, from the upper bound in (4.21) we deduce that  $\varrho$  has a bounded density on  $[\tau, \infty)$ .

## 5. Entrywise local law for diagonal $\Sigma$

In this section we prove Theorem 3.22, hence performing Step (A) of the proof mentioned in the introduction. The proof of Theorem 3.22 is similar to previous proofs of local entrywise laws, such as [7, 32]. We follow the basic approach of [7, Section 4], and only give the details where the argument departs significantly from that of [7].

The main novel observation of this section is that the equation (2.11) arises very easily from the random matrix model by a double application of Schur's complement formula. Heuristically, this may be seen using the identities (4.4) and (4.6). Indeed, suppose that  $G_{\mu\mu} \approx m$  for  $\mu \in \mathcal{I}_N$ . We ignore the random fluctuations in (4.4) to get

$$\frac{1}{m} \approx \frac{1}{G_{\mu\mu}} = -z - (X^* G^{(\mu)} X)_{\mu\mu} \approx -z - \frac{1}{N} \sum_{i \in \mathcal{I}_M} G_{ii}^{(\mu)} \approx -z - \frac{1}{N} \sum_{i \in \mathcal{I}_M} G_{ii}. \quad (5.1)$$

Similarly, ignoring the random fluctuations in (4.6), we get

$$\frac{1}{G_{ii}} = -\frac{1}{\sigma_i} - (X G^{(i)} X^*)_{ii} \approx -\frac{1}{\sigma_i} - \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} G_{\mu\mu}^{(i)} \approx -\frac{1}{\sigma_i} - \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} G_{\mu\mu} \approx -\frac{1}{\sigma_i} - m. \quad (5.2)$$

Plugging (5.2) into (5.1) yields (2.11). In this section we give a rigorous justification of these approximations.

**5.1. Weak entrywise law.** In this subsection we establish the following weaker version of Theorem 3.22. It is analogous to [7, Proposition 4.2].

**PROPOSITION 5.1 (WEAK ENTRYWISE LAW).** *Suppose that the assumptions of Theorem 3.22 hold. Then  $\Lambda \prec (N\eta)^{-1/4}$  uniformly in  $z \in \mathbf{S}$ .*

The rest of this subsection is devoted to the proof of Proposition 5.1. For each  $i \in \mathcal{I}_M$  we define

$$m_i := \frac{-\sigma_i}{1 + m\sigma_i}. \quad (5.3)$$

Recalling (2.11), we find that the functions  $m$  and  $m_i$  satisfy

$$\frac{1}{m} = -z - \frac{1}{N} \sum_{i \in \mathcal{I}_M} m_i, \quad \frac{1}{m_i} = -\frac{1}{\sigma_i} - m.$$

Note that (3.20) implies

$$|m_i| \leq C\sigma_i \quad \text{for } z \in \mathbf{S} \text{ and } i \in \mathcal{I}_M. \quad (5.4)$$

Next, we define the random control parameters

$$\Lambda := \max_{s,t \in \mathcal{I}} |(\underline{\Sigma}^{-1}(G(z) - \Pi(z))\underline{\Sigma}^{-1})_{st}|, \quad \Lambda_o := \max_{s \neq t \in \mathcal{I}} |(\underline{\Sigma}^{-1}G(z)\underline{\Sigma}^{-1})_{st}|.$$

We extend the definitions of  $\sigma_i$  and  $m_i$  for  $i \in \mathcal{I}_M$  by setting  $\sigma_\mu := 1$  and  $m_\mu := m$  for  $\mu \in \mathcal{I}_N$ . We may therefore write

$$\Lambda = \max_{s,t \in \mathcal{I}} \frac{|G_{st} - \delta_{st}m_s|}{\sigma_s\sigma_t}, \quad \Lambda_o = \max_{s \neq t \in \mathcal{I}} \frac{|G_{st}|}{\sigma_s\sigma_t}.$$

Moreover, we define the averaged control parameters

$$\Theta := \Theta_M + \Theta_N, \quad \Theta_M := \left| \frac{1}{M} \sum_{i \in \mathcal{I}_M} (G_{ii} - m_i) \right|, \quad \Theta_N := \left| \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} (G_{\mu\mu} - m) \right| = |m_N - m|.$$

We have the trivial bound

$$\Theta \leq C\Lambda. \quad (5.5)$$

For  $s \in \mathcal{I}$  we introduce the conditional expectation

$$\mathbb{E}_s[\cdot] := \mathbb{E}[\cdot | H^{(s)}]. \quad (5.6)$$

Using (4.6) we get for  $i \in \mathcal{I}_M$

$$\frac{1}{G_{ii}} = -\frac{1}{\sigma_i} - \frac{1}{N} \text{Tr } G_N^{(i)} - Z_i, \quad Z_i := (1 - \mathbb{E}_i)(XG^{(i)}X^*)_{ii}, \quad (5.7)$$

and using (4.4) we get for  $\mu \in \mathcal{I}_N$

$$\frac{1}{G_{\mu\mu}} = -z - \frac{1}{N} \text{Tr } G_M^{(\mu)} - Z_\mu, \quad Z_\mu := (1 - \mathbb{E}_\mu)(X^*G^{(\mu)}X)_{\mu\mu}. \quad (5.8)$$

In analogy to [7, Section 4], we define the  $z$ -dependent event  $\Xi := \{\Lambda \leq (\log N)^{-1}\}$  and the control parameter

$$\Psi_\Theta := \sqrt{\frac{\text{Im } m + \Theta}{N\eta}}.$$

The following estimate is analogous to [7, Lemma 4.4].

LEMMA 5.2. *Suppose that the assumptions of Theorem 3.22 hold. Then for  $s \in \mathcal{I}$  and  $z \in \mathbf{S}$  we have*

$$\mathbf{1}(\Xi)(|Z_s| + \Lambda_o) \prec \Psi_\Theta \quad (5.9)$$

as well as

$$\mathbf{1}(\eta \geq 1)(|Z_s| + \Lambda_o) \prec \Psi_\Theta. \quad (5.10)$$

PROOF. The proof relies on the identities from Lemma 4.4 and large deviation estimates, like that of [7, Lemma 4.4] and [32, Theorems 6.8 and 6.9]. Note first that (5.4), combined with (4.21) and (5.3), yields

$$|m_s| \asymp \sigma_s \quad (s \in \mathcal{I}). \quad (5.11)$$

Using (4.11) and a simple induction argument, it is not hard to conclude that

$$\mathbf{1}(\Xi)|G_{tt}^{(S)}| \asymp \sigma_t \quad (5.12)$$

for any  $S \subset \mathcal{I}$  and  $t \in \mathcal{I} \setminus S$  satisfying  $|S| \leq C$ .

Let us first estimate  $\Lambda_o$  in (5.9). We shall in fact prove that

$$\mathbf{1}(\Xi)|G_{st}| \prec \sigma_s \sigma_t \left( \frac{\text{Im } m + \Theta + \Lambda_o^2}{N\eta} \right)^{1/2} \quad (5.13)$$

for all  $s \neq t \in \mathcal{I}$ . From (5.13) it is easy to deduce that  $\mathbf{1}(\Xi)\Lambda_o \prec \Psi_\Theta$ .

Let us start with  $G_{ij}$  for  $i \neq j \in \mathcal{I}_M$ . Using (4.7), (5.12), and a large deviation estimate (see [7, Lemma 3.1]), we find

$$\mathbf{1}(\Xi)|G_{ij}| \leq \mathbf{1}(\Xi)|G_{ii}G_{jj}^{(i)}| \left| \sum_{\mu, \nu \in \mathcal{I}_N} X_{i\mu} G_{\mu\nu}^{(ij)} X_{\nu j}^* \right| \prec \mathbf{1}(\Xi)\sigma_i\sigma_j \left( \frac{1}{N^2} \sum_{\mu, \nu \in \mathcal{I}_N} |G_{\mu\nu}^{(ij)}|^2 \right)^{1/2}. \quad (5.14)$$

The term in parentheses is

$$\mathbf{1}(\Xi) \frac{1}{N^2} \sum_{\mu, \nu \in \mathcal{I}_N} |G_{\mu\nu}^{(ij)}|^2 = \mathbf{1}(\Xi) \frac{1}{N^2\eta} \sum_{\mu \in \mathcal{I}_N} \text{Im } G_{\mu\mu}^{(ij)} \prec \frac{1}{N^2\eta} \sum_{\mu \in \mathcal{I}_N} \text{Im } G_{\mu\mu} + \frac{\Lambda_o^2}{N\eta} \leq \frac{\text{Im } m + \Theta_N + \Lambda_o^2}{N\eta},$$

where in the first step we used (4.11) and (5.12). This yields (5.13) for  $s, t \in \mathcal{I}_M$ .

Next,  $\mathbf{1}(\Xi)G_{\mu\nu}$  for  $\mu \neq \nu$  is estimated similarly, using (4.5), (4.18), (4.22), and the bound  $\text{Im } m_i \leq C\sigma_i^2 \text{Im } m$  for all  $i \in \mathcal{I}_M$ , as follows easily from (5.3). Finally,  $\mathbf{1}(\Xi)G_{i\mu}$  with  $i \in \mathcal{I}_M$  and  $\mu \in \mathcal{I}_N$  is estimated similarly, using (4.9), (4.19), Lemma 4.8, and (4.17). This concludes the estimate of  $\mathbf{1}(\Xi)\Lambda_o$ .

An analogous argument for  $Z_s$  (see e.g. [14, Lemma 5.2]) completes the proof of (5.9).

In order to prove (5.10), we proceed similarly. For  $\eta \geq 1$ , we proceed as above to get  $|Z_s| \prec N^{-1/2}$ , where we used that  $\|G^{(s)}\| \leq C$  by (4.16). Similarly, as in (5.14) we get

$$|G_{ij}| \leq |G_{ii}G_{jj}^{(i)}| \left| \sum_{\mu, \nu \in \mathcal{I}_N} X_{i\mu} G_{\mu\nu}^{(ij)} X_{\nu j}^* \right| \prec |G_{ii}G_{jj}^{(i)}| N^{-1/2},$$

where in the last step we used that  $\text{Im } G_{\mu\mu}^{(ij)} \leq C$ , by (4.16). Moreover, from (4.1) and (4.3) we get immediately that  $|G_{ii}| \leq C\sigma_i$ , and a similar argument for  $G^{(i)}$  implies that  $|G_{jj}^{(i)}| \leq C\sigma_j$ . This concludes the proof.  $\square$

Recall the definition of  $m_N$  from (3.9). From (5.7) combined with (4.11) and Lemma 5.2 we get, for  $i \in \mathcal{I}_M$  and  $z \in \mathbf{S}$ ,

$$\mathbf{1}(\Xi)G_{ii} = \mathbf{1}(\Xi) \frac{-\sigma_i}{1 + m_N\sigma_i + \sigma_i Z_i + O_{\prec}(\sigma_i \Psi_\Theta^2)}. \quad (5.15)$$

Similarly, from (5.8) we get, for  $\mu \in \mathcal{I}_N$  and  $z \in \mathbf{S}$ ,

$$\mathbf{1}(\Xi) \frac{1}{G_{\mu\mu}} = \mathbf{1}(\Xi) \left( -z - \frac{1}{N} \sum_{i \in \mathcal{I}_M} G_{ii} - Z_\mu + O_{\prec}(\Psi_\Theta^2) \right). \quad (5.16)$$

As in [7, Lemma 4.7], it is easy to derive from (5.16) and Lemma 5.2 that

$$\mathbf{1}(\Xi)|G_{\mu\mu} - m_N| \prec \Psi_\Theta \quad (5.17)$$

for  $\mu \in \mathcal{I}_N$  and  $z \in \mathbf{S}$ . Hence, expanding  $G_{\mu\mu} = m_N + (G_{\mu\mu} - m_N)$  and using (4.21) yields

$$\mathbf{1}(\Xi) \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} \frac{1}{G_{\mu\mu}} = \mathbf{1}(\Xi) \frac{1}{m_N} + O_{\prec}(\Psi_\Theta^2).$$

Plugging this and (5.15) into (5.16) yields

$$\mathbf{1}(\Xi) \frac{1}{m_N} = \mathbf{1}(\Xi) \left( -z + \frac{1}{N} \sum_{i \in \mathcal{I}_M} \frac{\sigma_i}{1 + m_N \sigma_i + \sigma_i Z_i + O_{\prec}(\sigma_i \Psi_{\Theta}^2)} - \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} Z_{\mu} + O_{\prec}(\Psi_{\Theta}^2) \right). \quad (5.18)$$

From (5.18), Lemma 5.2, and the estimate  $\mathbf{1}(\Xi)|1 + m_N \sigma_1| \geq c$  (as follows from (3.20)), we conclude the following result.

LEMMA 5.3. *Suppose that the assumptions of Theorem 3.22 hold. Define*

$$[Z]_N := \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} Z_{\mu}, \quad [Z]_M := \frac{1}{N} \sum_{i \in \mathcal{I}_M} \frac{\sigma_i^2}{(1 + m_N \sigma_i)^2} Z_i.$$

*Then for  $z \in \mathbf{S}$  we have*

$$\mathbf{1}(\Xi)(f(m_N) - z) = \mathbf{1}(\Xi)([Z]_N + [Z]_M + O_{\prec}(\Psi_{\Theta}^2)), \quad (5.19)$$

*where  $f$  was defined in (2.12).*

Next, we give the precise stability condition for the equation (2.11). Roughly, it says that if  $f(u(z)) - z$  is small and  $u(\tilde{z}) - m(\tilde{z})$  is small for  $\tilde{z} := z + iN^{-5}$ , then  $u(z) - m(z)$  is small. (Recall that  $f(m(z)) - z = 0$ .)

DEFINITION 5.4 (STABILITY OF (2.11) ON  $\mathbf{S}$ ). *Let  $\mathbf{S} \subset \mathbf{D}$  be a spectral domain (see Definition 3.20). We say that (2.11) is stable on  $\mathbf{S}$  if the following holds for some large enough constant  $C > 0$ . Suppose that  $\delta : \mathbf{S} \rightarrow (0, \infty)$  satisfies  $N^{-2} \leq \delta(z) \leq (\log N)^{-1}$  for  $z \in \mathbf{S}$  and that  $\delta$  is Lipschitz continuous with Lipschitz constant  $N^2$ . Suppose moreover that for each fixed  $E$ , the function  $\eta \mapsto \delta(E + i\eta)$  is nonincreasing for  $\eta > 0$ . Suppose that  $u : \mathbf{S} \rightarrow \mathbb{C}$  is the Stieltjes transform of a probability measure supported in  $[0, C]$ . Let  $z \in \mathbf{S}$  and suppose that*

$$|f(u(z)) - z| \leq \delta(z).$$

*If  $\text{Im } z < 1$  suppose also that*

$$|u - m| \leq \frac{C\delta}{\text{Im } m + \sqrt{\delta}} \quad (5.20)$$

*holds at  $z + iN^{-5}$ . Then (5.20) holds at  $z$ .*

This condition has previously appeared, in somewhat different guises, in the works [5, 7, 32], where it was established under various assumptions on  $\pi$ . For instance, in [5], it was established for  $\mathbf{S} = \mathbf{D}_1^e$  under the assumption that the edge  $k = 1$  is regular (see Definition 2.7 (i)). In Appendix A, we establish it for each of the subdomains  $\mathbf{D}_k^e$ ,  $\mathbf{D}_k^b$ , and  $\mathbf{D}^o$  separately, under the regularity assumptions from Definition 2.7.

In accordance with the assumptions of Theorem 3.22, we suppose throughout this section that (2.11) is stable on  $\mathbf{S}$ . Using (5.10) and (4.16), it is easy to obtain the following result, which is analogous to [7, Lemma 4.6].

LEMMA 5.5. *Suppose that the assumptions of Theorem 3.22 hold. Then we have  $\Lambda \prec N^{-1/4}$  uniformly in  $z \in \mathbf{S}$  satisfying  $\eta \geq 1$ .*

Exactly as in [7, Section 4], we use a stochastic continuity argument to estimate  $\Lambda$ , using Lemmas 5.3 and 5.5. The major input is the stability of (2.11) on  $\mathbf{S}$  in the sense of Definition 5.4, which is analogous to [7, Lemma 4.5]. Proposition 5.1 now follows by estimating the right-hand side of (5.19) by  $O_{\prec}(\Psi_{\Theta})$ , as follows from Lemma 5.2. In its proof, the error  $\mathbf{1}(\Xi)(G_{\mu\mu} - m)$  is controlled using (5.17) by  $\Theta + \Psi_{\Theta}$ ; similarly, the error  $\mathbf{1}(\Xi)(G_{ii} - m_i)$  is controlled using (5.15) by  $\sigma_i^2 \Psi_{\Theta}$ . We omit further details. This concludes the proof of Proposition 5.1.

**5.2. Fluctuation averaging and proof of Theorem 3.22.** The weak law, Proposition 5.1, may be upgraded to the strong law, Theorem 3.22, using improved estimates for the averaged quantities  $[Z]_M$  and  $[Z]_N$ . We follow the arguments of [7, Section 4.2] to the letter. The key input is the following result, which is very similar to [7, Lemma 4.9], combined with the observation that  $Z_s = (1 - \mathbb{E}_s) \frac{1}{G_{ss}}$  for  $s \in \mathcal{I}$ , as follows from (5.7) and (5.8).

LEMMA 5.6 (FLUCTUATION AVERAGING). *Suppose that the assumptions of Theorem 3.22 hold. Suppose that  $\Upsilon$  is a positive,  $N$ -dependent, deterministic function on  $\mathbf{S}$  satisfying  $N^{-1/2} \leq \Upsilon \leq N^{-c}$  for some constant  $c > 0$ . Suppose moreover that  $\Lambda \prec N^{-c}$  and  $\Lambda_o \prec \Upsilon$  on  $\mathbf{S}$ . Then on  $\mathbf{S}$  we have*

$$\frac{1}{N} \sum_{\mu \in \mathcal{I}_N} (1 - \mathbb{E}_{\mu}) \frac{1}{G_{\mu\mu}} = O_{\prec}(\Upsilon^2) \quad (5.21)$$

and

$$\frac{1}{M} \sum_{i \in \mathcal{I}_M} \frac{\sigma_i^2}{(1 + m_N \sigma_i)^2} (1 - \mathbb{E}_i) \frac{1}{G_{ii}} = O_{\prec}(\Upsilon^2). \quad (5.22)$$

PROOF. The estimate (5.21) is a trivial extension of [7, Lemma 4.9] and [14, Theorem 4.7]. The estimate (5.22) may be proved using exactly the same method, explained in [14, Appendix B]. The only complication in the proof is that the coefficients  $\frac{\sigma_i^2}{(1 + m_N \sigma_i)^2}$  are random and depend on  $i$ . Using (4.11), this is dealt with in the proof of [14, Appendix B] by writing, for any  $j \in \mathcal{I}_N$ ,

$$m_N = \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} G_{\mu\mu}^{(j)} + \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} \frac{G_{\mu j} G_{j\mu}}{G_{jj}},$$

and continuing in this manner with any further indices  $k, l, \dots \in \mathcal{I}_N$  that we wish to include as superscripts of  $G_{\mu\mu}$ . We refer to [14, Appendix B] for the full details.  $\square$

Using Lemmas 5.6 and (5.2) combined with (5.19) we get  $\mathbf{1}(\Xi)|f(m_N) - z| \prec \Psi_{\Theta}^2$ . Then we may follow the argument [7, Section 4.2] verbatim to get  $\Theta \prec (N\eta)^{-1}$  on  $\mathbf{S}$ . This concludes the proof of the averaged local law in Theorem 3.22. Moreover, the entrywise local law follows immediately from Proposition 5.1, Lemma 5.2, (5.15), and (5.17). This concludes the proof of Theorem 3.22.

## 6. Anisotropic local law for Gaussian $X$

We now begin the proof of Theorem 3.21, which consists of Sections 6–9. In this section we perform the first step of the proof, by establishing Theorem 3.21 for the special case that  $X = X^{\text{Gauss}}$  is Gaussian. This corresponds to Step (B) of the proof mentioned in the introduction.

PROPOSITION 6.1. *Theorem 3.21 holds if  $X = X^{\text{Gauss}}$  is Gaussian.*



The rest of this section is devoted to the proof of Proposition 6.1. We shall in fact prove the following result.

LEMMA 6.2. *Suppose that the assumptions of Theorem 3.21 hold and that  $X = X^{\text{Gauss}}$  is Gaussian. If the entrywise local law holds with parameters  $(X, D, \mathbf{S})$ , then the entrywise local law holds with parameters  $(X, \Sigma, \mathbf{S})$ . (Recall the definition of  $D \equiv D(\Sigma)$  from (3.21).)*

Before proving Lemma 6.2, we show how it implies Proposition 6.1. In order to prove the anisotropic local law, we have to estimate the left-hand side of (3.19). We split  $\mathbf{v} = \mathbf{v}_M + \mathbf{v}_N$  and  $\mathbf{w} = \mathbf{w}_M + \mathbf{w}_N$ , where  $\mathbf{v}_M, \mathbf{w}_M \in \mathbb{R}^{\mathcal{I}_M}$  and  $\mathbf{v}_N, \mathbf{w}_N \in \mathbb{R}^{\mathcal{I}_N}$ . Plugging this into the left-hand side of (3.19), we find that it suffices to control, for arbitrary deterministic orthogonal matrices  $O_M \in O(M)$  and  $O_N \in O(N)$ , the entries of the matrix

$$\begin{pmatrix} O_M & 0 \\ 0 & O_N \end{pmatrix} G^\Sigma \begin{pmatrix} O_M^* & 0 \\ 0 & O_N^* \end{pmatrix} \stackrel{\text{d}}{=} G^{O_M \Sigma O_M^*}, \quad (6.1)$$

where we used that  $X \stackrel{\text{d}}{=} O_M X O_M^*$  since  $X$  is Gaussian. Applying Lemma 6.2 to the matrices  $\tilde{\Sigma} = O_M \Sigma O_M^*$  and  $D = D(\Sigma) = D(\tilde{\Sigma})$ , we obtain the anisotropic local law with parameters  $(X, \Sigma, \mathbf{S})$ . Moreover, the averaged local law follows by writing  $\Sigma = U D U^*$  and setting  $O_M = U^*$  and  $O_N = 1$  in (6.1). This concludes the proof of Proposition 6.1.

PROOF OF LEMMA 6.2. In this proof we abbreviate  $G^D \equiv G$ . Using (6.1) with  $O_M = U^*$  and  $O_N = 1$ , we find that it suffices to prove

$$\left[ \begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} D^{-1} & 0 \\ 0 & 1 \end{pmatrix} \left[ G - \begin{pmatrix} -D(1+mD)^{-1} & 0 \\ 0 & m \end{pmatrix} \right] \begin{pmatrix} D^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U^* & 0 \\ 0 & 1 \end{pmatrix} \right]_{st} = O_{\prec}(\Psi)$$

for all  $s, t \in \mathcal{I}$ . In components, this reads

$$\left| \sum_{k,l \in \mathcal{I}_M} U_{ik} \frac{G_{kl} - \delta_{kl} m_k}{\sigma_k \sigma_l} U_{lj}^* \right| \prec \Psi, \quad (6.2)$$

$$\left| \sum_{k \in \mathcal{I}_M} U_{ik} \frac{G_{k\mu}}{\sigma_k} \right| + \left| \sum_{k \in \mathcal{I}_M} \frac{G_{\mu k}}{\sigma_k} U_{ki} \right| \prec \Psi, \quad (6.3)$$

$$|G_{\mu\nu} - \delta_{\mu\nu} m_\mu| \prec \Psi, \quad (6.4)$$

for  $i, j \in \mathcal{I}_M$  and  $\mu, \nu \in \mathcal{I}_N$ .

The estimate (6.4) is trivial by assumption. What remains is the proof of (6.2) and (6.3). It is based on the polynomialization method developed in [7, Section 5]. The argument is very similar to that of [7], and we only outline the differences.

Let us begin with (6.2). By the assumption  $|G_{kk} - m_k| \prec \Psi \sigma_k^2$  and orthogonality of  $U$ , we have

$$\sum_{k,l} U_{ik} \frac{G_{kl} - \delta_{kl} m_k}{\sigma_k \sigma_l} U_{lj}^* = \sum_k U_{ik} \frac{G_{kk} - m_k}{\sigma_k^2} U_{kj}^* + \sum_{k \neq l} U_{ik} \frac{G_{kl}}{\sigma_k \sigma_l} U_{lj} = O_{\prec}(\Psi) + \mathcal{Z},$$

where we defined  $\mathcal{Z} := \sum_{k \neq l} (\sigma_k \sigma_l)^{-1} U_{ik} G_{kl} U_{lj}$ . We need to prove that  $|\mathcal{Z}| \prec \Psi$ , which, following [7, Section 5], we do by estimating the moment  $\mathbb{E}|\mathcal{Z}|^p$  for fixed  $p \in 2\mathbb{N}$ . The argument from [7, Section

5] may be taken over with minor changes. We use the identities (4.11),

$$\sum_{k \neq l \in \mathcal{I}_M \setminus T} (\sigma_k \sigma_l)^{-1} U_{ik} G_{kl}^{(T)} U_{lj} = \sum_{k \neq l \in \mathcal{I}_M \setminus T} (\sigma_k \sigma_l)^{-1} U_{ik} U_{lj} G_{kk}^{(T)} G_{ll}^{(kT)} (X G^{(klT)} X^*)_{kl} \quad (6.5)$$

for  $T \subset \mathcal{I}_M$  (which follows from (4.7)), (4.6), and

$$G_{ii}^{(T)} = \frac{-\sigma_i}{1 + m\sigma_i + \sigma_i((X G^{(iT)} X^*)_{ii} - m)} = \sum_{\ell=0}^{L-1} m_i^{\ell+1} ((X G^{(iT)} X^*)_{ii} - m)^\ell + O_{\prec}(\sigma_i^{3L+1} \Psi^L),$$

as follows from (4.6), (5.4), and  $(X G^{(iT)} X^*)_{ii} - m = O_{\prec}(\sigma_i^2 \Psi)$  (which may itself be deduced from (4.6)). We omit further details.

Finally, the proof of (6.3) is similar to that of (6.2). Writing  $\mathcal{Z}' := \sum_{k \in \mathcal{I}_M} \sigma_k^{-1} U_{ik} G_{k\mu}$ , we estimate  $\mathbb{E}|\mathcal{Z}'|^p$  for  $p \in 2\mathbb{N}$  using the method of [7, Section 5]. Instead of (6.5) we use

$$\sum_{k \in \mathcal{I}_M \setminus T} \sigma_k^{-1} U_{ik} G_{k\mu}^{(T)} = - \sum_{k \in \mathcal{I}_M \setminus T} \sigma_k^{-1} U_{ik} G_{kk}^{(T)} (X G^{(kT)})_{k\mu}.$$

The rest of the argument is the same as before.  $\square$

## 7. Self-consistent comparison I: the main argument

In this section we establish Theorem 3.21 (i) under the additional assumption that the third moment of all entries of  $X$  is zero.

**PROPOSITION 7.1.** *Suppose that the assumptions of Theorem 3.21 hold. Suppose moreover that  $X$  satisfies the additional condition*

$$\mathbb{E} X_{i\mu}^3 = 0. \quad (7.1)$$

*If the anisotropic local law holds with parameters  $(X^{\text{Gauss}}, \Sigma, \mathbf{S})$ , then the anisotropic local law holds with parameters  $(X, \Sigma, \mathbf{S})$ .*

The rest of this section is devoted to the proof of Proposition 7.1. This is the heart of the proof of the anisotropic local law.

**7.1. Sketch of the proof.** Before giving the full proof of Proposition 7.1, we outline the key ideas of the self-consistent comparison argument on which it relies. For simplicity, we ignore the factors  $\underline{\Sigma}^{-1}$  on the left-hand side of (3.19) (hence obtaining a weaker bound; see (2.7)), so that by polarization it suffices to estimate  $G_{\mathbf{v}\mathbf{v}} - \Pi_{\mathbf{v}\mathbf{v}}$  for a deterministic unit vector  $\mathbf{v} \in \mathbb{R}^{\mathcal{I}}$ .

We introduce a family of interpolating matrices  $(X^\theta)_{\theta \in [0,1]}$  satisfying  $X^0 = G^{\text{Gauss}}$  and  $X^1 = X$ . In order to prove  $|G_{\mathbf{v}\mathbf{v}}(X^1) - \Pi_{\mathbf{v}\mathbf{v}}| \prec \Psi$ , it suffices to prove a high moment bound  $\mathbb{E}|G_{\mathbf{v}\mathbf{v}}(X^1) - \Pi_{\mathbf{v}\mathbf{v}}|^p \leq (N^{C\delta} \Psi)^p$  for any fixed  $p \in \mathbb{N}$  and  $\delta > 0$ , and large enough  $N$ . Since, by assumption, this moment bound holds for  $X^1$  replaced by  $X^0$ , using Grönwall's inequality it suffices to prove that

$$\frac{d}{d\theta} \mathbb{E}|G_{\mathbf{v}\mathbf{v}}(X^\theta) - \Pi_{\mathbf{v}\mathbf{v}}|^p \leq (N^{C\delta} \Psi)^p + C \mathbb{E}|G_{\mathbf{v}\mathbf{v}}(X^\theta) - \Pi_{\mathbf{v}\mathbf{v}}|^p. \quad (7.2)$$

Note that we estimate the derivative of the quantity  $\mathbb{E}|G_{\mathbf{v}\mathbf{v}}(X^\theta) - \Pi_{\mathbf{v}\mathbf{v}}|^p$  in terms of itself. It is therefore important that the arguments  $X^\theta$  are the same on both sides. In particular, a Lindeberg-type replacement of the matrix entries one by one would not work, and a continuous interpolation is necessary. A common choice of continuous interpolation of random matrices, arising for example from Dyson Brownian motion, is  $X^\theta = \sqrt{\theta}X^1 + \sqrt{1-\theta}X^0$ , where  $X^0$  and  $X^1$  are defined on a common probability space and are independent. With this choice of interpolation, however, the differentiation on the left-hand side of (7.2) leads to very complicated expressions that are hard to control. Instead, we interpolate by setting  $X_{i\mu}^\theta := \chi_{i\mu}^\theta X_{i\mu}^1 + (1 - \chi_{i\mu}^\theta)X_{i\mu}^0$ , where  $(\chi_{i\mu}^\theta)$  is a family of i.i.d. Bernoulli random variables, independent of  $X^0$  and  $X^1$ , satisfying  $\mathbb{P}(\chi_{i\mu}^\theta = 1) = \theta$  and  $\mathbb{P}(\chi_{i\mu}^\theta = 0) = 1 - \theta$ . This may also be interpreted as a linear interpolation between the laws of  $X_{i\mu}^0$  and  $X_{i\mu}^1$ . It gives rise to formulas that are simple enough for our comparison argument to work. In fact, after some calculations we find that it suffices to prove

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \left( \frac{\partial}{\partial X_{i\mu}^\theta} \right)^n |G_{\mathbf{v}\mathbf{v}}(X^\theta) - \Pi_{\mathbf{v}\mathbf{v}}|^p \right| \leq (N^{C\delta}\Psi)^p + C \mathbb{E}|G_{\mathbf{v}\mathbf{v}}(X^\theta) - \Pi_{\mathbf{v}\mathbf{v}}|^p \quad (7.3)$$

for all  $n = 4, \dots, 4p$ .

Computing the derivatives on the left-hand side of (7.3) leads to a product of terms of the form

$$(i) \quad |G_{\mathbf{v}\mathbf{v}} - \Pi_{\mathbf{v}\mathbf{v}}|, \quad (ii) \quad G_{\mathbf{v}i}, G_{\mathbf{v}\mu}, G_{i\mathbf{v}}, G_{\mu\mathbf{v}}, \quad (iii) \quad G_{i\mu}, G_{\mu i}, \quad (7.4)$$

and their complex conjugates (here we omit the argument  $X^\theta$ ). Terms of type (i) are simply kept as they are; they will be put into the second term on the right-hand side of (7.3). Terms of type (ii) are the key to the gain that allows us to compensate losses from several other terms, such as the terms of type (iii) (see below). The gain is obtained in combination with the summation over  $i$  and  $\mu$  on the left-hand side of (7.3), according to estimates of the form

$$\frac{1}{N} \sum_{i \in \mathcal{I}_M} |G_{\mathbf{v}i}|^2 \prec \frac{\text{Im } G_{\mathbf{v}\mathbf{v}} + \eta}{N\eta}. \quad (7.5)$$

Note that the right-hand side of (7.5) cannot be estimated by either  $N^{C\delta}\Psi$  or  $|G_{\mathbf{v}\mathbf{v}} - \Pi_{\mathbf{v}\mathbf{v}}|$ , so that a further ingredient is required to obtain (7.3). This ingredient takes the form of an a priori bound

$$\text{Im } G_{\mathbf{v}\mathbf{v}} \prec \text{Im } m + N^{C\delta}\Psi, \quad (7.6)$$

which allows us to estimate the right-hand side of (7.5) by  $N^{C\delta}\Psi^2$ . The a priori bound (7.6) will be obtained from a bootstrapping, as explained below.

An estimate similar to (7.5) may be obtained for  $\frac{1}{N} \sum_{i \in \mathcal{I}_M} |G_{\mathbf{v}i}|$  by a simple application of Cauchy-Schwarz to (7.5). However, for  $\frac{1}{N} \sum_{i \in \mathcal{I}_M} |G_{\mathbf{v}i}|^d$  with  $d \geq 3$ , we have to estimate  $d - 2$  factors  $|G_{\mathbf{v}i}|$  pointwise to recover (7.5). The trivial bound  $|G_{\mathbf{v}i}| \leq \eta^{-1}$  is far too rough for small  $\eta$ . A more reasonable attempt is to write  $|G_{\mathbf{v}i}| \leq |G_{\mathbf{v}i} - \Pi_{\mathbf{v}i}| + |\Pi_{\mathbf{v}i}| \prec \max_{|\mathbf{w}|=1} |G_{\mathbf{w}\mathbf{w}} - \Pi_{\mathbf{w}\mathbf{w}}| + 1$ , by polarization and  $\|\Pi\| \leq C$ . This, however, leads to bounds of the form  $\mathbb{E} \max_{|\mathbf{w}|=1} |G_{\mathbf{w}\mathbf{w}} - \Pi_{\mathbf{w}\mathbf{w}}|^p$ , which are not affordable. Note that, as explained in (1.9), bounds of the form  $\max_{|\mathbf{w}|=1} \mathbb{E}|G_{\mathbf{w}\mathbf{w}} - \Pi_{\mathbf{w}\mathbf{w}}|^p$  would be affordable, and in fact we shall need such more general bounds in Section 8.

The correct way to deal with the pointwise estimate of  $|G_{\mathbf{v}i}|$  is a second type of a priori bound,

$$|G_{\mathbf{x}\mathbf{y}}| \prec N^{2\delta}, \quad (7.7)$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are deterministic unit vectors. As (7.6), the estimate (7.7) will follow from the bootstrapping explained below. The need for an a priori bound of the form (7.7) on the entries of  $G$  is also apparent for terms of type (iii), which are also estimated using (7.7).

The preceding discussion was rather cavalier with the various constants  $C$  in the factors  $N^{C\delta}$ . In fact, some care has to be taken to ensure that the factors of  $N^{C\delta}$  arising from the use of the a priori bounds (7.6) and (7.7) are compensated by a sufficiently large number of factors of the form (7.5). Moreover, in order to be able to iterate this argument from large to small scales, as explained in the next paragraph, we need an explicit bound of the constant  $C$  in (7.3) in terms of the constant  $C$  in (7.6).

Finally, we outline the bootstrapping. The above argument can be carried out for  $\eta = 1$  using the trivial a priori bounds where the right-hand sides of (7.6) and (7.7) are replaced by 1. The idea of the bootstrapping is to fix a small exponent  $\delta > 0$  and to proceed from larger scales  $\eta$  to smaller scales in multiplicative increments of  $N^{-\delta}$ , i.e.  $\eta = 1, N^{-\delta}, N^{-2\delta}, \dots, N^{-1}$ . Suppose that we have proved the anisotropic local law at the scale  $\eta = N^{-\delta l}$ . In particular, since  $|\Pi_{\mathbf{x}\mathbf{y}}| \leq C$ , we have proved for  $\eta = N^{-\delta l}$  that  $|G_{\mathbf{x}\mathbf{y}}(X, E + i\eta)| \prec 1$  for any  $X$  satisfying the assumptions (2.4) and (2.5). Starting from this bound, it is not hard to derive the a priori bound (7.7) at  $z = E + iN^{-\delta}\eta$ . Similarly, by using simple monotonicity properties of the Stieltjes transform, we obtain the other a priori bound (7.6) at  $z = E + iN^{-\delta}\eta$ . These estimates hold for  $X = X^\theta$  for all  $\theta \in [0, 1]$ , since  $X^\theta$  satisfies the assumptions (2.4) and (2.5) uniformly in  $\theta$ . Note that this bootstrapping on  $\eta$  is different from the stochastic continuity argument commonly used in establishing local laws (see e.g. [7, Section 4.1]), since the multiplicative steps of size  $N^{-\delta}$  are much too large for a continuity argument to work; all that they provide are the rather crude a priori bounds, (7.6) and (7.7).

Since  $\delta > 0$  is fixed, the bootstrapping consists of  $O(\delta^{-1})$  steps. The resulting estimates contain an extra factor  $N^{C\delta}$ . Since  $\delta > 0$  can be made arbitrarily small, the claim will follow on all scales.

**7.2. Bootstrapping on the spectral scale.** We now move on to the proof of Proposition 7.1. By polarization, it suffices to prove that for any deterministic unit vector  $\mathbf{v} \in \mathbb{R}^Z$  we have

$$\langle \mathbf{v}, \underline{\Sigma}^{-1}(G(z) - \Pi(z))\underline{\Sigma}^{-1}\mathbf{v} \rangle = O_{\prec}(\Psi(z)) \quad (7.8)$$

for  $z \in \mathbf{S}$ . In fact, we shall prove (7.8) for all  $z \in \widehat{\mathbf{S}}$  in some discrete subset  $\widehat{\mathbf{S}} \subset \mathbf{S}$ . We take  $\widehat{\mathbf{S}}$  to be an  $N^{-10}$ -net in  $\mathbf{S}$  (i.e. for every  $z \in \mathbf{S}$  there is a  $w \in \widehat{\mathbf{S}}$  such that  $|z - w| \leq N^{-10}$ ) satisfying  $|\widehat{\mathbf{S}}| \leq N^{20}$ . The function  $z \mapsto \Sigma^{-1}(G(z) - \Pi(z))\underline{\Sigma}^{-1}$  is Lipschitz continuous (with respect to the operator norm) in  $\mathbf{S}$  with Lipschitz constant  $(C + \|X^*X\|)N^2$ , as follows from Lemma 4.9 and the bound (3.20). Using Lemma 4.8 it is therefore not hard to see that (7.8) for all  $z \in \mathbf{S}$  follows provided we can prove (7.8) for all  $z \in \widehat{\mathbf{S}}$ .

The proof consists of a bootstrap argument from larger scales to smaller scales in multiplicative increments of  $N^{-\delta}$ . Here

$$\delta \in \left(0, \frac{\tau}{2C_0}\right), \quad (7.9)$$

is fixed, where  $C_0 > 0$  is a universal constant that will be chosen large enough in the proof ( $C_0 = 25$  will work). For any  $\eta \geq N^{-1}$  we define  $\eta_0 \leq \eta_1 \leq \dots \leq \eta_L$ , where

$$L \equiv L(\eta) := \max\{l \in \mathbb{N} : \eta N^{\delta(l-1)} < 1\}, \quad (7.10)$$

through

$$\eta_l := \eta N^{\delta l} \quad (l = 0, \dots, L-1), \quad \eta_L := 1. \quad (7.11)$$

Note that  $L \leq \delta^{-1} + 1$ .

We shall always work with a net  $\widehat{\mathbf{S}}$  satisfying the following condition.

DEFINITION 7.2. *Let  $\widehat{\mathbf{S}}$  be an  $N^{-10}$ -net of  $\mathbf{S}$  satisfying  $|\widehat{\mathbf{S}}| \leq N^{20}$  and the condition*

$$E + i\eta \in \widehat{\mathbf{S}} \quad \implies \quad E + i\eta_l \in \widehat{\mathbf{S}} \quad \text{for } l = 1, \dots, L(\eta).$$

The bootstrapping is formulated in terms of two scale-dependent properties,  $(\mathbf{A}_m)$  and  $(\mathbf{C}_m)$ , formulated on the subsets

$$\widehat{\mathbf{S}}_m := \{z \in \widehat{\mathbf{S}} : \text{Im } z \geq N^{-\delta m}\}.$$

We denote by  $\mathbb{S} := \{\mathbf{v} \in \mathbb{R}^{\mathcal{I}} : |\mathbf{v}| = 1\}$  the unit sphere of  $\mathbb{R}^{\mathcal{I}}$ , and abbreviate

$$\bar{\mathbf{v}} := \underline{\Sigma}^{-1} \mathbf{v} \tag{7.12}$$

for  $\mathbf{v} \in \mathbb{S}$ .

$(\mathbf{A}_m)$  For all  $z \in \widehat{\mathbf{S}}_m$ , all deterministic  $\mathbf{v} \in \mathbb{S}$ , and all  $X$  satisfying (2.4) and (2.5), we have

$$\text{Im } G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z) \prec \text{Im } m(z) + N^{C_0\delta} \Psi(z). \tag{7.13}$$

$(\mathbf{C}_m)$  For all  $z \in \widehat{\mathbf{S}}_m$ , all deterministic  $\mathbf{v} \in \mathbb{S}$ , and all  $X$  satisfying (2.4) and (2.5), we have

$$|G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z) - \Pi_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z)| \prec N^{C_0\delta} \Psi(z). \tag{7.14}$$

The bootstrapping is started by the following result.

LEMMA 7.3. *Property  $(\mathbf{A}_0)$  holds.*

PROOF. Using the definition of  $\Pi$ , the assumption (3.20), and Lemmas 4.9 and 4.8, we find for  $z \in \widehat{\mathbf{S}}_0$

$$\text{Im } G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z) \leq \text{Im } \Pi_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z) + \|\underline{\Sigma}^{-1}(G(z) - \Pi(z))\underline{\Sigma}^{-1}\| \prec 1.$$

The claim now follows from Lemma 4.10.  $\square$

The key step is the following result.

LEMMA 7.4. *Under the assumptions of Proposition 7.1, for any  $1 \leq m \leq \delta^{-1}$ , property  $(\mathbf{A}_{m-1})$  implies property  $(\mathbf{C}_m)$ .*

Using the definition of  $\Pi$  and the assumption (3.20), it is easy to see that property  $(\mathbf{C}_m)$  implies property  $(\mathbf{A}_m)$ . We therefore conclude from Lemmas 7.3 and 7.4 that (7.14) holds for all  $z \in \widehat{\mathbf{S}}$ . Since  $\delta$  can be chosen arbitrarily small under the condition (7.9), we conclude that (7.8) follows for all  $z \in \widehat{\mathbf{S}}$ , and the proof of Proposition 7.1 is complete.

What remains is the proof of Lemma 7.4. We shall estimate the random variable

$$F_{\mathbf{v}}(X, z) := |G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z) - \Pi_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z)|, \tag{7.15}$$

where  $\mathbf{v} \in \mathbb{S}$  is deterministic. By Markov's inequality and Definition 3.4, we conclude that (7.14) follows provided we can prove the following result.

LEMMA 7.5. *Fix  $p \in 2\mathbb{N}$  and  $m \leq \delta^{-1}$ . Suppose that (7.1) and  $(\mathbf{A}_{m-1})$  hold. Then we have*

$$\mathbb{E} F_{\mathbf{v}}^p(X, z) \leq (N^{C_0\delta} \Psi(z))^p$$

for all  $z \in \widehat{\mathbf{S}}_m$  and all deterministic  $\mathbf{v} \in \mathbb{S}$ .

**7.3. Interpolation.** We use the interpolation outlined in Section 7.1.

DEFINITION 7.6 (INTERPOLATING MATRICES). Introduce the notations  $X^0 := G^{\text{Gauss}}$  and  $X^1 := X$ . For  $u \in \{0, 1\}$ ,  $i \in \mathcal{I}_M$ , and  $\mu \in \mathcal{I}_N$ , denote by  $\rho_{i\mu}^u$  the law of  $X_{i\mu}^u$ . For  $\theta \in [0, 1]$  we define the law

$$\rho_{i\mu}^\theta := \theta \rho_{i\mu}^1 + (1 - \theta) \rho_{i\mu}^0.$$

Let  $(X^0, X^\theta, X^1)$  be a triple of independent  $\mathcal{I}_M \times \mathcal{I}_N$  random matrices, where for  $u \in \{0, \theta, 1\}$  the matrix  $X^u = (X_{i\mu}^u)$  has law

$$\prod_{i \in \mathcal{I}_M} \prod_{\mu \in \mathcal{I}_N} \rho_{i\mu}^u(\mathrm{d}X_{i\mu}^u).$$

For  $i \in \mathcal{I}_M$ ,  $\mu \in \mathcal{I}_N$ , and  $\lambda \in \mathbb{R}$  we define the matrix

$$(X_{(i\mu)}^{\theta, \lambda})_{j\nu} := \begin{cases} \lambda & \text{if } (j, \nu) = (i, \mu) \\ X_{j\nu}^\theta & \text{if } (j, \nu) \neq (i, \mu). \end{cases}$$

We also introduce the matrices

$$G^\theta(z) := G^\Sigma(X^\theta, z), \quad G_{(i\mu)}^{\theta, \lambda}(z) := G^\Sigma(X_{(i\mu)}^{\theta, \lambda}, z),$$

(recall the notation (3.2)).

Throughout the following we shall need to deduce bounds of the form  $\mathbb{E}\xi \prec \Gamma$  from  $\xi \prec \Gamma$ . This is the content of the following lemma, whose proof is a simple application of Cauchy-Schwarz.

LEMMA 7.7. Let  $\Gamma$  be deterministic and satisfy  $\Gamma \geq N^{-C}$  for some constant  $C > 0$ . Let  $\xi$  be a random variable satisfying  $\xi \prec \Gamma$  and  $\mathbb{E}\xi^2 \leq N^C$ . Then  $\mathbb{E}\xi \prec \Gamma$ .

In the following applications of Lemma 7.7, we shall often not mention the assumption  $\mathbb{E}\xi^2 \leq N^C$ ; it may be always easily verified using rough bounds from Lemma 4.9.

We shall prove Lemma 7.5 by interpolation between the ensembles  $X^0$  and  $X^1$ . The bound for the Gaussian case  $X^0$  is given by the following result.

LEMMA 7.8. Lemma 7.5 holds if  $X$  is replaced with  $X^0$ .

PROOF. By assumption of Proposition 7.1, we have  $F_{\mathbf{V}}^p(X^0, z) \prec \Psi^p$ . In order to conclude the proof using Lemma 7.7, we need a rough bound of the form  $\mathbb{E}(F_{\mathbf{V}}^p(X^0, z))^2 \leq N^{C_p}$  for some constant  $C_p$  depending on  $p$ . This easily follows from the first identity in (4.3) combined with the bound

$$\|\Sigma^{-1}(-\Sigma + \Sigma(1 + m\Sigma)^{-1})\Sigma^{-1}\| \leq C.$$

We omit further details. □

The basic interpolation formula is given by the following lemma, which follows from the fundamental theorem of calculus.

LEMMA 7.9. For  $F : \mathbb{R}^{\mathcal{I}_M \times \mathcal{I}_N} \rightarrow \mathbb{C}$  we have

$$\mathbb{E}F(X^1) - \mathbb{E}F(X^0) = \int_0^1 \mathrm{d}\theta \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left[ \mathbb{E}F(X_{(i\mu)}^{\theta, X_{i\mu}^1}) - \mathbb{E}F(X_{(i\mu)}^{\theta, X_{i\mu}^0}) \right] \quad (7.16)$$

provided all the expectations exist.

We shall apply Lemma 7.9 with  $F(X) = F_{\mathbf{v}}^p(X, z)$ , where  $F_{\mathbf{v}}(X, z)$  was defined in (7.15). The main work is to derive the following self-consistent estimate for the right-hand side of (7.16). We emphasize that our estimates are always uniform in quantities, such as  $\mathbf{v}$ ,  $\Sigma$ ,  $\theta$ , and  $z$ , that are not explicitly fixed.

LEMMA 7.10. *Fix  $p \in 2\mathbb{N}$  and  $m \leq \delta^{-1}$ . Suppose that (7.1) and  $(\mathbf{A}_{m-1})$  hold. Then we have*

$$\sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left[ \mathbb{E} F_{\mathbf{v}}^p \left( X_{(i\mu)}^{\theta, X_{i\mu}^1}, z \right) - \mathbb{E} F_{\mathbf{v}}^p \left( X_{(i\mu)}^{\theta, X_{i\mu}^0}, z \right) \right] = O \left( (N^{C_0 \delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p (X^\theta, z) \right) \quad (7.17)$$

for all  $\theta \in [0, 1]$ , all  $z \in \widehat{\mathbf{S}}_m$ , and all deterministic  $\mathbf{v} \in \mathbb{S}$ .

Combining Lemmas 7.8, 7.9, and 7.10 with a Grönwall argument, we conclude the proof of Lemma 7.5, and hence of Proposition 7.1. Note that for this Grönwall argument to work, it is essential that the error term  $\sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p (X^\theta, z)$  on the right-hand side of (7.17) be multiplied by a factor that is bounded (as is implied by the notation  $O(\cdot)$ ). Even a factor  $\log N$  multiplying  $\sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p (X^\theta, z)$  would render (7.17) useless.

We remark that, in the proof Proposition 7.1 developed in the rest of this section, we in fact obtain a stronger version of (7.17) where the quantity  $\sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p (X^\theta, z)$  on the right-hand side of (7.17) is replaced with  $\mathbb{E} F_{\mathbf{v}}^p (X^\theta, z)$ . However, we keep the more general form (7.17) because it will be needed for the proof of Proposition 8.1 in Section 8.

In order to prove Lemma 7.10, we compare the ensembles  $X_{(i\mu)}^{\theta, X_{i\mu}^0}$  and  $X_{(i\mu)}^{\theta, X_{i\mu}^1}$  via  $X_{(i\mu)}^{\theta, 0}$ . Clearly, it suffices to prove the following result.

LEMMA 7.11. *Fix  $p \in 2\mathbb{N}$  and  $m \leq \delta^{-1}$ . Suppose that (7.1) and  $(\mathbf{A}_{m-1})$  hold. Then there exists some function  $A(\cdot, z)$  such that for  $u \in \{0, 1\}$  we have*

$$\sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left[ \mathbb{E} F_{\mathbf{v}}^p \left( X_{(i\mu)}^{\theta, X_{i\mu}^u}, z \right) - \mathbb{E} A \left( X_{(i\mu)}^{\theta, 0}, z \right) \right] = O \left( (N^{C_0 \delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p (X^\theta, z) \right) \quad (7.18)$$

for all  $\theta \in [0, 1]$ , all  $z \in \widehat{\mathbf{S}}_m$ , and all deterministic  $\mathbf{v} \in \mathbb{S}$ .

In the remainder of this section, we prove Lemma 7.11 for  $u = 1$ . In order to make use of the assumption  $(\mathbf{A}_{m-1})$ , which involves spectral parameters in  $\widehat{\mathbf{S}}_{m-1}$ , to obtain estimates for spectral parameters in  $\widehat{\mathbf{S}}_m$ , we use the following rough bound.

LEMMA 7.12. *For any  $z = E + i\eta \in \mathbf{S}$  and  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}}$  we have*

$$|G_{\mathbf{xy}}(z) - \Pi_{\mathbf{xy}}(z)| \prec N^{2\delta} \sum_{l=1}^{L(\eta)} \left( \operatorname{Im} G_{\mathbf{xx}}(E + i\eta_l) + \operatorname{Im} G_{\mathbf{yy}}(E + i\eta_l) \right) + |\underline{\Sigma} \mathbf{x}| |\underline{\Sigma} \mathbf{y}|,$$

where we recall the definitions of  $L(\eta)$  and  $\eta_l$  from (7.10) and (7.11).

PROOF. From (3.5) and (4.15) we get

$$G - \Pi = \sum_k \frac{\mathbf{u}_k \mathbf{u}_k^*}{\lambda_k - z} - \begin{pmatrix} m\Sigma^2(1 + m\Sigma)^{-1} & 0 \\ 0 & I_N \end{pmatrix}.$$



Using (3.20) we therefore get

$$|G_{\mathbf{xy}}(z) - \Pi_{\mathbf{xy}}(z)| \leq \sum_k \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|} + \sum_k \frac{\langle \mathbf{y}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|} + C|\underline{\Sigma}\mathbf{x}||\underline{\Sigma}\mathbf{y}|.$$

It suffices to estimate the first term. Setting  $\eta_{-1} := 0$  and  $\eta_{L+1} := \infty$ , we define the subsets of indices

$$U_l := \{k : \eta_{l-1} \leq |\lambda_k - E| < \eta_l\} \quad (l = 0, 1, \dots, L+1).$$

We split the summation

$$\sum_k \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|} = \sum_{l=0}^{L+1} \sum_{k \in U_l} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|}$$

and treat each  $l$  separately. For  $l = 1, \dots, L$  we find

$$\begin{aligned} \sum_{k \in U_l} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|} &\leq \sum_{k \in U_l} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2 \eta_l}{(\lambda_k - E)^2} \leq 2 \sum_{k \in U_l} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2 \eta_l}{(\lambda_k - E)^2 + \eta_{l-1}^2} \\ &\leq \frac{2\eta_l}{\eta_{l-1}} \operatorname{Im} G_{\mathbf{xx}}(E + i\eta_{l-1}) \leq 2N^\delta \operatorname{Im} G_{\mathbf{xx}}(E + i\eta_{l-1}), \end{aligned}$$

where the third step easily follows from (4.15). Using that the map  $y \mapsto y \operatorname{Im} G_{\mathbf{xx}}(E + iy)$  is nondecreasing, we find for  $l = 1, \dots, L$  that

$$\sum_{k \in U_l} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|} \leq 2N^{2\delta} \operatorname{Im} G_{\mathbf{xx}}(E + i\eta_{l \vee 1}),$$

as desired.

Next, we estimate

$$\sum_{k \in U_0} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|} \leq \sum_{k \in U_0} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2 2\eta}{(\lambda_k - E)^2 + \eta^2} = 2 \operatorname{Im} G_{\mathbf{xx}}(E + i\eta) \leq 2N^\delta \operatorname{Im} G_{\mathbf{xx}}(E + i\eta_1).$$

Finally, we estimate

$$\sum_{k \in U_{L+1}} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2}{|\lambda_k - z|} \leq 2 \sum_{k \in U_{L+1}} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2 |\lambda_k - E| \eta_L}{(\lambda_k - E)^2 + \eta_L^2} \prec \sum_{k \in U_{L+1}} \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle^2 \eta_L}{(\lambda_k - E)^2 + \eta_L^2} \leq \operatorname{Im} G_{\mathbf{xx}}(E + i\eta_L),$$

where in the second step we used that  $\lambda_k \prec 1$ , as follows from (2.7) and Lemma 4.8. This concludes the proof.  $\square$

LEMMA 7.13. *Suppose that  $(\mathbf{A}_{m-1})$  holds. Then we have*

$$\underline{\Sigma}^{-1}(G(z) - \Pi(z))\underline{\Sigma}^{-1} = O_{\prec}(N^{2\delta}) \quad (7.19)$$

for all  $z \in \widehat{\mathbf{S}}_m$ . Moreover, for any deterministic  $\mathbf{v} \in \mathbb{S}$  we have

$$\operatorname{Im} G_{\widehat{\mathbf{v}}\widehat{\mathbf{v}}}(z) \prec N^{2\delta}(\operatorname{Im} m(z) + N^{C_0\delta}\Psi(z)) \quad (7.20)$$

for all  $z \in \widehat{\mathbf{S}}_m$ .

PROOF. Let  $E + i\eta \in \widehat{\mathbf{S}}_m$ . Then we have  $E + i\eta_l \in \widehat{\mathbf{S}}_{m-1}$  for  $l = 1, \dots, L(\eta)$ . Therefore (7.13) yields, for any deterministic  $\mathbf{v} \in \mathbb{S}$ ,

$$\operatorname{Im} G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(E + i\eta_l) \prec |\underline{\Sigma}\bar{\mathbf{v}}|^2 + \operatorname{Im}\langle \bar{\mathbf{v}}, \Pi(E + i\eta_l)\bar{\mathbf{v}} \rangle \leq C|\underline{\Sigma}\bar{\mathbf{v}}|^2 = C,$$

where in the second step we used (3.20) and the definition of  $\Pi$  from (3.5). The estimate (7.19) now follows easily using Lemma 7.12.

In order to prove (7.20), we remark that if  $s(z)$  is the Stieltjes transform of any probability measure, the map  $\eta \mapsto \eta \operatorname{Im} s(E + i\eta)$  is nondecreasing and the map  $\eta \mapsto \eta^{-1} \operatorname{Im} s(E + i\eta)$  is nonincreasing. Abbreviating  $z = E + i\eta \in \widehat{\mathbf{S}}_m$  and  $z_1 = E + i\eta_1 \in \widehat{\mathbf{S}}_{m-1}$ , we therefore find

$$\operatorname{Im} G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z) \leq N^\delta \operatorname{Im} G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}(z_1) \prec N^\delta (\operatorname{Im} m(z_1) + N^{C_0\delta} \Psi(z_1)) \leq N^{2\delta} (\operatorname{Im} m(z) + N^{C_0\delta} \Psi(z)),$$

where in the second step we used the assumption  $(\mathbf{A}_{m-1})$ .  $\square$

**7.4. Expansion.** We now develop the main expansion which underlies the proof of Lemma 7.11. Throughout the rest of this section we suppose that  $(\mathbf{A}_{m-1})$  holds. The rest of the proof is performed at a single  $z \in \widehat{\mathbf{S}}_m$ , and from now on we therefore consistently omit the argument  $z$  from our notation.

Let  $i \in \mathcal{I}_M$  and  $\mu \in \mathcal{I}_N$ . Define the  $\mathcal{I} \times \mathcal{I}$  matrix  $\Delta_{(i\mu)}^\lambda$  through

$$(\Delta_{(i\mu)}^\lambda)_{st} := \lambda \delta_{is} \delta_{\mu t} + \lambda \delta_{it} \delta_{\mu s}.$$

Thus we get the resolvent expansion, for  $\lambda, \lambda' \in \mathbb{R}$  and any  $K \in \mathbb{N}$ ,

$$G_{(i\mu)}^{\theta, \lambda'} = G_{(i\mu)}^{\theta, \lambda} + \sum_{k=1}^K G_{(i\mu)}^{\theta, \lambda} (\Delta_{(i\mu)}^{\lambda-\lambda'} G_{(i\mu)}^{\theta, \lambda})^k + G_{(i\mu)}^{\theta, \lambda'} (\Delta_{(i\mu)}^{\lambda-\lambda'} G_{(i\mu)}^{\theta, \lambda})^{K+1}. \quad (7.21)$$

The following result provides a priori bounds for the entries of  $G_{(i\mu)}^{\theta, \lambda}$ .

LEMMA 7.14. *Suppose that  $y$  is a random variable satisfying  $|y| \prec N^{-1/2}$ . Then*

$$\underline{\Sigma}^{-1} (G_{(i\mu)}^{\theta, y} - \Pi) \underline{\Sigma}^{-1} = O_{\prec}(N^{2\delta}) \quad (7.22)$$

for all  $i \in \mathcal{I}_M$  and  $\mu \in \mathcal{I}_N$ .

PROOF. We use (7.21) with  $K := 10$ ,  $\lambda' := y$ , and  $\lambda := X_{i\mu}^\theta$ , so that  $G_{(i\mu)}^{\theta, \lambda} = G_{(i\mu)}^\theta$ . By (3.20), we have

$$\|\underline{\Sigma}^{-1} \Pi\| + \|\Pi \underline{\Sigma}^{-1}\| \leq C. \quad (7.23)$$

By assumption  $(\mathbf{A}_{m-1})$ , the conclusions of Lemma 7.13 hold for the matrix ensemble  $X^\theta$  (since it satisfies (2.4) and (2.5)). Hence, (7.22) holds for  $y = \lambda$ , and in particular  $\underline{\Sigma}^{-1} G_{(i\mu)}^{\theta, \lambda} = O_{\prec}(N^{2\delta})$  and  $G_{(i\mu)}^{\theta, \lambda} \underline{\Sigma}^{-1} = O_{\prec}(N^{2\delta})$ . Plugging these estimates into (7.21) and using that  $|\lambda - \lambda'| \prec N^{-1/2}$ , it is easy to estimate the contribution to (7.22) of all terms of (7.21) except the rest term. In order to handle the rest term, we use the rough bounds  $\underline{\Sigma}^{-1} G_{(i\mu)}^{\theta, \lambda'} = O_{\prec}(N)$  and  $G_{(i\mu)}^{\theta, \lambda'} \underline{\Sigma}^{-1} = O_{\prec}(N)$ , which may be deduced from Lemma 4.9 and a simple modification of Lemma 4.8.  $\square$

To simplify notation, we introduce the function

$$f_{(i\mu)}(\lambda) := F_{\mathbf{v}}^p(X_{(i\mu)}^{\theta,\lambda}), \quad (7.24)$$

where we omit the dependence on  $\theta$ ,  $p$ ,  $\mathbf{v}$ , and  $z$  from our notation. (Recall that  $p$  is fixed and all estimates are uniform in  $z \in \widehat{\mathbf{S}}_m$ ,  $\mathbf{v} \in \mathbb{S}$ , and  $\theta \in [0, 1]$ .) We denote by  $f_{(i\mu)}^{(n)}$  the  $n$ -th derivative of  $f_{(i\mu)}$ .

The following result is easy to deduce from (7.21) and Lemma 7.14.

LEMMA 7.15. *Suppose that  $y$  is a random variable satisfying  $|y| \prec N^{-1/2}$ . Then for any fixed  $n \in \mathbb{N}$  we have*

$$|f_{(i\mu)}^{(n)}(y)| \prec N^{2\delta(p+n)}. \quad (7.25)$$

By Taylor expansion, we therefore have

$$f_{(i\mu)}(y) = \sum_{n=1}^{4p} \frac{y^n}{n!} f_{(i\mu)}^{(n)}(0) + O_{\prec}(\Psi^p).$$

From Lemma 7.15 and Lemma 7.7, we get

$$\begin{aligned} \left[ \mathbb{E} F_{\mathbf{v}}^p(X_{(i\mu)}^{\theta, X_{i\mu}^1}) - \mathbb{E} F_{\mathbf{v}}^p(X_{(i\mu)}^{\theta, 0}) \right] &= \mathbb{E}[f_{(i\mu)}(X_{i\mu}^1) - f_{(i\mu)}(0)] \\ &= \mathbb{E} f_{(i\mu)}(0) + \frac{1}{2N} \mathbb{E} f_{(i\mu)}^{(2)}(0) + \sum_{n=4}^{4p} \frac{1}{n!} \mathbb{E} f_{(i\mu)}^{(n)}(0) \mathbb{E}(X_{i\mu}^1)^n + O_{\prec}(\Psi^p), \end{aligned} \quad (7.26)$$

where we used that  $X_{i\mu}^1$  has vanishing first and third moments, by (7.1), and its variance is equal to  $1/N$ . Recalling our goal (7.18), we therefore find that we only have to prove

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} |\mathbb{E} f_{(i\mu)}^{(n)}(0)| = O\left((N^{C_0\delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X^{\theta})\right) \quad (7.27)$$

for  $n = 4, \dots, 4p$ . Here we used the bounds (2.5).

In order to obtain a self-consistent estimate in terms of the matrix  $X^{\theta}$  on the right-hand side of (7.27), we need to replace the matrix  $X_{(i\mu)}^{\theta, 0}$  in  $f_{(i\mu)}(0) = F_{\mathbf{v}}^p(X_{(i\mu)}^{\theta, 0})$  (and its derivatives) with  $X^{\theta} = X_{(i\mu)}^{\theta, X_{i\mu}^{\theta}}$ . We shall do this by an other application of (7.21).

LEMMA 7.16. *Suppose that*

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} |\mathbb{E} f_{(i\mu)}^{(n)}(X_{i\mu}^{\theta})| = O\left((N^{C_0\delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X^{\theta})\right) \quad (7.28)$$

*holds for  $n = 4, \dots, 4p$ . Then (7.27) holds for  $n = 4, \dots, 4p$ .*

PROOF. To simplify notation, we abbreviate  $f_{(i\mu)} \equiv f$  and  $X_{i\mu}^{\theta} \equiv \xi$ . The proof consists of a repeated application of the identity

$$\mathbb{E} f^{(l)}(0) = \mathbb{E} f^{(l)}(\xi) - \sum_{k=1}^{4p-l} \mathbb{E} f^{(l+k)}(0) \frac{\mathbb{E} \xi^k}{k!} + O_{\prec}(N^{l/2} \Psi^p) \quad (7.29)$$

for  $l \leq 4p$ , which follows from Lemmas 7.7 and 7.15. Fix  $n = 4, \dots, 4p$ . Using (7.29) we get

$$\begin{aligned}
\mathbb{E}f^{(n)}(0) &= \mathbb{E}f^{(n)}(\xi) - \sum_{k_1 \geq 1} \mathbf{1}(n + k_1 \leq 4p) \mathbb{E}f^{(n+k_1)}(0) \frac{\mathbb{E}\xi^{k_1}}{k_1!} + O_{\prec}(N^{n/2}\Psi^p) \\
&= \mathbb{E}f^{(n)}(\xi) - \sum_{k_1 \geq 1} \mathbf{1}(n + k_1 \leq 4p) \mathbb{E}f^{(n+k_1)}(\xi) \frac{\mathbb{E}\xi^{k_1}}{k_1!} \\
&\quad + \sum_{k_1, k_2 \geq 1} \mathbf{1}(n + k_1 + k_2 \leq 4p) \mathbb{E}f^{(n+k_1+k_2)}(0) \frac{\mathbb{E}\xi^{k_1}}{k_1!} \frac{\mathbb{E}\xi^{k_2}}{k_2!} + O_{\prec}(N^{n/2}\Psi^p) \\
&= \dots = \sum_{q=0}^{4p-n} (-1)^q \sum_{k_1, \dots, k_q \geq 1} \mathbf{1}\left(n + \sum_j k_j \leq 4p\right) \mathbb{E}f^{(n+\sum_j k_j)}(\xi) \prod_{j=1}^q \frac{\mathbb{E}\xi^{k_j}}{k_j!} + O_{\prec}(N^{n/2}\Psi^p).
\end{aligned}$$

The claim now follows easily using (2.5).  $\square$

What therefore remains is to prove (7.28). Since it only involves the matrix ensemble  $X^\theta$ , for the remainder of the proof we abbreviate  $X^\theta \equiv X$ . Recalling the notation (7.24), we find from Lemma 7.16 that it suffices to prove the following result.

LEMMA 7.17. *for any  $n = 4, \dots, 4p$  we have*

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \left( \frac{\partial}{\partial X_{i\mu}} \right)^n F_{\mathbf{v}}^p(X) \right| = O\left((N^{C_0\delta}\Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E}F_{\mathbf{w}}^p(X)\right). \quad (7.30)$$

REMARK 7.18. We conclude this subsection with a general digression on the algebra of the Bernoulli interpolation method underlying the proof of Proposition 7.1, and in particular establish the formula (1.8) from the introduction. Let  $(X_\alpha^0)_\alpha$  and  $(X_\alpha^1)_\alpha$  be arbitrary independent finite families of independent random variables. Define  $X^\theta = (X_\alpha^\theta)_\alpha$  as in (1.6). Then we get

$$\frac{d}{d\theta} \mathbb{E}F(X^\theta) = \sum_{\alpha} \left[ \mathbb{E}F(X_{(\alpha)}^{\theta, X_\alpha^1}) - \mathbb{E}F(X_{(\alpha)}^{\theta, X_\alpha^0}) \right],$$

where  $X_{(\alpha)}^{\theta, x}$  denotes the family obtained from  $X^\theta$  by replacing  $X_\alpha^\theta$  with  $x$ . Fix  $\alpha$  and abbreviate  $f(x) := F(X_{(\alpha)}^{\theta, x})$ ,  $\zeta := X_\alpha^1$ ,  $\zeta' := X_\alpha^0$ , and  $\xi := X_\alpha^\theta$ . We may assume that  $\zeta$ ,  $\zeta'$ , and  $\xi$  are independent. We want to compute the difference

$$\mathbb{E}[f(\zeta) - f(\zeta')] = \mathbb{E}[f(\zeta) - f(0)] - \mathbb{E}[f(\zeta') - f(0)].$$

Since we are only interested in the algebra of the interpolation, we assume for simplicity that all random variables have finite exponential moments and that  $f$  is analytic. (Otherwise, as in the computations above, the expansions have to be truncated.) Repeating the steps of the proof of Lemma 7.16, we find

$$\mathbb{E}(f(\zeta) - f(0)) = \sum_{q \geq 0} (-1)^q \sum_{n, k_1, \dots, k_q \geq 1} \mathbb{E}f^{(n+k_1+\dots+k_q)}(\xi) \frac{\mathbb{E}\xi^n}{n!} \prod_{j=1}^q \frac{\mathbb{E}\xi^{k_j}}{k_j!} = \sum_{m \geq 1} K_m(\zeta, \xi) \mathbb{E}f^{(m)}(z),$$

where

$$K_m(\zeta, \xi) := \sum_{q \geq 0} (-1)^q \sum_{n, k_1, \dots, k_q \geq 1} \mathbf{1}(n + k_1 + \dots + k_q = m) \frac{\mathbb{E}\zeta^n}{n!} \prod_{j=1}^q \frac{\mathbb{E}\xi^{k_j}}{k_j!} = \frac{1}{m!} \left( \frac{d}{dt} \right)^m \Big|_{t=0} \frac{\mathbb{E}e^{t\zeta} - 1}{\mathbb{E}e^{t\xi}}.$$

Together with the same computation for the difference  $\mathbb{E}(f(\zeta') - f(0))$ , this yields the identity (1.8) from the introduction, with  $K_{m,\alpha}^\theta := K_m(X_\alpha^1, X_\alpha^\theta) - K_m(X_\alpha^0, X_\alpha^\theta)$ .

**7.5. Introduction of words and conclusion of the proof.** In order to prove (7.30), we shall have to exploit the detailed structure of the derivatives in on the left-hand side of (7.30). The following definition introduces the basic algebraic objects that we shall use.

**DEFINITION 7.19 (WORDS).** *We consider words  $w \in \mathcal{W}$  of even length in the two letters  $\{\mathbf{i}, \boldsymbol{\mu}\}$ . We denote by  $2n(w)$  the length of the word  $w$ , where  $n(w) = 0, 1, 2, \dots$ . We always use bold symbols to denote the letters of words. We use the notation*

$$w = \mathbf{t}_1 \mathbf{s}_2 \mathbf{t}_2 \mathbf{s}_3 \cdots \mathbf{t}_n \mathbf{s}_{n+1} \quad (7.31)$$

for a word of length  $2n = 2n(w)$ . For  $n = 0, 1, 2, \dots$  we introduce the subset  $\mathcal{W}_n := \{w \in \mathcal{W} : n(w) = n\}$  of words of length  $2n$ . We require that each word  $w \in \mathcal{W}_n$  satisfy  $\mathbf{t}_l \mathbf{s}_{l+1} \in \{\mathbf{i}\boldsymbol{\mu}, \boldsymbol{\mu}\mathbf{i}\}$  for all  $1 \leq l \leq n$ .

Next, we assign to each letter  $*$  its value  $[*] \equiv [*]_{i,\mu} \in \mathcal{I}$  through

$$[\mathbf{i}] := i, \quad [\boldsymbol{\mu}] := \mu.$$

(Our choice of the names of the two letters is suggestive of their value. Note, however, that it is important to distinguish the abstract letter from its value, which is an index in  $\mathcal{I}$  and may be used as a summation index.)

Finally, to each word  $w \in \mathcal{W}$  we assign a random variable  $A_{\mathbf{v},i,\mu}(w)$  as follows. If  $n(w) = 0$  (i.e.  $w$  is the empty word) we define

$$A_{\mathbf{v},i,\mu}(w) := G_{\bar{\mathbf{v}}\bar{\mathbf{v}}} - \Pi_{\bar{\mathbf{v}}\bar{\mathbf{v}}}.$$

If  $n(w) \geq 1$  with  $w$  as in (7.31) we define

$$A_{\mathbf{v},i,\mu}(w) := G_{\bar{\mathbf{v}}[\mathbf{t}_1]} G_{[\mathbf{s}_2][\mathbf{t}_2]} \cdots G_{[\mathbf{s}_n][\mathbf{t}_n]} G_{[\mathbf{s}_{n+1}]\bar{\mathbf{v}}}. \quad (7.32)$$

Definition 7.19 is constructed so that

$$\left( \frac{\partial}{\partial X_{i\mu}} \right)^n (G_{\bar{\mathbf{v}}\bar{\mathbf{v}}} - \Pi_{\bar{\mathbf{v}}\bar{\mathbf{v}}}) = (-1)^n \sum_{w \in \mathcal{W}_n} A_{\mathbf{v},i,\mu}(w)$$

for any  $n = 0, 1, 2, \dots$ . This may be easily deduced from (7.21). Using Leibnitz's rule we conclude that

$$\begin{aligned} \left( \frac{\partial}{\partial X_{i\mu}} \right)^n F_{\mathbf{v}}^p(X) &= (-1)^n \sum_{n_1, \dots, n_{p/2} \geq 0} \sum_{\tilde{n}_1, \dots, \tilde{n}_{p/2} \geq 0} \mathbf{1} \left( \sum_r (n_r + \tilde{n}_r) = n \right) \frac{n!}{\prod_r n_r! \tilde{n}_r!} \\ &\quad \times \prod_r \left( \sum_{w_r \in \mathcal{W}_{n_r}} \sum_{\tilde{w}_r \in \mathcal{W}_{\tilde{n}_r}} A_{\mathbf{v},i,\mu}(w_r) \overline{A_{\mathbf{v},i,\mu}(\tilde{w}_r)} \right). \end{aligned}$$

To prove (7.30), it therefore suffices to prove that

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \prod_{r=1}^{p/2} (A_{\mathbf{v}, i, \mu}(w_r) \overline{A_{\mathbf{v}, i, \mu}(\tilde{w}_r)}) \right| = O\left((N^{C_0 \delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X)\right) \quad (7.33)$$

for  $4 \leq n \leq 4p$  and words  $w_r, \tilde{w}_r \in \mathcal{W}$  satisfying  $\sum_r (n(w_r) + n(\tilde{w}_r)) = n$ . To avoid irrelevant notational complications arising from the complex conjugates, we in fact prove that

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \prod_{r=1}^p A_{\mathbf{v}, i, \mu}(w_r) \right| = O\left((N^{C_0 \delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X)\right) \quad (7.34)$$

for  $4 \leq n \leq 4p$  and words  $w_r \in \mathcal{W}$  satisfying  $\sum_r n(w_r) = n$ . (The proof of (7.33) is the same with slightly heavier notation.) Treating words  $w_r$  with  $n(w_r) = 0$  separately, we find that it suffices to prove

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \mathbb{E} \left| A_{\mathbf{v}, i, \mu}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{v}, i, \mu}(w_r) \right| = O\left((N^{C_0 \delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X)\right) \quad (7.35)$$

for  $4 \leq n \leq 4p$ ,  $1 \leq q \leq p$ , and words  $w_r \in \mathcal{W}$  satisfying  $\sum_r n(w_r) = n$ ,  $n(w_0) = 0$ , and  $n(w_r) \geq 1$  for  $r \geq 1$ . Note that we also have the bound  $q \leq n$ .

In order to estimate (7.35), we introduce the quantity

$$R_s := |G_{\mathbf{v}s}| + |G_{s\mathbf{v}}|.$$

LEMMA 7.20. *For  $w \in \mathcal{W}$  we have the rough bound*

$$|A_{\mathbf{v}, i, \mu}(w)| \prec N^{2\delta(n(w)+1)}. \quad (7.36)$$

Moreover, for  $n(w) \geq 1$  we have

$$|A_{\mathbf{v}, i, \mu}(w)| \prec (R_i^2 + R_\mu^2) N^{2\delta(n(w)-1)}. \quad (7.37)$$

Finally, for  $n(w) = 1$  we have the sharper bound

$$|A_{\mathbf{v}, i, \mu}(w)| \prec R_i R_\mu. \quad (7.38)$$

PROOF. The estimates (7.36) and (7.37) follow easily from Lemma 7.13 and the definition (7.32). The estimate (7.38) follows from the constraint  $\mathbf{t}_1 \neq \mathbf{s}_2$  in Definition 7.19.  $\square$

In addition to the high-probability bounds from Lemma 7.20, we have the rough estimate

$$\mathbb{E} |A_{\mathbf{v}, i, \mu}(w)|^2 \leq N^C \quad (7.39)$$

for some  $C > 0$  and all  $w \in \mathcal{W}$  satisfying  $n(w) \leq 4p$ ; this follows easily from Definition 7.19 and Lemma 4.9.

By pigeonholing on the words appearing on the left-hand side of (7.35), if  $n \leq 2q - 2$  then there exist at least two words  $w_r$  satisfying  $n(w_r) = 1$ . Using Lemma 7.20 we therefore get

$$\left| A_{\mathbf{v}, i, \mu}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{v}, i, \mu}(w_r) \right| \prec N^{2\delta(n+q)} F_{\mathbf{v}}^{p-q}(X) \left( \mathbf{1}(n \geq 2q - 1)(R_i^2 + R_\mu^2) + \mathbf{1}(n \leq 2q - 2)R_i^2 R_\mu^2 \right). \quad (7.40)$$

From Lemma 4.6 combined with Lemma 4.8 we get

$$\frac{1}{N} \sum_{i \in \mathcal{I}_M} R_i^2 + \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} R_\mu^2 \prec \frac{\operatorname{Im} G_{\bar{\mathbf{v}}\mathbf{v}} + \eta}{N\eta} \prec N^{2\delta} \frac{\operatorname{Im} m + N^{C_0\delta}\Psi}{N\eta} \leq N^{(C_0+2)\delta}\Psi^2, \quad (7.41)$$

where in the second step we used (7.20), and in the last step the definition of  $\Psi$ . Inserting (7.41) into (7.40), we get using Lemma 7.7 and (7.39) that the left-hand side of (7.35) is bounded by

$$N^{-n/2+2} N^{3\delta(n+q)} \mathbb{E} F_{\mathbf{v}}^{p-q}(X) \left( \mathbf{1}(n \geq 2q - 1) (N^{C_0\delta/2}\Psi)^2 + \mathbf{1}(n \leq 2q - 2) (N^{C_0\delta/2}\Psi)^4 \right).$$

Using that  $\Psi \geq cN^{-1/2}$ , we find that the left-hand side of (7.35) is bounded by

$$\begin{aligned} & N^{3\delta(n+q)} \mathbb{E} F_{\mathbf{v}}^{p-q}(X) \left( \mathbf{1}(n \geq 2q - 1) (N^{C_0\delta/2}\Psi)^{n-2} + \mathbf{1}(n \leq 2q - 2) (N^{C_0\delta/2}\Psi)^n \right) \\ & \leq \mathbb{E} F_{\mathbf{v}}^{p-q}(X) \left( \mathbf{1}(n \geq 2q - 1) (N^{(C_0/2+12)\delta}\Psi)^{n-2} + \mathbf{1}(n \leq 2q - 2) (N^{(C_0/2+12)\delta}\Psi)^n \right), \end{aligned}$$

where we used that  $q \leq n$  and  $n \geq 4$ . Choose  $C_0 \geq 25$ . Then, by assumption (7.9) on  $\delta$  we have  $N^{(C_0/2+12)\delta}\Psi \leq 1$ . Moreover, if  $n \geq 4$  and  $n \geq 2q - 1$  then  $n \geq q + 2$ . We conclude that the left-hand side of (7.35) is bounded by

$$\mathbb{E} F_{\mathbf{v}}^{p-q}(X) (N^{C_0\delta}\Psi)^q. \quad (7.42)$$

Now (7.35) follows using Young's inequality. This concludes the proof of (7.30), and hence of Lemma 7.11. The proof of Proposition 7.1 is therefore complete.

## 8. Self-consistent comparison II: general $X$

In this section and the next we prove the following result, which is Proposition 7.1 without the condition (7.1).

**PROPOSITION 8.1.** *Suppose that the assumptions of Theorem 3.21 hold. If the anisotropic local law holds with parameters  $(X^{\text{Gauss}}, \Sigma, \mathbf{S})$ , then the anisotropic local law holds with parameters  $(X, \Sigma, \mathbf{S})$ .*

**8.1. Overview of the proof of Proposition 8.1.** The proof of Proposition 8.1 builds on that of Proposition 7.1. Throughout this section, we take over the notations of Section 7 without further comment. In particular, as in (7.9),  $C_0$  is some large enough constant (which may depend only on  $\tau$ ) and  $\delta > 0$  is fixed and small, satisfying (7.9).

As in Section 7.4, we suppose that  $(\mathbf{A}_{m-1})$  holds, and we fix  $z \in \widehat{\mathbf{S}}_m$  which we consistently omit from our notation. From the assumption  $(\mathbf{A}_{m-1})$  and Lemma 7.13 we get the bounds (7.19) and (7.20), which we summarize here for easy reference:

$$\underline{\Sigma}^{-1}(G - \Pi)\underline{\Sigma}^{-1} = O_{\prec}(N^{2\delta}), \quad \text{Im } G_{\bar{\mathbf{v}}\mathbf{v}} \prec N^{2\delta}(\text{Im } m + N^{C_0\delta}\Psi) \quad (8.1)$$

for all deterministic  $\mathbf{v} \in \mathbb{S}$ .

The assumption (7.1) was used in the proof of Proposition 7.1 only in (7.26), where it ensured that the summation over  $n$  starts from 4 instead of 3. Without the assumption (7.1), we in addition have to estimate the term  $n = 3$  in (7.26). It therefore suffices to prove the following result.

LEMMA 8.2. *Let  $z \in \widehat{\mathbf{S}}_m$ , and suppose that (8.1) holds at  $z$ . There exists a constant  $C_0$  such that, for any  $\delta$  satisfying (7.9), the estimate (7.27) holds for  $n = 3$ .*

We shall in fact prove a slightly stronger bound than (7.27), where the term  $\sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X^\theta, z)$  on the right-hand side of (7.27) is replaced with  $\sup\{\mathbb{E} F_{\mathbf{w}}^p(X^\theta, z) : \mathbf{w} \in \{\mathbf{v}\} \cup \{\mathbf{e}_\mu : \mu \in \mathcal{I}_N\}\}$ . However, (7.27) is simpler to state and strong enough for our purposes.

As in Lemma 7.16, one may easily replace the matrix  $X_{i\mu}^{\theta,0}$  in the definition of  $f_{i\mu}^{(3)}(0)$  with  $X^\theta$ . As in Lemma 7.17, from now on we abbreviate  $X^\theta \equiv X$ . Thus, we find that in order to prove Lemma 8.2 it suffices to prove the following result, which complements Lemma 7.17.

LEMMA 8.3. *Let  $z \in \widehat{\mathbf{S}}_m$ , and suppose that (8.1) holds at  $z$ . There exists a constant  $C_0$  such that, for any  $\delta$  satisfying (7.9), the estimate (7.30) holds for  $n = 3$ .*

Next, recall the definition of words from Definition 7.19. As in (7.33)–(7.35), Lemma 8.3 is proved provided we can show the following result, which is analogous to (7.35).

LEMMA 8.4. *Let  $z \in \widehat{\mathbf{S}}_m$ , and suppose that (8.1) holds at  $z$ . Let  $1 \leq q \leq 3$  and choose words  $w_1, \dots, w_p \in \mathcal{W}$  satisfying  $n(w_r) \geq 1$  for  $r \leq q$ ,  $n(w_r) = 0$  for  $r \geq q + 1$ , and  $\sum_r n(w_r) = 3$ . Then (7.34) holds with  $n = 3$ .*

We now explain some key ingredients of the proof of Lemma 8.4. The first main difficulty is to extract an additional factor  $N^{-1/2}$  from the left-hand side of (7.34) with  $n = 3$ , so as to obtain in total a factor  $N^{-2}$  that will cancel the number of terms in the summation. The second main difficulty is to extract a factor  $\Psi^q$  from the expectation, so as to apply estimates of the form  $\mathbb{E}|A_{\mathbf{v},i,\mu}(w_0)^{p-q}|\Psi^q \leq (N^{C_\delta}\Psi)^p + \mathbb{E}F_{\mathbf{v}}^p(X)$ . In order to extract the factor  $N^{-1/2}$ , we analyse the dependence of each factor  $A_{\mathbf{v},i,\mu}(w_r)$  on the  $\mu$ -th column of  $X$ , i.e. the variables  $X_\mu := (X_{i\mu})_{i \in \mathcal{I}_M}$ . We express each term, up to negligible error terms, as a polynomial in  $X_\mu$  whose coefficients are independent of  $X_\mu$  (i.e.  $X^{(\mu)}$ -measurable). Hence we may take the conditional expectation  $\mathbb{E}(\cdot | X^{(\mu)})$  of the product of the variables  $X_\mu$ , which results in a partition of all entries of  $X_\mu$ . The extra factor of  $N^{-1/2}$  then follows using a parity argument similar to the one first used in [7]: the degree of the polynomial is odd, so that a perfect matching, which would give a contribution of order  $N^0$ , is not possible.

Beyond the factor of  $N^{-1/2}$ , the need to obtain the bound from (7.34) that is strong enough to close the self-consistent estimate presents significant difficulties, which we outline briefly. These difficulties are roughly of two types. (a) The off-diagonal entries of  $G^{(\mu)}$  are in general not small; only entries of  $G^{(\mu)} - \Pi$  are small. (b) A priori, using the estimate (7.19), the entries of  $G^{(\mu)}$  are not bounded by  $O_{\prec}(1)$  but by  $O_{\prec}(N^{2\delta})$ . This would yield a bound on the coefficients of the polynomial in  $X_\mu$  that is a high power of  $N^{2\delta}$ , which may become too large to conclude the proof.



We deal with difficulty (a) not by estimating individual entries  $G_{\mathbf{v}t}^{(\mu)}$  but by exploiting the extra summations generated by the polynomial expansion, expressing our error bounds in terms of as many summed entries of the form

$$\frac{1}{N} \sum_{j \in \mathcal{I}_M} |G_{\mathbf{v}j}^{(\mu)}|^d \quad (8.2)$$

as possible, where  $d \in \mathbb{N}$ . For instance, similarly to (7.5), if  $d = 2$  then (8.2) may be estimated by  $\frac{\text{Im } G_{\mathbf{v}\mathbf{v}} + \eta}{N\eta}$ , which is much smaller than the naive bound  $N^{4\delta}$ . (See (8.28) below for a precise statement.) The factor  $\text{Im } G_{\mathbf{v}\mathbf{v}}$  may be estimated using the a priori bound (7.20). The proof relies on a careful balance of the integers  $d$  from (8.2) versus the number of entries of  $G$  that cannot be estimated in this way. We also note that, while  $d = 2$  gives a stronger bound than  $d = 1$  in (8.2), for  $d \geq 3$  the bound (8.2) cannot be improved over the corresponding bound for  $d = 2$ . In this sense, powers of  $d$  larger than 2 lead to wasted factors of  $\Psi$ , and we have to make sure that the final combined power of  $\Psi$  and  $F_{\mathbf{v}}$  is large enough despite the possibility of indices  $d$  larger than 2.

Next, we deal with difficulty (b) by showing that polynomials whose coefficients consist of many entries of  $G^{(\mu)}$ , each estimated by  $O_{\prec}(N^{2\delta})$ , also give rise to many factors of the form (8.2) with large enough  $d$ , which compensate the powers of  $O_{\prec}(N^{2\delta})$ . Hence, controlling the number of factors of the form (8.2) and the corresponding indices  $d$  is crucial to obtain a sufficiently high power of  $\Psi + F_{\mathbf{v}}$  and a sufficiently small prefactor. The details of this counting are explained in Section 8.5.

We conclude this sketch by noting that, in order to get rid of the omnipresent upper indices  $(\mu)$  generated by the expansion in  $X_{\mu}$ , we use identities from Lemma 4.4, which generate error terms of the form  $F_{\mathbf{e}_{\mu}}(X)$  in addition to  $F_{\mathbf{v}}(X)$  (see (8.10) below). Since  $\mu$  may take any value in  $\mathcal{I}_N$ , we need the general self-consistent error term  $\sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X)$  on the right-hand side of (7.34) instead of the simple  $\mathbb{E} F_{\mathbf{v}}^p(X)$ .

**8.2. Tagged words.** We now move on to the actual proof of Lemma 8.4. The index  $\mu \in \mathcal{I}_N$  will play a distinguished role. In a preliminary step, we show that, up to an affordable error, the random variables  $A_{\mathbf{v},i,\mu}(w)$  in (7.34) may be replaced with

$$\widehat{A}_{\mathbf{v},i,\mu}(w) := \begin{cases} G_{\bar{\mathbf{v}}\bar{\mathbf{v}}} - \Pi_{\bar{\mathbf{v}}\bar{\mathbf{v}}} - \bar{\mathbf{v}}(\mu)^2 (G_{\mu\mu} - \Pi_{\mu\mu}) & \text{if } n(w) = 0 \\ A_{\mathbf{v},i,\mu}(w) & \text{if } n(w) \geq 1. \end{cases}$$

The motivation behind the definition of  $\widehat{A}_{\mathbf{v},i,\mu}(w)$  stems from expansion formula (8.3) below, which will play a key role in the proof of Proposition 8.1.

LEMMA 8.5. *Lemma 8.4 holds provided it holds with  $A$  replaced by  $\widehat{A}$  in (7.34).*

PROOF. We write, dropping the subscripts from  $A$  and using  $w_0$  to denote the empty word,

$$\begin{aligned} \left| \prod_{r=1}^p A(w_r) - \prod_{r=1}^p \widehat{A}(w_r) \right| &= \left| (A(w_0)^{p-q} - \widehat{A}(w_0)^{p-q}) \prod_{r=1}^q A(w_r) \right| \\ &\leq p(|A(w_0)| + |\widehat{A}(w_0)|)^{p-q-1} |G_{\mu\mu} - \Pi_{\mu\mu}| \bar{\mathbf{v}}(\mu)^2 \prod_{r=1}^q |A(w_r)| \\ &\prec (F_{\mathbf{v}}(X) + F_{\mathbf{e}_{\mu}}(X))^{p-q-1} F_{\mathbf{e}_{\mu}}(X) \bar{\mathbf{v}}(\mu)^2 \prod_{r=1}^q |A(w_r)|. \end{aligned}$$

Next, from Definition 7.19 we conclude, analogously to the proof of Lemma 7.20, that if  $n(w_r) \geq 1$  for  $1 \leq r \leq q$  and  $\sum_{r=1}^q n(w_r) = 3$  then

$$\sum_{i \in \mathcal{I}_M} \prod_{r=1}^q |A(w_r)| \prec \sum_{i \in \mathcal{I}_M} N^{C\delta} R_i^{q-1} \prec NN^{C\delta} (N^{C_0/2} \Psi)^{q-1},$$

where in the last step we used (7.41). We therefore conclude using Lemma 7.7 that

$$\begin{aligned} & N^{-3/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \left[ \prod_{r=1}^p A(w_r) - \prod_{r=1}^p \hat{A}(w_r) \right] \right| \\ & \leq N^{C\delta} N^{-1/2} \sum_{\mu \in \mathcal{I}_N} \bar{\mathbf{v}}(\mu)^2 (N^{C_0/2} \Psi)^{q-1} \mathbb{E} (F_{\mathbf{v}}(X) + F_{\mathbf{e}_\mu}(X))^{p-q-1} F_{\mathbf{e}_\mu}(X) \\ & \leq (N^{(C+C_0/2)\delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X), \end{aligned}$$

where in the last step we used Young's inequality and the estimate  $N^{-1/2} \leq \Psi$ . Here we also used that  $\sum_{\mu \in \mathcal{I}_N} \bar{\mathbf{v}}(\mu)^2 \leq 1$ , as follows from the definition of  $\bar{\mathbf{v}}$  from (7.12) and (3.6). This concludes the proof.  $\square$

Next, we introduce a family of *tagged words*, which refine the words  $w$  from Definition 7.19. They are designed to encode formulas which refine the definition of  $A_{\mathbf{v},i,\mu}(w)$  from Definition 7.19. In order to state them, we shall need a decomposition of entries  $G_{\mathbf{xy}}$  into three pieces, denoted by  $[G_{\mathbf{xy}}]^0$ ,  $[G_{\mathbf{xy}}]^1$ , and  $[G_{\mathbf{xy}}]^2$ . The identity behind this decomposition is

$$G_{\mathbf{xy}} = \mathbf{x}(\mu)\mathbf{y}(\mu)G_{\mu\mu} + G_{\mathbf{xy}}^{(\mu)} + G_{\mu\mu}(G^{(\mu)}X)_{\mathbf{x}\mu}(X^*G^{(\mu)})_{\mu\mathbf{y}} - \mathbf{x}(\mu)G_{\mu\mu}(X^*G^{(\mu)})_{\mu\mathbf{y}} - \mathbf{y}(\mu)G_{\mu\mu}(G^{(\mu)}X)_{\mathbf{x}\mu}, \quad (8.3)$$

which follows from Lemma 4.4. (Recall that an expression of the form  $G_{\mathbf{xy}}^{(\mu)}$  is by definition equal to  $\mathbf{x}^*G^{(\mu)}\mathbf{y}$ , where the matrix multiplication is performed according to Definition 4.1.)

**DEFINITION 8.6.** *Let  $\mathbf{x}, \mathbf{y} \in \{\bar{\mathbf{v}}, \mathbf{e}_\mu, \mathbf{e}_i\}$  with  $\mathbf{x}$  and  $\mathbf{y}$  not both equal to  $\bar{\mathbf{v}}$ . Then the quantities  $[G_{\mathbf{xy}}]^0$ ,  $[G_{\mathbf{xy}}]^1$ , and  $[G_{\mathbf{xy}}]^2$  are uniquely defined by requiring that they do not contain factors of  $G_{\mu\mu}$  or  $\bar{\mathbf{v}}(\mu)$ , and that (8.3) reads*

$$G_{\mathbf{xy}} = [G_{\mathbf{xy}}]^0 + G_{\mu\mu}[G_{\mathbf{xy}}]^1 + \bar{\mathbf{v}}(\mu)G_{\mu\mu}[G_{\mathbf{xy}}]^2. \quad (8.4)$$

*In particular,  $[G_{\mathbf{xy}}]^0$  is  $X^{(\mu)}$ -measurable.*

Explicitly, this decomposition reads

$$\begin{aligned} G_{\bar{\mathbf{v}}\mu} &= \bar{\mathbf{v}}(\mu)G_{\mu\mu} - G_{\mu\mu}(G^{(\mu)}X)_{\bar{\mathbf{v}}\mu} = \bar{\mathbf{v}}(\mu)G_{\mu\mu}[G_{\bar{\mathbf{v}}\mu}]^2 + G_{\mu\mu}[G_{\bar{\mathbf{v}}\mu}]^1, \\ G_{\bar{\mathbf{v}}i} &= G_{\bar{\mathbf{v}}i}^{(\mu)} + G_{\mu\mu}(G^{(\mu)}X)_{\bar{\mathbf{v}}\mu}(X^*G^{(\mu)})_{\mu i} - \bar{\mathbf{v}}(\mu)G_{\mu\mu}(X^*G^{(\mu)})_{\mu i} = [G_{\bar{\mathbf{v}}i}]^0 + G_{\mu\mu}[G_{\bar{\mathbf{v}}i}]^1 + \bar{\mathbf{v}}(\mu)G_{\mu\mu}[G_{\bar{\mathbf{v}}i}]^2, \\ G_{i\mu} &= -G_{\mu\mu}(G^{(\mu)}X)_{i\mu} = G_{\mu\mu}[G_{i\mu}]^1, \\ G_{ii} &= G_{ii}^{(\mu)} + G_{\mu\mu}(G^{(\mu)}X)_{i\mu}(X^*G^{(\mu)})_{\mu i} = [G_{ii}]^0 + G_{\mu\mu}[G_{ii}]^1. \end{aligned}$$

Throughout the following we abbreviate  $\mathbb{Z}_k := \llbracket 0, k-1 \rrbracket$ .

DEFINITION 8.7 (TAGGED WORDS). Let  $w \in \mathcal{W}$  be a word from Definition 7.19. A tagged word is a pair  $(w, \sigma)$ , where  $\sigma = (\sigma(l))_{l=1}^{n(w)+1} \in \mathbb{Z}_3^{n(w)+1}$ . We assign to each tagged word  $(w, \sigma)$  with  $w \in \mathcal{W}_n$  of the form (7.31) a random variable  $A_{\mathbf{v}, i, \mu}(w, \sigma)$  according to the following construction.

(i) If  $n = 0$  (i.e.  $w$  is the empty word) we define

$$\begin{aligned} A_{\mathbf{v}, i, \mu}(w, 0) &:= G_{\bar{\mathbf{v}}\bar{\mathbf{v}}}^{(\mu)} - \Pi_{\bar{\mathbf{v}}\bar{\mathbf{v}}}^{(\mu)}, \\ A_{\mathbf{v}, i, \mu}(w, 1) &:= (G^{(\mu)} X)_{\bar{\mathbf{v}}\mu} (X^* G^{(\mu)})_{\mu\bar{\mathbf{v}}}, \\ A_{\mathbf{v}, i, \mu}(w, 2) &:= -(X^* G^{(\mu)})_{\mu\bar{\mathbf{v}}} - (G^{(\mu)} X)_{\bar{\mathbf{v}}\mu}, \end{aligned}$$

where we introduced the matrix  $\Pi^{(\mu)} := (\Pi_{st} : s, t \in \mathcal{I} \setminus \{\mu\})$ .

(ii) If  $n \geq 1$  we define

$$A_{\mathbf{v}, i, \mu}(w, \sigma) := [G_{\bar{\mathbf{v}}[\mathbf{t}_1]}]^{\sigma(1)} [G_{[\mathbf{s}_2][\mathbf{t}_2]}]^{\sigma(2)} \cdots [G_{[\mathbf{s}_n][\mathbf{t}_n]}]^{\sigma(n)} [G_{[\mathbf{s}_{n+1}]\bar{\mathbf{v}}}]^{\sigma(n+1)}. \quad (8.5)$$

(iii) We define

$$|\sigma|_{\boldsymbol{\mu}} := \sum_l \mathbf{1}(\sigma(l) \geq 1), \quad |\sigma|_{\mathbf{v}} := \sum_l \mathbf{1}(\sigma(l) = 2).$$

(Note that the choice of indices  $\boldsymbol{\mu}$  and  $\mathbf{v}$  is merely suggestive of the meaning of  $|\cdot|_{\boldsymbol{\mu}}$  and  $|\cdot|_{\mathbf{v}}$ ; these quantities clearly do not depend on  $\mu$  and  $\mathbf{v}$ .)

(iv) By construction,  $A_{\mathbf{v}, i, \mu}(w, \sigma)$  is a homogeneous polynomial in the variables  $\{X_{k\mu}\}_{k \in \mathcal{I}_M}$ , whose coefficients are  $X^{(\mu)}$ -measurable. We use  $\deg(A_{\mathbf{v}, i, \mu}(w, \sigma)) \in \mathbb{N}$  to denote its (total) degree.

From (8.4), (8.3), and Definition 8.7 we find that for  $w \in \mathcal{W}_n$  we have

$$\hat{A}_{\mathbf{v}, i, \mu}(w) = \sum_{\sigma \in \mathbb{Z}_3^{n+1}} A_{\mathbf{v}, i, \mu}(w, \sigma) (G_{\mu\mu})^{|\sigma|_{\boldsymbol{\mu}}} \bar{\mathbf{v}}(\mu)^{|\sigma|_{\mathbf{v}}}. \quad (8.6)$$

Recalling Lemma 8.5, we conclude that, in order to prove Lemma 8.4, it suffices to prove the following result.

LEMMA 8.8. Suppose that (8.1) holds. Let  $1 \leq q \leq 3$  and choose words  $w_1, \dots, w_p \in \mathcal{W}$  satisfying  $n(w_r) \geq 1$  for  $r \leq q$ ,  $n(w_r) = 0$  for  $r \geq q+1$ , and  $\sum_r n(w_r) = 3$ . For  $r = 1, \dots, p$  let  $\sigma_r \in \mathbb{Z}_3^{n(w_r)+1}$ . Then we have for all  $\mathbf{v} \in \mathbb{S}$

$$N^{-3/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \left( (G_{\mu\mu})^{d_{\boldsymbol{\mu}}} \bar{\mathbf{v}}(\mu)^{d_{\mathbf{v}}} \prod_{r=1}^p A_{\mathbf{v}, i, \mu}(w_r, \sigma_r) \right) \right| = O \left( (N^{C_0 \delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X) \right), \quad (8.7)$$

where we abbreviated

$$d_* \equiv d_*(\sigma_1, \dots, \sigma_p) := \sum_{r=1}^p |\sigma_r|_* \quad (8.8)$$

for  $* = \boldsymbol{\mu}, \mathbf{v}$ .

**8.3. Preliminary estimates and the entries  $G_{\mu\mu}$ .** We shall require the following rough bounds, which refine those of Lemma 7.20.

LEMMA 8.9 (ROUGH BOUNDS ON  $A_{\mathbf{v},i,\mu}(w, \sigma)$ ). *Suppose that (8.1) holds. Let  $(w, \sigma)$  be a tagged word. Then*

$$|A_{\mathbf{v},i,\mu}(w, \sigma)| \prec N^{2\delta(n(w)+1)}. \quad (8.9)$$

Moreover, if  $n(w) = 0$  then

$$|A_{\mathbf{v},i,\mu}(w, \sigma)| \prec N^{(C_0/2+1)\delta}\Psi + F_{\mathbf{v}}(X) + F_{\mathbf{e}_\mu}(X). \quad (8.10)$$

PROOF. Using a large deviation estimate (see [7, Lemma 3.1]) combined with Lemmas 4.6 and 4.8 we get

$$|(G^{(\mu)}X)_{\bar{\mathbf{w}}\mu}|^2 + |(X^*G^{(\mu)})_{\mu\bar{\mathbf{w}}}|^2 \prec \frac{1}{N} \sum_{j \in \mathcal{I}_M} |G_{\bar{\mathbf{w}}j}^{(\mu)}|^2 + \frac{1}{N} \sum_{j \in \mathcal{I}_M} |G_{j\bar{\mathbf{w}}}^{(\mu)}|^2 \prec \frac{\operatorname{Im} G_{\bar{\mathbf{w}}\bar{\mathbf{w}}}^{(\mu)} + \eta}{N\eta} \quad (8.11)$$

for any deterministic  $\mathbf{w}$  satisfying  $|\mathbf{w}| \leq C$ . Using (8.3) with  $\mathbf{x} = \mathbf{y} = \bar{\mathbf{w}}$ , (8.11), (8.1),  $\Psi \geq cN^{-1/2}$ , and the trivial estimate  $|\bar{\mathbf{w}}(\mu)| \leq 1$ , we find

$$\operatorname{Im} G_{\bar{\mathbf{w}}\bar{\mathbf{w}}}^{(\mu)} = \operatorname{Im} G_{\bar{\mathbf{w}}\bar{\mathbf{w}}} + O_{\prec} \left( \operatorname{Im} G_{\mu\mu} + N^{2\delta} \sqrt{\frac{\operatorname{Im} G_{\bar{\mathbf{w}}\bar{\mathbf{w}}}^{(\mu)} + \eta}{N\eta}} \right),$$

We conclude that

$$\operatorname{Im} G_{\bar{\mathbf{w}}\bar{\mathbf{w}}}^{(\mu)} \prec \operatorname{Im} G_{\bar{\mathbf{w}}\bar{\mathbf{w}}} + \operatorname{Im} G_{\mu\mu} + N^{3\delta}\Psi \prec N^{2\delta} \operatorname{Im} m + N^{(C_0+2)\delta}\Psi,$$

where in the second step we used (8.1). Therefore,

$$\frac{\operatorname{Im} G_{\bar{\mathbf{w}}\bar{\mathbf{w}}}^{(\mu)}}{N\eta} \prec N^{(C_0+2)\delta}\Psi^2. \quad (8.12)$$

Now (8.10) with  $\sigma \geq 1$  follows easily from (8.11) and (8.12) with  $\mathbf{w} = \mathbf{v}$ . Moreover, (8.10) with  $\sigma = 0$  follows using (8.3) and (8.1).

Next, in order to prove (8.9), we deduce Definition 8.6 and (8.1), as well as from (8.11) and (8.12) with  $\mathbf{w} = \mathbf{v}, \mathbf{e}_\mu, \Sigma \mathbf{e}_i$ , that

$$|[G_{\bar{\mathbf{v}}i}]^\sigma| + |[G_{i\bar{\mathbf{v}}}]^\sigma| + |[G_{\bar{\mathbf{v}}\mu}]^\sigma| + |[G_{\mu\bar{\mathbf{v}}}]^\sigma| + |[G_{i\mu}]^\sigma| + |[G_{\mu i}]^\sigma| \prec N^{2\delta}, \quad (8.13)$$

for all  $\sigma = 0, 1, 2$ . This concludes the proof.  $\square$

In order to analyse the left-hand side of (8.7), we need to exhibit the precise  $X$ -dependence of the factor  $(G_{\mu\mu})^{d_\mu}$ . To that end, we write, using (4.4),  $G_{\mu\mu} = (-z - Y_\mu - Z_\mu)^{-1}$ , where we defined

$$Z_\mu := (X^*G^{(\mu)}X)_{\mu\mu} - Y_\mu, \quad Y_\mu := \mathbb{E} \left[ (X^*G^{(\mu)}X)_{\mu\mu} | X^{(\mu)} \right] = \frac{1}{N} \sum_{j \in \mathcal{I}_M} G_{jj}^{(\mu)}.$$

Using the large deviation estimates from [7, Lemma 3.1], we find  $|Z_\mu| \prec N^{-\tau/2+2\delta}$ . Since  $|G_{\mu\mu}| \prec N^{2\delta}$  by (8.1), we therefore deduce using  $-\tau/2 + 2\delta < -2\delta$  that

$$|-z - Y_\mu|^{-1} \prec N^{2\delta}. \quad (8.14)$$

Hence there exists a constant  $K \equiv K(\tau)$  such that

$$G_{\mu\mu} = \sum_{k=0}^K (-z - Y_\mu)^{-k-1} Z_\mu^k + O_{\prec}(N^{-10}).$$

Plugging in the definition of  $Z_\mu$ , we find

$$G_{\mu\mu} = \sum_{k=0}^K \mathcal{Y}_{\mu,k} (X^* G^{(\mu)} X)_{\mu\mu}^k + O_{\prec}(N^{-10}), \quad (8.15)$$

where the coefficients  $\mathcal{Y}_{\mu,k}$  are  $X^{(\mu)}$ -measurable and satisfy the bound  $|\mathcal{Y}_{\mu,k}| \prec N^{C\delta}$ . Here we used (8.14) and the estimate  $|G_{jj}^{(\mu)}| \prec N^{2\delta}$ , which follows from (8.1), (8.3), (8.11), and (8.12) with  $\bar{\mathbf{w}} = \mathbf{x} = \mathbf{y} = \mathbf{e}_j$ . The precise form of  $\mathcal{Y}_{\mu,k}$  is unimportant. In order to apply Lemma 7.7 in the following, we shall also need the rough bound

$$\mathbb{E}|\mathcal{Y}_{\mu,k}|^q \leq N^{C_{p,q}} \quad (8.16)$$

for all  $q \in \mathbb{N}$ , which may be easily deduced from  $\mathbb{E}(|Y_\mu| + |-z - Y_\mu|^{-1})^\ell \leq N^{C_p}$  for any  $\ell \leq 16K$ . This latter estimate follows easily from the deterministic estimate  $|Y_\mu| + |-z - Y_\mu|^{-1} \leq CN$ , where we used that  $\text{Im}(-z - Y_\mu) \leq \text{Im}(-z) \leq -N^{-1}$ .

Now we replace the factors  $G_{\mu\mu}$  on the left-hand side of (8.7) with the leading term in (8.15). From (8.15), (7.19), (8.16), and Lemma 7.7 we get

$$\begin{aligned} \mathbb{E} \prod_{r=1}^p |A_{\mathbf{v},i,\mu}(w_r, \sigma_r)| \left| (G_{\mu\mu})^{d_\mu} - \left( \sum_{k=0}^K \mathcal{Y}_{\mu,k} (X^* G^{(\mu)} X)_{\mu\mu}^k \right)^{d_\mu} \right| &\prec N^{-10} N^{2\delta d_\mu} \mathbb{E} \prod_{r=1}^p |A_{\mathbf{v},i,\mu}(w_r, \sigma_r)| \\ &\prec N^{-7} \left( (N^{(C_0/2+4)\delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X) \right), \end{aligned}$$

where in the last step we used  $d_\mu \leq p$ , Lemma 8.9 and that at most three  $r \in \llbracket 1, p \rrbracket$  satisfy  $n(w_r) \geq 1$ . We conclude that, to prove Lemma 8.8, it suffices to prove under the assumptions of Lemma 8.8 that

$$\begin{aligned} N^{-3/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \bar{\mathbf{v}}(\mu)^{d_\mu} \left| \mathbb{E} \left[ \prod_{r=1}^p A_{\mathbf{v},i,\mu}(w_r, \sigma_r) \left( \sum_{k=0}^K \mathcal{Y}_{\mu,k} (X^* G^{(\mu)} X)_{\mu\mu}^k \right)^{d_\mu} \right] \right| \\ = O \left( (N^{C_0\delta} \Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X) \right). \quad (8.17) \end{aligned}$$

Next, we expand

$$\left( \sum_{k=0}^K \mathcal{Y}_{\mu,k} (X^* G^{(\mu)} X)_{\mu\mu}^k \right)^{d_\mu} = \sum_{k=0}^{K d_\mu} \mathcal{Z}_{\mu,k} (X^* G^{(\mu)} X)_{\mu\mu}^k,$$

where the coefficients  $\mathcal{Z}_{\mu,k}$  are  $X^{(\mu)}$ -measurable and satisfy the bound

$$|\mathcal{Z}_{\mu,k}| \prec N^{C d_\mu \delta}. \quad (8.18)$$

Moreover, for any tagged word  $(w, \sigma)$  we split

$$A_{\mathbf{v}, i, \mu}(w, \sigma) = A_{\mathbf{v}, i, \mu}^0(w, \sigma) A_{\mathbf{v}, i, \mu}^+(w, \sigma), \quad (8.19)$$

where  $A_{\mathbf{v}, i, \mu}^0(w, \sigma)$  is equal to  $A_{\mathbf{v}, i, \mu}^0(w, 0)$  if  $n(w) = 0$  and contains all factors  $[G_{\mathbf{xy}}]^0$  from (8.5) if  $n(w) \geq 1$ . Hence,  $A_{\mathbf{v}, i, \mu}^0(w, \sigma)$  is  $X^{(\mu)}$ -measurable and  $A_{\mathbf{v}, i, \mu}^+(w, \sigma)$  is a product of terms of the form

$$(G^{(\mu)}X)_{\mathbf{x}\mu} \quad \text{or} \quad (X^*G^{(\mu)})_{\mu\mathbf{x}} \quad \text{where} \quad \mathbf{x} \in \{\bar{\mathbf{v}}, \mathbf{e}_i\}. \quad (8.20)$$

From (8.17), Lemma 7.7 and (8.18), we conclude that, to prove Lemma 8.8, it suffices to prove under the assumptions of Lemma 8.8 and for any nonnegative  $k \leq Cd_\mu$  that

$$\begin{aligned} N^{-3/2} N^{Cd_\mu\delta} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \bar{\mathbf{v}}(\mu)^{d_{\mathbf{v}}} \mathbb{E} \left( \left\| \prod_{r=1}^p A_{\mathbf{v}, i, \mu}^0(w_r, \sigma_r) \right\| \mathbb{E}_\mu \left( \prod_{r=1}^p A_{\mathbf{v}, i, \mu}^+(w_r, \sigma_r) \right) (X^*G^{(\mu)}X)_{\mu\mu}^k \right) \\ = O \left( (N^{C_0\delta}\Psi)^p + \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} F_{\mathbf{w}}^p(X) \right), \end{aligned} \quad (8.21)$$

where we recall the definition of the conditional expectation  $\mathbb{E}_\mu$  from (5.6). (Here the applicability of Lemma 7.7 may be checked using the estimates  $\|G_M\| \leq CN$  and  $\mathbb{E}|\mathcal{Z}_{\mu, k}|^8 \leq N^{C_p}$ . The latter estimate follows from (8.16).)

**8.4. Degree counting.** For the following we choose  $q, w_1, \dots, w_p$ , and  $\sigma_1, \dots, \sigma_r$  as in Lemma 8.8. We introduce the abbreviation

$$d_{\mathbf{X}} \equiv d_{\mathbf{X}}(w_1, \sigma_1, \dots, w_p, \sigma_p) := \sum_{r=1}^p \deg(A_{\mathbf{v}, i, \mu}(w_r, \sigma_r)) = \sum_{r=1}^p \deg(A_{\mathbf{v}, i, \mu}^+(w_r, \sigma_r)). \quad (8.22)$$

(Recall the definition of  $\deg(\cdot)$  from Definition 8.7 (iv).) By definition, the polynomial given by  $\prod_{r=1}^p A_{\mathbf{v}, i, \mu}^+(w_r, \sigma_r)$  is a product of  $d_{\mathbf{X}}$  factors from the list

$$(G^{(\mu)}X)_{i\mu}, \quad (X^*G^{(\mu)})_{\mu i}, \quad (G^{(\mu)}X)_{\bar{\mathbf{v}}\mu}, \quad (X^*G^{(\mu)})_{\mu\bar{\mathbf{v}}}. \quad (8.23)$$

LEMMA 8.10. *If  $d_{\mathbf{v}} = 0$  then  $d_{\mathbf{X}}$  is odd.*

PROOF. This may be checked directly using Definitions 8.6 and 8.7. The condition  $d_{\mathbf{v}} = 0$  means that we only take factors  $[\cdot]^0$  and  $[\cdot]^1$  in (8.5). The key observation is that the parity of  $\deg(A_{\mathbf{v}, i, \mu}(w_r, \sigma_r))$  for  $n \geq 1$  is the same as that of the number of indices  $\mu$  appearing on the right-hand side of (8.5). The claim then follows from the observation that in  $\prod_{r=1}^p A_{\mathbf{v}, i, \mu}(w_r)$ , which is a product of entries of  $G$ , the index  $\mu$  appears exactly three times.  $\square$

We may now estimate the conditional expectation  $\mathbb{E}_\mu(\cdot)$  in (8.21).

LEMMA 8.11. *Under the assumptions of Lemma 8.8 and for nonnegative  $k \leq Cd_\mu$  we have*

$$\left| \mathbb{E}_\mu \left( \prod_{r=1}^p A_{\mathbf{v}, i, \mu}^+(w_r, \sigma_r) \right) (X^*G^{(\mu)}X)_{\mu\mu}^k \right| \prec N^{-1(d_{\mathbf{v}}=0)/2} (N^{(C_0/2+C)\delta}\Psi)^{d_{\mathbf{X}}-1(d_{\mathbf{X}} \geq 3)\mathbf{1}(d_{\mathbf{v}}=0)}. \quad (8.24)$$

PROOF. To streamline notation, we abbreviate  $d := d_{\mathbf{x}}$ . Recalling the form of  $A^+(w_r, \sigma_r)$  from (8.20), we find that

$$\left( \prod_{r=1}^p A_{\mathbf{v}, i, \mu}^+(w_r, \sigma_r) \right) (X^* G^{(\mu)} X)_{\mu\mu}^k = \sum_{j_1, \dots, j_{d+2k} \in \mathcal{I}_N} \mathcal{G}_{j_1 \dots j_d} \tilde{\mathcal{G}}_{j_{d+1} \dots j_{d+2k}} \prod_{l=1}^{d+2k} X_{j_l \mu}, \quad (8.25)$$

where  $\mathcal{G}_{j_1 \dots j_d}$  is the product of  $d$  terms in  $\{G_{j_l \mathbf{x}}^{(\mu)}, G_{\mathbf{x} j_l}^{(\mu)} : l \in \llbracket 1, d \rrbracket, \mathbf{x} \in \{\bar{\mathbf{v}}, \mathbf{e}_i\}\}$  and  $\tilde{\mathcal{G}}_{j_{d+1} \dots j_{d+2k}}$  is the product of  $k$  terms in  $\{G_{j_l j_{l'}}^{(\mu)} : l, l' \in \llbracket d+1, d+k \rrbracket\}$ . In particular,  $\tilde{\mathcal{G}}_{j_{d+1} \dots j_{d+2k}}$  is  $H^{(\mu)}$ -measurable and satisfies the bound  $|\tilde{\mathcal{G}}_{j_{d+1} \dots j_{d+2k}}| \prec N^{2\delta k}$ . Since  $\mathbb{E}_\mu X_{j\mu} = 0$ , the contribution of the indices  $j_1, \dots, j_{d+2k}$  after taking the conditional expectation  $\mathbb{E}_\mu(\cdot)$  is nonzero only if each index appears at least twice. We classify the summation on the right-hand side of (8.25) according to the coincidences of the indices, which results in a sum over partitions of the set  $\{1, \dots, d+2k\}$  whose blocks have size at least two. We denote by  $L$  the number of blocks (i.e. independent summation indices). For a block  $b \subset \{1, \dots, d+2k\}$  we define  $d_b := |b \cap \llbracket 1, d \rrbracket|$  and  $k_b := |b \cap \llbracket d+1, d+2k \rrbracket|$ . The number  $d_b$  denotes the number of coinciding summation indices in the block  $b$  that originate from the first factor on the left-hand side of (8.25), and the number  $k_b$  the number of coinciding summation indices in the block  $b$  that originate from the second factor. Indexing the blocks as  $\llbracket 1, L \rrbracket$  (and hence writing  $d_l$  and  $k_l$  with  $l \in \llbracket 1, L \rrbracket$  instead of  $d_b$  and  $k_b$ ) and renaming the independent summation indices  $j_1, \dots, j_L$ , we therefore get

$$\begin{aligned} & \left| \mathbb{E}_\mu \left( \prod_{r=1}^p A_{\mathbf{v}, i, \mu}^+(w_r, \sigma_r) \right) (X^* G^{(\mu)} X)_{\mu\mu}^k \right| \\ & \leq C_p N^{2\delta k} \max_L \max_{\{d_l\}} \max_{\{k_l\}} \sum_{j_1, \dots, j_L} \prod_{l=1}^L \left( \mathbb{E} |X_{j_l \mu}|^{d_l + k_l} \left( |G_{\bar{\mathbf{v}} j_l}^{(\mu)}| + |G_{j_l \bar{\mathbf{v}}}^{(\mu)}| + |G_{j_l i}^{(\mu)}| + |G_{i j_l}^{(\mu)}| \right)^{d_l} \right), \quad (8.26) \end{aligned}$$

where the maxima are taken over  $L \in \mathbb{N}$  and  $d_l, k_l \geq 0$  satisfying

$$\sum_{l=1}^L d_l = d, \quad \sum_{l=1}^L k_l = 2k, \quad d_l + k_l \geq 2. \quad (8.27)$$

Here the constant  $C_p$  accounts for the immaterial constants depending on  $p$  arising from the combinatorics of all partitions. For  $L$ ,  $\{d_l\}$ , and  $\{k_l\}$  as above, we may estimate

$$\begin{aligned} & C_p N^{2\delta k} \sum_{j_1, \dots, j_L} \prod_{l=1}^L \left( \mathbb{E} |X_{j_l \mu}|^{d_l + k_l} \left( |G_{\bar{\mathbf{v}} j_l}^{(\mu)}| + |G_{j_l \bar{\mathbf{v}}}^{(\mu)}| + |G_{j_l i}^{(\mu)}| + |G_{i j_l}^{(\mu)}| \right)^{d_l} \right) \\ & \leq C_p N^{2\delta k} N^{-d/2-k} \prod_{l=1}^L \left( \sum_j \left( |G_{\bar{\mathbf{v}} j}^{(\mu)}| + |G_{j \bar{\mathbf{v}}}^{(\mu)}| + |G_{j i}^{(\mu)}| + |G_{i j}^{(\mu)}| \right)^{d_l} \right), \end{aligned}$$

where we used (2.5). Now we use the estimate

$$\sum_{j \in \mathcal{I}_M} |G_{\bar{\mathbf{w}} j}^{(\mu)}|^d \prec N (N^{(C_0/2+1)\delta} \Psi)^{d \wedge 2} N^{2\delta[d-2]_+}, \quad (8.28)$$

which follows from (8.11), (8.12), and (8.1). Hence we get

$$\left| \mathbb{E}_\mu \left( \prod_{r=1}^p A^+(w_r, \sigma_r) \right) (X^* G^{(\mu)} X)_{\mu\mu}^k \right| \prec \max_L \max_{\{d_l\}} \max_{\{k_l\}} N^{-d/2-k+L} N^{2\delta k} N^{2\delta \sum_l [d_l-2]_+} (N^{(C_0/2+1)\delta} \Psi)^{\sum_l (2 \wedge d_l)}, \quad (8.29)$$

where the maxima are subject to the same conditions as above.

Next, from  $\sum_l (d_l + k_l) = 2k + d$  and  $d_l + k_l \geq 2$  we deduce that

$$L = k + \frac{d}{2} - \frac{1}{2} \sum_l [d_l + k_l - 2]_+.$$

Together with  $\sum_l [d_l - 2]_+ \leq \sum_l d_l = d$ , this gives

$$\begin{aligned} N^{-d/2-k+L} N^{2\delta k} N^{2\delta \sum_l [d_l-2]_+} (N^{(C_0/2+1)\delta} \Psi)^{\sum_l (2 \wedge d_l)} \\ \leq N^{2\delta(d+k)} (N^{(C_0/2+1)\delta} \Psi)^{\sum_l (2 \wedge d_l)} N^{-\frac{1}{2} \sum_l [d_l + k_l - 2]_+}. \end{aligned} \quad (8.30)$$

Suppose first that  $d_{\mathbf{v}} = 0$ . From Lemma 8.10 and (8.27) we find that  $\sum_l [d_l + k_l - 2]_+ \geq 1$ . Using  $\Psi \geq N^{-1/2}$ , we therefore get

$$\begin{aligned} N^{-d/2-k+L} N^{2\delta k} N^{2\delta \sum_l [d_l-2]_+} (N^{(C_0/2+1)\delta} \Psi)^{\sum_l (2 \wedge d_l)} \\ \leq N^{2\delta(d+k)} N^{-1/2} (N^{(C_0/2+1)\delta} \Psi)^{\sum_l (2 \wedge d_l) + \sum_l [d_l + k_l - 2]_+ - 1} \\ \prec N^{2\delta(d+k)} N^{-1/2} (N^{(C_0/2+1)\delta} \Psi)^{d-1(d \geq 3)} \end{aligned}$$

where in the last step we used that  $[d_l + k_l - 2]_+ \geq [d_l - 2]_+$ , and  $\sum_l (2 \wedge d_l) + \sum_l [d_l - 2]_+ = d$ . For the case  $d = 1$ , we also used that  $\sum_l (2 \wedge d_l) + \sum_l [d_l + k_l - 2]_+ - 1 \geq 1$ .

Next, we claim that

$$d_\mu \leq d + 6. \quad (8.31)$$

This follows from  $|\sigma|_\mu \leq \deg(A_{\mathbf{v}, i, \mu}(w, \sigma)) + 2 \mathbf{1}(n(w) \geq 1)$ , which may be checked using Definition 8.7. Since  $k \leq C d_\mu$ , we deduce that  $k \leq C d$ , and therefore conclude the proof of (8.24) for the case  $d_{\mathbf{v}} = 0$ .

Finally, if  $d_{\mathbf{v}} \geq 1$  then we only have the trivial lower bound  $\sum_l [d_l + k_l - 2]_+ \geq 0$ , from which (8.24) may be obtained by a minor modification of the above argument.  $\square$

We conclude that the left-hand side of (8.21) is bounded by

$$\begin{aligned} N^{-3/2-1(d_{\mathbf{v}}=0)/2} \sum_{\mu \in \mathcal{I}_N} |\bar{\mathbf{v}}(\mu)|^{d_{\mathbf{v}}} \sum_{i \in \mathcal{I}_M} \mathbb{E} \left| \prod_{r=1}^p A_{\mathbf{v}, i, \mu}^0(w_r, \sigma_r) \right| (N^{(C_0/2+C)\delta} \Psi)^{d_{\mathbf{X}}-1(d_{\mathbf{X}} \geq 3) \mathbf{1}(d_{\mathbf{v}}=0)} \\ \leq \max_{\mu \in \mathcal{I}_N} N^{-1} \sum_{i \in \mathcal{I}_M} \mathbb{E} \left| \prod_{r=1}^p A_{\mathbf{v}, i, \mu}^0(w_r, \sigma_r) \right| (N^{(C_0/2+C)\delta} \Psi)^{d_{\mathbf{X}}-1(d_{\mathbf{X}} \geq 3) \mathbf{1}(d_{\mathbf{v}}=0) + \mathbf{1}(d_{\mathbf{v}} \geq 2)}, \end{aligned} \quad (8.32)$$

where we used (8.31) with  $d = d_{\mathbf{X}}$  and  $\Psi \geq N^{-1/2}$ .



Next, we define  $f_{\mathbf{i}}$  to be the number of factors of the form  $[G_{\mathbf{v}i}]^0$  or  $[G_{i\mathbf{v}}]^0$  in the polynomial  $\prod_{r=1}^q A_{\mathbf{v},i,\mu}(w_r, \sigma_r)$  (recall Definition 8.7 (ii)). (Here we use the letter  $f$  instead of  $d$  to emphasize that the definition of  $f_{\mathbf{i}}$  is not analogous to that of  $d_{\mu}$ .) Then we get from (8.9), (8.11), and (8.12) that

$$N^{-1} \sum_{i \in \mathcal{I}_M} \left| \prod_{r=1}^q A_{\mathbf{v},i,\mu}^0(w_r, \sigma_r) \right| \prec N^{C\delta} N^{-1} \sum_{i \in \mathcal{I}_M} (|G_{i\mathbf{v}}^{(\mu)}|^{f_{\mathbf{i}} \wedge 2} + |G_{\mathbf{v}i}^{(\mu)}|^{f_{\mathbf{i}} \wedge 2}) \prec N^{C\delta} (N^{(C_0/2+1)\delta} \Psi)^{f_{\mathbf{i}} \wedge 2}. \quad (8.33)$$

Moreover, by (8.10) we have

$$\left| \prod_{r=q+1}^p A_{\mathbf{v},i,\mu}^0(w_r, \sigma_r) \right| \prec \left( N^{(C_0/2+1)\delta} \Psi + F_{\mathbf{v}}(X) + F_{\mathbf{e}_{\mu}}(X) \right)^{p-q-\sum_{r=q+1}^p |\sigma_r|_{\mu}} \quad (8.34)$$

Plugging (8.33) and (8.34) into (8.32), and recalling Lemma 7.7, yields

$$\begin{aligned} \text{LHS of (8.21)} &\leq (N^{(C_0/2+C)\delta} \Psi)^{d_{\mathbf{X}} - 1(d_{\mathbf{X}} \geq 3)1(d_{\mathbf{v}}=0) + 1(d_{\mathbf{v}} \geq 2) + f_{\mathbf{i}} \wedge 2} \\ &\quad \times \sup_{\mathbf{w} \in \mathbb{S}} \mathbb{E} \left( N^{(C_0/2+1)\delta} \Psi + F_{\mathbf{w}}(X) \right)^{p-q-\sum_{r=q+1}^p |\sigma_r|_{\mu}}, \end{aligned} \quad (8.35)$$

where we absorbed the factor  $N^{C\delta}$  on the right-hand side of (8.33) into the parentheses on the first line using that  $d_{\mathbf{X}} + f_{\mathbf{i}} > 0$ , as follows from Definition 8.7 and the assumptions of Lemma 8.8.

Recalling that  $N^{(C_0/2+C)\delta} \Psi + N^{-3\delta} F_{\mathbf{w}}(X) \prec N^{-\delta}$  for small enough  $\delta$  by (8.1), we find using Young's inequality that in order to prove (8.21), it suffices to prove

$$d_{\mathbf{X}} - 1(d_{\mathbf{X}} \geq 3)1(d_{\mathbf{v}}=0) + 1(d_{\mathbf{v}} \geq 2) + f_{\mathbf{i}} \wedge 2 - q - \sum_{r=q+1}^p |\sigma_r|_{\mu} \geq 0. \quad (8.36)$$

**8.5. Power counting: proof of (8.36).** The proof of (8.36) requires precise information about the structure of the factors of  $A_{\mathbf{v},i,\mu}(w_r, \sigma_r)$ . To that end, we introduce the quantities

$$d_{\mathbf{X}}^{\leq} := \sum_{r=1}^q \deg(A_{\mathbf{v},i,\mu}(w_r, \sigma_r)), \quad d_{\mathbf{X}}^{\geq} := \sum_{r=q+1}^p \deg(A_{\mathbf{v},i,\mu}(w_r, \sigma_r))$$

and

$$d_*^{\leq} := \sum_{r=1}^q |\sigma_r|_*, \quad d_*^{\geq} := \sum_{r=q+1}^p |\sigma_r|_*$$

for  $* = \mu, \mathbf{v}$ . Hence,  $d_* = d_*^{\leq} + d_*^{\geq}$  for  $* = \mu, \mathbf{v}, \mathbf{X}$ . (Recall the definitions (8.8) and (8.22).) Moreover,

$$2d_{\mu}^{\geq} = d_{\mathbf{X}}^{\geq} + d_{\mathbf{v}}^{\geq}, \quad d_{\mu}^{\geq} \geq d_{\mathbf{v}}^{\geq}, \quad (8.37)$$

as may be easily checked from Definition 8.7. (Recall that  $n(w_r) = 0$  for  $r \geq q+1$ .) We conclude that (8.36) holds provided we can show that

$$d_{\mathbf{X}}^{\leq} + f_{\mathbf{i}} \wedge 2 - q - 1(d_{\mathbf{X}} \geq 3)1(d_{\mathbf{v}}=0) + 1(d_{\mathbf{v}} \geq 2) \geq 0. \quad (8.38)$$

The following arguments rely on the algebraic structure of  $A_{\mathbf{v},i,\mu}(w_r)$  and  $A_{\mathbf{v},i,\mu}(w_r, \sigma_r)$  from Definitions 7.19 and 8.7 respectively, to which we refer tacitly throughout the rest of the proof.

We claim that

$$d_{\mathbf{X}}^{\leq} + f_{\mathbf{i}} \geq 1. \quad (8.39)$$

To see this, we note that each factor  $[G_{\bar{\mathbf{v}}i}]^1$  contributes one to  $d_{\mathbf{X}}^{\leq}$  and each factor  $[G_{\bar{\mathbf{v}}i}]^0$  contributes one to  $f_{\mathbf{i}}$ ; the same holds for  $[G_{i\bar{\mathbf{v}}}]^*$  (where  $*$  = 0, 1). Since there are exactly three indices  $i$  in the factors of  $\prod_{r=1}^q A_{\mathbf{v},i,\mu}(w_r)$ , we find that  $\prod_{r=1}^q A_{\mathbf{v},i,\mu}(w_r)$  must contain at least one of the factors  $G_{\bar{\mathbf{v}}i}$ ,  $G_{i\bar{\mathbf{v}}}$ ,  $G_{i\mu}$ , or  $G_{\mu i}$ . We therefore conclude that, no matter the choice of  $\sigma_1, \dots, \sigma_r$ , we always have (8.39).

Furthermore, if  $q = 3$  then there are exactly three factors  $G_{\bar{\mathbf{v}}i}$  or  $G_{i\bar{\mathbf{v}}}$  in  $\prod_{r=1}^q A_{\mathbf{v},i,\mu}(w_r)$ . We deduce that (8.39) may be improved to

$$d_{\mathbf{X}}^{\leq} + f_{\mathbf{i}} \wedge 2 \geq 1 + \mathbf{1}(q = 3).$$

We conclude that (8.38) holds provided that (i)  $d_{\mathbf{v}} \geq 2$  or (ii)  $d_{\mathbf{v}} = 1$  and  $q = 1$ .

Next, as noted above, each factor  $G_{\bar{\mathbf{v}}i}$  or  $G_{i\bar{\mathbf{v}}}$  of  $\prod_{r=1}^q A_{\mathbf{v},i,\mu}(w_r)$  contributes one to  $d_{\mathbf{X}}^{\leq} + f_{\mathbf{i}}$ . Moreover, each factor  $G_{\bar{\mathbf{v}}\mu}$  or  $G_{\mu\bar{\mathbf{v}}}$  contributes one to  $d_{\mathbf{X}}^{\leq} + d_{\mathbf{v}}^{\leq}$ . Since the number of factors  $G_{\bar{\mathbf{v}}i}$ ,  $G_{i\bar{\mathbf{v}}}$ ,  $G_{\bar{\mathbf{v}}\mu}$  or  $G_{\mu\bar{\mathbf{v}}}$  in  $\prod_{r=1}^q A_{\mathbf{v},i,\mu}(w_r)$  is exactly  $2q$ , we find

$$d_{\mathbf{X}}^{\leq} + f_{\mathbf{i}} \wedge 2 + \mathbf{1}(f_{\mathbf{i}} = 3) + d_{\mathbf{v}}^{\leq} \geq 2q,$$

where we used that  $f_{\mathbf{i}} \leq 3$ . We conclude that (8.38) holds provided that

$$q - d_{\mathbf{v}}^{\leq} - \mathbf{1}(f_{\mathbf{i}} = 3) - \mathbf{1}(d_{\mathbf{X}} \geq 3)\mathbf{1}(d_{\mathbf{v}} = 0) \geq 0, \quad (8.40)$$

which is easy to check in the case  $d_{\mathbf{v}} \leq 1$  and  $q \geq 2$ .

All that remain therefore is the case  $d_{\mathbf{v}} = 0$  and  $q = 1$ . In that case we have  $f_{\mathbf{i}} \leq 2$  and  $d_{\mathbf{v}}^{\leq} = 0$ , so that (8.40) holds also in this case. This concludes the proof of (8.36).

The proof of (8.21), and hence of Lemma 8.8, is therefore complete. This concludes the proof of Lemma 8.4.

We have hence proved Proposition 8.1. Recalling Proposition 6.1, we conclude the proof of Theorem 3.21 (i).

## 9. Averaged local law

In this section we complete the proof of Theorem 3.21 by proving its part (ii). Bearing applications to eigenvalue rigidity in Section 10 below in mind, we in fact prove a somewhat stronger result. Recall the definition of  $m_N$  from (3.9). We generalize Definition 3.20 (iii) by saying that the *averaged local law holds with parameters*  $(X, \Sigma, \mathbf{S}, \Phi)$  if  $|m_N(z) - m(z)| \prec \Phi(z)$  uniformly in  $z \in \mathbf{S}$ . Here  $\Phi : \mathbf{S} \rightarrow (0, \infty)$  is a deterministic control parameter satisfying

$$c\Psi^2 \leq \Phi \leq N^{-\tau/2}. \quad (9.1)$$

for some constant  $c > 0$ . In particular, we recover Definition 3.20 (iii) by setting  $\Phi = (N\eta)^{-1}$ . In this section we prove the following result, which also completes the proof of Theorem 3.21.

PROPOSITION 9.1. *Suppose that the assumptions of Theorem 3.21 hold. Let  $\Phi$  satisfy (9.1). Suppose that the entrywise local law holds with parameters  $(X^{\text{Gauss}}, D, \mathbf{S})$  and that the averaged local law holds with parameters  $(X^{\text{Gauss}}, D, \mathbf{S}, \Phi)$ . Then the averaged local law holds with parameters  $(X, \Sigma, \mathbf{S}, \Phi)$ .*

PROOF. The proof is similar to that of Propositions 7.1 and 8.1, and we only explain the differences. Note that now there is no bootstrapping, since the necessary a priori bounds are obtained from the anisotropic local law. In analogy to (7.15), we define

$$\tilde{F}(X, z) := |m_N(z) - m(z)| = \left| \frac{1}{N} \sum_{\nu \in \mathcal{I}_N} G_{\nu\nu}(z) - m(z) \right|.$$

Following the argument leading up to (7.30) and Lemma 8.3 to the letter, we find that it suffices to prove that

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \left( \frac{\partial}{\partial X_{i\mu}} \right)^n \tilde{F}^p(X) \right| = O\left((N^\delta \Psi^2)^p + \mathbb{E} \tilde{F}^p(X)\right)$$

for any  $n = 3, \dots, 4p$  and  $z \in \mathbf{S}$ . Here  $\delta > 0$  is an arbitrary positive constant. We use the words  $w$  from Definition 7.19. Analogously to (7.34) and Lemma 8.4, it suffices to prove, for all  $n = 3, \dots, 4p$ , that

$$N^{-n/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \mathbb{E} \prod_{r=1}^p \left( \frac{1}{N} \sum_{\nu \in \mathcal{I}_N} A_{\mathbf{e}_\nu, i, \mu}(w_r) \right) \right| = O\left((N^\delta \Psi^2)^p + \mathbb{E} \tilde{F}^p(X)\right) \quad \text{for } \sum_r n(w_r) = n. \quad (9.2)$$

Next, from the part (i) of Theorem 3.21 we get  $|G_{st} - \Pi_{st}| \prec \Psi$  for  $s, t \in \mathcal{I}$  and  $z \in \mathbf{S}$ . From (7.32) we deduce that

$$\left| \frac{1}{N} \sum_{\nu \in \mathcal{I}_N} A_{\mathbf{e}_\nu, i, \mu}(w) \right| \prec \Psi^2 \quad \text{for } n(w) \geq 1. \quad (9.3)$$

For  $n \geq 4$ , the claim (9.2) easily follows from (9.3) and an application of Young's inequality to the factors  $r$  satisfying  $n(w_r) = 0$  and  $n(w_r) \geq 1$ ; we omit the details. What remains, therefore, is to verify (9.2) for  $n = 3$ .

Let  $n = 3$ . As in Lemma 8.5, it is easy to check that it suffices to prove (9.2) with  $A_{\mathbf{e}_\nu, i, \mu}(w_r)$  replaced by  $\hat{A}_{\mathbf{e}_\nu, i, \mu}(w_r)$ . This means that we replace the sum  $\sum_{\nu \in \mathcal{I}_N}$  with  $\sum_{\nu \in \mathcal{I}_N \setminus \{\mu\}}$ . From (8.6) and Definition 8.7, we conclude that it suffices to prove that

$$N^{-3/2} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \left| \frac{1}{N^p} \sum_{\nu_1, \dots, \nu_p \in \mathcal{I}_M \setminus \{\mu\}} \mathbb{E} \prod_{r=1}^p A_{\mathbf{e}_{\nu_r}, i, \mu}(w_r, \sigma_r) (G_{\mu\mu})^{d_\mu} \right| = O\left((N^\delta \Psi^2)^p + \mathbb{E} \tilde{F}^p(X)\right),$$

where  $\sum_r n(w_r) = 3$ ,  $n(w_r) \geq 1$  for  $r \in \llbracket 1, q \rrbracket$  and  $n(w_r) = 0$  for  $r \geq q+1$ , and  $\sigma_r \in \mathbb{Z}_2^{n(w_r)+1}$ . Here we recall the index  $d_\mu$  defined in Lemma 8.8. Note that, since  $\nu_r \neq \mu$ , in (8.4) we have  $\bar{\mathbf{v}}(\mu) = \delta_{\mu\nu_r} = 0$ , so that we only need to consider  $\sigma_r \in \mathbb{Z}_2^{n(w_r)+1}$  instead of  $\sigma_r \in \mathbb{Z}_3^{n(w_r)+1}$ . In

other words, we always have  $d_{\mathbf{v}} = 0$ . As in (8.21), we find that it suffices to prove, under the same assumptions and for any  $k \leq Cd_{\mu}$ ,

$$N^{-3/2} N^{d_{\mu}\delta} \sum_{i \in \mathcal{I}_M} \sum_{\mu \in \mathcal{I}_N} \mathbb{E} \left| \frac{1}{N^p} \sum_{\nu_1, \dots, \nu_p \in \mathcal{I}_M \setminus \{\mu\}} \left( \prod_{r=1}^p A^0(r) \right) \mathbb{E}_{\mu} \left( \prod_{r=1}^p A^+(r) \right) (X^* G^{(\mu)} X)_{\mu\mu}^k \right| \\ = O\left((N^{\delta} \Psi^2)^p + \mathbb{E} \tilde{F}^p(X)\right), \quad (9.4)$$

where we abbreviated  $A^*(r) \equiv A_{\mathbf{e}_{\nu_r, i, \mu}}^*(w_r, \sigma_r)$  for  $* = 0, +$ .

Next, note that each factor  $A^+(r)$  is a product of factors  $(G^{(\mu)} X)_{s\mu}$  and  $(X^* G^{(\mu)})_{\mu s}$  where  $s \in \{i, \nu_r\}$ . We denote by  $d_{\mathbf{X}, i, r}$  the number of factors of  $A^+(r)$  for which  $s = i$ , and abbreviate  $d_{\mathbf{X}, i} := \sum_{r=1}^p d_{\mathbf{X}, i, r}$ . It is not hard to check that  $d_{\mathbf{X}, i} \leq 3$ . Recall the definition of  $d_{\mathbf{X}}$  from (8.22). We claim that

$$\left| \mathbb{E}_{\mu} \left( \prod_{r=1}^p A^+(r) \right) (X^* G^{(\mu)} X)_{\mu\mu}^k \right| \prec N^{-1/2} \Psi^{d_{\mathbf{X}} - 1(d_{\mathbf{X}, i} = 3)}, \quad (9.5)$$

which may be regarded as an improvement of Lemma 8.11. The proof of (9.5) is similar to that of Lemma 8.11, and we merely give a sketch. We have to estimate an expression of the form

$$\mathbb{E}_{\mu} \left( \prod_{l=1}^{d_{\mathbf{X}}} (G^{(\mu)} X)_{s_l \mu}^{b_l} (X^* G^{(\mu)})_{\mu s_l}^{1-b_l} \right) (X^* G^{(\mu)} X)_{\mu\mu}^k,$$

where  $s_l \in \{i, \nu_1, \dots, \nu_p\}$  and  $b_l \in \{0, 1\}$ . (This is simply the general form of a polynomial in  $X$  of the correct degree.) The proof of (9.5) relies on the crucial observations that, by Theorem 3.21 (i),

$$|G_{\nu_j}^{(\mu)}| + |G_{j\nu}^{(\mu)}| \prec \Psi \quad \text{and} \quad \frac{1}{N} \sum_{j \in \mathcal{I}_M} (|G_{ij}^{(\mu)}|^a + |G_{ji}^{(\mu)}|^a) \prec \Psi^{a \wedge 2}$$

for  $a \in \mathbb{N}$ . Using that  $d_{\mathbf{X}}$  is odd by Lemma 8.10 (recall that  $d_{\mathbf{v}} = 0$ ), it is not hard to conclude (9.5). (Note that, thanks to the a priori information provided by Theorem 3.21 (i), the estimate (9.5), including the exponent on the right-hand side, is sharper than Lemma 8.11. This improvement will prove crucial for the conclusion of the argument.)

Next, let  $r \geq q+1$  satisfy  $\sigma_r = 0$ , so that the left-hand side of (9.5) does not depend on  $\nu_r$ . There are  $p - q - \sum_{s=q+1}^p |\sigma_s|$  such indices  $r$ , and for each such  $r$  we find

$$\frac{1}{N} \sum_{\nu_r \in \mathcal{I}_N \setminus \{\mu\}} A^0(r) = \tilde{F}(X) + O_{\prec}(\Psi^2). \quad (9.6)$$

Here we wrote  $A^0(r) = G_{\nu_r \nu_r} - \Pi_{\nu_r \nu_r} - A_{\mathbf{e}_{\nu_r, i, \mu}}(w_r, 1)$  and estimated the term  $A_{\mathbf{e}_{\nu_r, i, \mu}}(w_r, 1)$  using (8.11) and (8.12). For the remaining  $\sum_{s=q+1}^p |\sigma_s|$  indices  $r \geq q+1$  we have  $A^0(r) = 1$ . Moreover, for  $r \leq q$  it is easy to verify directly using Definition 8.7 that

$$|A^0(r)| \prec \Psi^{(2-d_{\mathbf{X}, r})_+}, \quad (9.7)$$

where we abbreviated  $d_{\mathbf{X}, r} := \deg(A^+(r))$ . From (9.5), (9.6), and (9.7) we find that the left-hand side of (9.4) is bounded by

$$N^{d_{\mu}\delta} \Psi^{d_{\mathbf{X}} - 1(d_{\mathbf{X}, i} = 3)} \Psi^{\sum_{r=1}^q (2-d_{\mathbf{X}, r})_+} \mathbb{E}(\tilde{F}(X) + \Psi^2)^{p-q-\sum_{r=q+1}^p |\sigma_r|}.$$

As in the argument following Lemma 8.11, we conclude that the claim holds provided that

$$d_{\mathbf{X}} - \mathbf{1}(d_{\mathbf{X},i} = 3) + \sum_{r=1}^q (2 - d_{\mathbf{X},r})_+ - 2q - 2 \sum_{r=q+1}^p |\sigma_r| \geq 0. \quad (9.8)$$

In order to establish (9.8), we make the following observations about  $d_{\mathbf{X},i,r}$ , which may be checked case by case directly from Definition 8.7. First, if  $n(w_r) = 0$  then  $d_{\mathbf{X},i,r} = 0$ . Second, if  $n(w_r) \in \llbracket 1, 3 \rrbracket$  we have the implication

$$d_{\mathbf{X},r} \leq 2 \implies d_{\mathbf{X},i,r} \leq \mathbf{1}(n(w_r) \geq 2).$$

We conclude that if  $d_{\mathbf{X},i} = 3$  then there exists an  $r \leq q$  such that  $d_{\mathbf{X},r} \geq 3$ . Hence, to establish (9.8) it is enough to establish

$$d_{\mathbf{X}} + \sum_{r=1}^q (2 - d_{\mathbf{X},r}) - 2q - 2 \sum_{r=q+1}^p |\sigma_r| \geq 0,$$

which is trivial by the identity

$$d_{\mathbf{X}} = \sum_{r=1}^p d_{\mathbf{X},r} = \sum_{r=1}^q d_{\mathbf{X},r} + 2 \sum_{r=q+1}^p |\sigma_r|.$$

This concludes the proof.  $\square$

This concludes the proof of Theorem 3.21. We conclude this section by drawing consequences from Theorems 3.21 and 3.22.

PROOF OF THEOREMS 3.14 (i), 3.15 (i), AND 3.16 (i). From Lemma A.4 we find that (A.7) and (A.8) hold on  $\mathbf{D}_k^e$ . From (A.8), after a possible decrease of  $\tau$ , we find that (3.20) holds. Moreover, from Lemma A.5 we find that (2.11) is strongly stable on  $\mathbf{D}_k^e$  in the sense of Definition A.2. Using (A.7), we deduce that (2.11) is stable on  $\mathbf{D}_k^e$  in the sense of Definition 5.4. We may therefore invoke Theorems 3.21 and 3.22 to conclude Theorem 3.14 (i).

Similarly, Theorems 3.15 (i) and 3.16 (i) follow from Lemmas A.6 and A.8 respectively, combined with Theorems 3.21 and 3.22  $\square$

PROOF OF THEOREM 3.6. Let  $\tau' > 0$  be chosen so that the conclusions of Theorem 3.14 (i) hold for  $k = 1, \dots, 2p$ . Since  $\mathbf{D} = \bigcup_{k=1}^{2p} \mathbf{D}_k^e \cup \bigcup_{k=1}^p \mathbf{D}_k^b \cup \mathbf{D}^o$ , we obtain Theorem 3.6 from Theorems 3.14 (i), 3.15 (i), and 3.16 (i).  $\square$

PROOF OF COROLLARY 3.9 AND THEOREMS 3.14 (ii), 3.15 (ii), AND 3.16 (ii) ASSUMING (2.9). These follow immediately from Theorem 3.6 and Theorems 3.14 (i), 3.15 (i), and 3.16 (i) respectively, combined with (3.3), (3.4), and (3.5).  $\square$

How to remove the assumption (2.9) in the above proof is explained in Section 11.1.

## 10. Eigenvalue rigidity and edge universality

In the first part of this section we establish eigenvalue rigidity (Theorems 3.12, 3.14 (iii), and 3.15 (iii)). As a consequence of Theorem 3.6 we prove Theorem 3.7. In the second part of this section we establish the edge universality from Theorem 3.18. We assume (2.9) throughout this section; how to remove it is explained in Section 11.1 below.

**10.1. Eigenvalue rigidity assuming (2.9).** For simplicity, we first prove Theorem 3.12, i.e. we assume that all edges and bulk components are regular. The proof consists of three steps. First, we prove that with high probability there are no eigenvalues at a distance greater than  $N^{-2/3+\varepsilon}$  from the support of  $\varrho$ . Second, we prove that a neighbourhood of the  $k$ -th bulk component of  $\varrho$  contains with high probability exactly  $N_k$  eigenvalues (recall the definition (3.16)). Third, we use the averaged local law from Theorem 3.6 together with the first two steps to complete the proof.

We begin with the first step. We define the *distance to the spectral edge* through

$$\kappa \equiv \kappa(E) := \text{dist}(E, \partial \text{supp } \varrho). \quad (10.1)$$

LEMMA 10.1. *Fix  $\varepsilon \in (0, 2/3)$ . Under (2.9) and the assumptions of Theorem 3.12 we have*

$$\text{spec}(Q) \cap \{E \geq \tau : \text{dist}(E, \text{supp } \varrho) \geq N^{-2/3+\varepsilon}\} = \emptyset$$

*with high probability (recall Definition 4.7).*

PROOF. The argument is similar to that of the previous works [20, 32], and we only explain how to adapt it. From [7, Theorem 2.10], we find that  $\|X^*X\| \leq C_0$  with high probability for some large enough constant  $C_0 > 0$  depending on  $\tau$ . From (2.7) we therefore deduce that  $\lambda_1 \leq C$  with high probability for some constant  $C > 0$  depending on  $\tau$ . It therefore suffices to prove that, after a possible decrease of  $\tau$ ,

$$\text{spec}(Q) \cap \{E \in [\tau, \tau^{-1}] : \text{dist}(E, \text{supp } \varrho) \geq N^{-2/3+\varepsilon}\} = \emptyset \quad (10.2)$$

with high probability. For the remainder of the proof we use a spectral parameter  $z = E + i\eta$  where  $E \in [\tau, \tau^{-1}]$  satisfies  $\text{dist}(E, \text{supp } \varrho) \geq N^{-2/3+\varepsilon}$ , and  $\eta := N^{-1/2-\varepsilon/4}\kappa^{1/4}$ . We omit  $z$  from our notation.

In order to prove (10.2), it suffices to prove that  $\text{Im } m_N \prec N^{-\varepsilon/2}(N\eta)^{-1}$ . By (A.7) and Lemmas A.4, A.6, and A.8, we have

$$\text{Im } m \asymp \eta(\kappa + \eta)^{-1/2}. \quad (10.3)$$

We conclude that it suffices to prove

$$|m_N - m| \prec (\kappa + \eta)^{-1/2} \Psi^2. \quad (10.4)$$

It is easy to check that the control parameter on the right-hand side of (10.4) satisfies (9.1), so that by Proposition 9.1 and Theorem 3.6 it suffices to prove (10.4) for diagonal  $\Sigma$ . The proof is identical to that of Section 5, except that we use the strong stability of (2.11) from Definition A.2, which is established in Lemmas A.5, A.6, and A.8. This concludes the proof of (10.4), and hence of (10.2).

With applications to the deformed Wigner matrices in Section 12 in mind, we give another proof of (10.2), which is based on [1, Section 6]. Using a partial fraction decomposition, one easily finds that there exist universal constants  $C_{n,k}$  such that

$$\frac{1}{N} \sum_i \prod_{k=1}^n \frac{1}{(\lambda_i - E)^2 + k\eta^2} = \sum_{k=1}^n C_{n,k} \eta^{-2n+1} \operatorname{Im} m_N(z_k), \quad z_k := E + i\eta_k, \quad \eta_k := \sqrt{k}\eta. \quad (10.5)$$

Similarly, we have (with the same  $C_{n,k}$  and  $z_k$ )

$$\int \prod_{k=1}^n \frac{1}{(x - E)^2 + k\eta^2} \varrho(dx) = \sum_{k=1}^n C_{n,k} \eta^{-2n+1} \operatorname{Im} m(z_k). \quad (10.6)$$

Let  $E \geq \gamma_+ + N^{-2/3+\varepsilon}$  and set  $\eta := N^{-\varepsilon}\kappa$ . It is not hard to deduce from (10.4) that, for any fixed  $n \in \mathbb{N}$ , we have

$$|\operatorname{Im} m_N(z_k) - \operatorname{Im} m(z_k)| \prec \frac{N^{-\varepsilon/2}}{N\eta} \quad (10.7)$$

for all  $k = 1, \dots, n$ . Setting  $n := \lceil 2/\varepsilon \rceil$ , we get from (10.5) and (10.6) that

$$\begin{aligned} \frac{1}{N} \sum_i \prod_{k=1}^n \frac{1}{(\lambda_i - E)^2 + k\eta^2} &= \int \prod_{k=1}^n \frac{1}{(x - E)^2 + k\eta^2} \varrho(dx) + O_{\prec}(N^{-1-\varepsilon/2}\eta^{-2n}) \\ &= O_{\prec}(N^{-1-\varepsilon/2}\eta^{-2n}), \end{aligned}$$

where in the last step we used that  $|x - E| \geq \kappa$  for  $x \in \operatorname{supp} \varrho$ . This immediately implies that with high probability there is no eigenvalue in  $[E - \eta, E + \eta]$ .  $\square$

The second step represents most of the work. It is a counting argument, based on a continuous deformation of the matrix  $Q$  to another matrix for which the claim is obvious. Since the eigenvalues depend continuously on the deformation parameter and each intermediate matrix satisfies a gap condition from Lemma 10.1, we shall be able to conclude that the number of eigenvalues in a neighbourhood of the  $i$ -th component does not change under the deformation. We shall in fact need two deformations: one which deforms the original matrix  $Q$  to a Gaussian one,  $Q^{\text{Gauss}}$ , with the same expectation  $\Sigma$  as  $Q$ , and another which deforms the Gaussian matrix  $Q^{\text{Gauss}}$  to another Gaussian matrix where some eigenvalues of  $\Sigma$  have been increased.

For  $k = 1, \dots, p-1$ , we introduce the number of eigenvalues to the right of the  $k$ -th gap,

$$\Upsilon_k := \sum_{i=1}^{M \wedge N} \mathbf{1} \left( \lambda_i \geq \frac{a_{2k} + a_{2k+1}}{2} \right).$$

**PROPOSITION 10.2.** *Under (2.9) and the assumptions of Theorem 3.12 we have  $\Upsilon_k = \sum_{l \leq k} N_l$  with high probability.*

As explained above, the first step in the proof of Proposition 10.2 is a deformation of the general matrix  $X$  to a Gaussian one.

**LEMMA 10.3.** *Let  $X$  be general and  $X^{\text{Gauss}}$  a Gaussian matrix of the same dimensions. Suppose that (2.9) and the assumptions of Theorem 3.12 hold. If  $\Upsilon_k = \sum_{l \leq k} N_l$  with high probability under the law of  $X^{\text{Gauss}}$ , then  $\Upsilon_k = \sum_{l \leq k} N_l$  with high probability under the law of  $X$ .*

PROOF. Let  $X_1 := X$  and  $X_0 := X^{\text{Gauss}}$  be independent. For  $t \in [0, 1]$  define  $X(t) := \sqrt{t}X_1 + \sqrt{1-t}X_0$  and denote by  $\lambda_i(t)$  the eigenvalues of  $X(t)\Sigma X(t)^*$ . We also write  $\Upsilon_k(t)$  for  $\Upsilon_k$  defined in terms of  $\lambda_i(t)$ . Note that Lemmas 4.8 and 10.1 hold for  $X(t)$  uniformly in  $t \in [0, 1]$ . Recalling (2.7), we deduce that there exists a constant  $C > 0$  such that

$$|\lambda_i(t) - \lambda_i(s)| \leq C\sqrt{|t-s|} \quad (10.8)$$

with high probability, uniformly in  $t, s \in [0, 1]$  and  $i$ . The claim now follows by considering  $t$  in the lattice  $0, 1/K, 2/K, \dots, 1$ , where  $K$  is chosen large enough that  $CK^{-1/2} \leq (a_{2k} - a_{2k+1})/10$  and  $C$  is the constant from (10.8). Indeed, suppose  $\Upsilon_k(t) = \sum_{l \leq k} N_l$  with high probability. Then we use (10.8) and the gap from Lemma 10.1 to deduce that  $\Upsilon_k(t + K^{-1}) = \sum_{l \leq k} N_l$  with high probability. The claim for  $t = 1 = K/K$  therefore follows by induction.  $\square$

Using Lemma 10.3, in order to prove Proposition 10.2 it suffices to prove the following result.

LEMMA 10.4. *Suppose that (2.9) and the assumptions of Theorem 3.12 hold, that  $X$  is Gaussian, and that  $\Sigma$  is diagonal. Then  $\Upsilon_k = \sum_{l \leq k} N_l$  with high probability.*

PROOF. Abbreviate  $d_k := \sum_{l \leq k} N_l$ . We write the diagonal matrix  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_M)$  in the block form  $\Sigma = \text{diag}(\Sigma_1, \Sigma_2)$ , where  $\Sigma_1$  contains the  $d_k$  top eigenvalues of  $\Sigma$ . We introduce the deformed covariance matrix  $\Sigma(t) := \text{diag}(t\Sigma_1, \Sigma_2)$ . In particular,  $\Sigma(1) = \Sigma$ . The idea is to increase  $t$  until the claim for  $\Sigma$  replaced by  $\Sigma(t)$  may be deduced from simple linear algebra. Then we use a continuity argument to compare  $\Sigma(t)$  to  $\Sigma$ .

We add the argument  $t$  to quantities to indicate that they are defined in terms of  $\Sigma(t)$  instead of  $\Sigma$ . Using (A.5) (see also Figure 2.1), it is not hard to find that the gap  $a_{2k}(t) - a_{2k+1}(t)$  is increasing in  $t$ , and that

$$a_{2k}(t) \geq ct, \quad a_{2k+1}(t) \leq C \quad (10.9)$$

for some constants  $c, C > 0$ . Using that  $\|\Sigma(t) - \Sigma(s)\| \leq C|t-s|$ , we deduce from Lemma 4.8 that  $\|Q(t) - Q(s)\| \leq C|t-s|$  with high probability. Now let  $T$  be a fixed large time to be chosen later. By a continuity argument using Lemma 10.1 we therefore find that there exists a constant  $K \equiv K(T)$  such that for any  $t \in [1, T]$  we have

$$|\text{spec}(Q(t)) \cap (a_{2k}(t) - N^{-2/3+\varepsilon}, \infty)| = |\text{spec}(Q(t + K^{-1})) \cap (a_{2k}(t + K^{-1}) - N^{-2/3+\varepsilon}, \infty)| \quad (10.10)$$

with high probability.

It therefore remains to show that there exists a large enough  $T$ , depending only on  $\tau$ , such that for  $t = T$  the left-hand side of (10.10) is equal to  $d_k$ . Writing  $XX^* = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}$  as a block matrix, we find

$$Q(t) = \Sigma(t)^{1/2} XX^* \Sigma(t)^{1/2} = \begin{pmatrix} t\Sigma_1^{1/2} E_{11} \Sigma_1^{1/2} & \sqrt{t}\Sigma_1^{1/2} E_{12} \Sigma_2^{1/2} \\ \sqrt{t}\Sigma_2^{1/2} E_{21} \Sigma_1^{1/2} & \Sigma_2^{1/2} E_{22} \Sigma_2^{1/2} \end{pmatrix}. \quad (10.11)$$

From (A.4) we deduce that  $d_k \leq N - N_p \leq (1-c)N$ , by the assumption of Theorem 3.12. From [7, Theorem 2.10], we therefore deduce that  $c \leq E_{11} \leq C$  with high probability for some positive constants  $c, C$ . Moreover, from Lemma 4.8 we deduce that  $\|E_{12}\| + \|E_{21}\| + \|E_{22}\| \leq C$  with high probability. We conclude that for large enough  $t$  (depending on the constants  $c$  and  $C$



above) the matrix (10.11) has with high probability exactly  $d_k$  eigenvalues of order  $\asymp t$ , and all other eigenvalues are of order  $O(\sqrt{t})$ . Since  $\text{spec}(Q(t)) \cap [a_{2k+1}(t) + N^{-2/3+\tau}, a_{2k} - N^{-2/3+\tau}] = \emptyset$  with high probability by Lemma 10.1, we conclude that for large enough  $t$  the left-hand side of (10.10) is equal to  $d_k$  with high probability. This concludes the proof.  $\square$

Proposition 10.2 follows immediately from Lemmas 10.3 and 10.4. This concludes the second step outlined above.

Finally, the third step – the conclusion of the proof of Theorem 3.12 – follows from Theorem 3.6, Lemma 10.1, and Proposition 10.2 by repeating the analysis of [20, 32] with merely cosmetic changes, as explained in the proof of Theorem 3.12. This concludes the proof of Theorem 3.12 under the assumption (2.9). Moreover, as explained in Section 11.1, the assumption (2.9) may be easily removed.

Next, we note that the proof Theorem 3.7 under the assumption (2.9) is an easy consequence of Theorems 3.6 and 3.12, following the proof of [7, Theorem 3.12]. The key tool is the spectral decomposition (4.15). We omit further details.

Finally, the proof of Theorems 3.14 (iii) and 3.15 (iii) under the assumption (2.9) is exactly the same as that of Theorem 3.12. Indeed, one can check that in the proof of Theorem 3.12 only the weaker assumptions of Theorems 3.14 (iii) and 3.15 (iii) are needed. We omit the details.

**10.2. Edge universality assuming (2.9).** Finally, we prove the edge universality from Theorem 3.18 under the assumption (2.9).

PROOF OF THEOREM 3.18 ASSUMING (2.9). First, we claim that the joint asymptotic distribution of  $\mathbf{q}_{k,l}$  does not depend on the distribution of the entries of  $X$ , provided they satisfy (2.4) and (2.5). This is a routine application of the Green function comparison method near the edge, developed in [20, Section 6]. The argument of [20, Section 6] may be easily adapted to our case, using the linearizing block matrix  $H(z)$  and its inverse  $G(z)$  to write  $m_N = \frac{1}{N} \sum_{\mu \in \mathcal{I}_N} G_{\mu\mu}$ . The key technical inputs are the anisotropic local law from Theorem 3.6, the eigenvalue rigidity from Theorem 3.14 (iii), and the resolvent identities from Lemma 4.4. We omit further details.

We may therefore without loss of generality assume that  $X$  is Gaussian. By orthogonal / unitary invariance of the law of  $X$ , we may furthermore assume that  $\Sigma$  is diagonal. This concludes the proof.  $\square$

## 11. General matrices: removing (2.9) and extension to $\dot{Q}$

In this section we explain how our results, proved under the assumption (2.9) and for the matrix  $Q$ , may be generalized to hold without the assumption (2.9) and for the matrix  $\dot{Q}$  from (1.4) as well.

**11.1. How to remove (2.9).** For simplicity, throughout the proofs up to now we made the assumption (2.9). As advertised, this assumption is not necessary. In this section we explain how to dispense with it. The argument relies on simple approximation and linear algebra. Roughly, if  $\Sigma$  has a zero eigenvalue, we consider  $\Sigma + \varepsilon$  instead and let  $\varepsilon \downarrow 0$ ; if  $T$  is not square, we augment it to a square matrix by padding it out with zeros. While this extension is simple, we emphasize that it relies crucially on the fact we do not assume that  $\Sigma$  has a lower bound (the assumption (2.9) only requires the qualitative bound  $\Sigma > 0$ ).

We distinguish the cases  $\widehat{M} \geq M$  and  $\widehat{M} < M$ . Suppose first that  $\widehat{M} \geq M$ . We extend  $T$  to an  $\widehat{M} \times \widehat{M}$  matrix by setting  $\widehat{T} := \begin{pmatrix} 0 \\ T \end{pmatrix}$ . Define the  $\widehat{M} \times \widehat{M}$  matrices

$$\widehat{\Sigma} := \widehat{T}\widehat{T}^* = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma \end{pmatrix}, \quad \widehat{S} := \widehat{T}^*\widehat{T} = T^*T.$$

By polar decomposition, we have  $\widehat{T} = \widehat{U}\widehat{S}^{1/2}$ , where  $\widehat{U}$  is orthogonal. Therefore  $\widehat{\Sigma} = \widehat{U}\widehat{S}\widehat{U}^*$ . Moreover, from (2.11) we get  $m \equiv m_{\Sigma, N} = m_{\widehat{\Sigma}, N} = m_{\widehat{S}, N}$ , which we use tacitly in the following.

We define the  $(\widehat{M} + N) \times (\widehat{M} + N)$  matrix

$$\widehat{G} := \lim_{\varepsilon \downarrow 0} \begin{pmatrix} \widehat{U} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -(\widehat{S} + \varepsilon)^{-1} & X \\ X^* & -z \end{pmatrix}^{-1} \begin{pmatrix} \widehat{U}^* & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} z\widehat{\Sigma}^{1/2}(\widehat{T}X X^* \widehat{T}^* - z)^{-1}\widehat{\Sigma}^{1/2} & \widehat{T}X R_N \\ R_N X^* \widehat{T}^* & R_N \end{pmatrix},$$

where we used (4.3). Note that  $\widehat{G}$  has the block form

$$\widehat{G} = \begin{pmatrix} 0 & 0 \\ 0 & G \end{pmatrix}, \quad G := \begin{pmatrix} z\Sigma^{1/2}R_M\Sigma^{1/2} & TXR_N \\ R_N X^* T^* & R_N \end{pmatrix}. \quad (11.1)$$

Using this representation of  $G$  as a block of  $\widehat{G}$ , it is easy to drop the assumption (2.9). For example, suppose that Theorem 3.6 has been proved under the assumption (2.9). Applying it to  $\widehat{G}$  and using a simple approximation argument in  $\varepsilon$ , we find the following generalization of Theorem 3.6.

**PROPOSITION 11.1.** *Fix  $\tau > 0$ . Suppose that (2.9) and Assumption 2.1 hold. Suppose moreover that every edge  $k = 1, \dots, 2p$  satisfying  $a_k \geq \tau$  and every bulk component  $k = 1, \dots, p$  is regular in the sense of Definition 2.7. Then (3.10) and (3.11) hold for  $G$  defined in (11.1) and  $R_N$  defined in (3.1).*

In particular, Corollary 3.9 and Theorems 3.12, 3.14, 3.15, 3.16, and 3.18 follow easily from their counterparts proved under the assumption (2.9). This concludes the discussion for the case  $\widehat{M} \geq M$ .

Finally, we consider the case  $\widehat{M} < M$ . We set  $\widetilde{T} := (T, 0)$  and  $\widetilde{X} := \begin{pmatrix} X \\ Y \end{pmatrix}$ , where  $Y$  is an  $(M - \widehat{M}) \times N$  matrix, independent of  $X$ , with independent entries satisfying (2.4) and (2.5). Hence,  $\widetilde{T}$  is  $M \times M$  and  $\widetilde{X}$  is  $M \times N$ . Now we have  $TX = \widetilde{T}\widetilde{X}$ , and we have reduced the problem to the case  $\widehat{M} = M$ , which was dealt with above.

**11.2. Extension to  $\dot{Q}$ .** In this section we explain how our results have to be modified for  $\dot{Q}$  from (1.4). For simplicity of presentation, we make the assumption (2.9); it may be easily removed as explained in Section 11.1.

The matrix  $\dot{Q}$  is obtained from  $Q$  by replacing  $X$  with  $\dot{X} := X(1 - \mathbf{e}\mathbf{e}^*)$ . Generally, a dot on any quantity depending on  $X$  means that  $X$  has been replaced by  $\dot{X}$  in its definition. For example, we have

$$\dot{G} = \begin{pmatrix} -\Sigma^{-1} & \dot{X} \\ \dot{X}^* & -z \end{pmatrix}^{-1}.$$

As before it is easy to obtain  $\dot{R}_M$  and  $\dot{R}_N$  from  $\dot{G}$ .

Moreover, in analogy to (3.5) we define  $\dot{\Pi} := \Pi - (m + z^{-1})\mathbf{e}\mathbf{e}^*$ . (Recall the convention that  $\mathbf{e} \in \mathbb{R}^{\mathcal{I}}$  is the natural embedding of  $\mathbf{e} \in \mathcal{R}^{\mathcal{I}_N}$  obtained by adding zeros.) Let  $\mathbf{S} \subset \mathbf{D}$  be a spectral domain. As in Definition 3.20, we say that the anisotropic local law holds for  $\dot{G}$  if

$$\underline{\Sigma}^{-1}(\dot{G}(z) - \dot{\Pi}(z))\underline{\Sigma}^{-1} = O_{\prec}(\Psi(z)) \quad (11.2)$$

uniformly in  $z \in \mathbf{S}$ , and that averaged local law for  $\dot{G}$  holds if

$$\dot{m}_N(z) - m(z) = O_{\prec}\left(\frac{1}{N\eta}\right) \quad (11.3)$$

uniformly in  $z \in \mathbf{S}$ .

The local law for  $\dot{G}$  reads as follows.

**THEOREM 11.2 (LOCAL LAWS FOR  $\dot{G}$ ).** *Fix  $\tau > 0$ . Suppose that  $X$  and  $\Sigma$  satisfy (2.9) and Assumption 2.1. Let  $\mathbf{S} \subset \mathbf{D}$  be a spectral domain. Then the entrywise and averaged local laws hold for  $\dot{G}$  in the sense of (11.2) and (11.3) provided they hold for  $G$  in the sense of Definition 3.20 (ii) and (iii).*

**PROOF.** To simplify notation, we omit the factor  $\frac{N}{N-1}$  in the definition of  $\dot{Q}$ . It may easily be put back by scaling the argument  $z$ . We prove the anisotropic local law by estimating the four blocks of  $\dot{G}$  individually. Using simple linear algebra and (4.1) and (4.3), we get

$$\dot{G}_M = G_M + \frac{G_M X \mathbf{e} \mathbf{e}^* X^* G_M}{z - \mathbf{e}^* X^* G_M X \mathbf{e}} = G_M + \frac{G_M X \mathbf{e} \mathbf{e}^* X^* G_M}{z^2 \mathbf{e}^* G_N \mathbf{e}}. \quad (11.4)$$

Since  $G$  satisfies the anisotropic local law by assumption, it is easy to deduce from Lemma 4.4 and (4.21) that

$$\langle \mathbf{v}, \dot{G}_M \mathbf{w} \rangle = \langle \mathbf{v}, G_M \mathbf{w} \rangle + O_{\prec}(\Psi^2 |\Sigma \mathbf{v}| |\Sigma \mathbf{w}|)$$

for  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{\mathcal{I}_M}$ .

Next, define the orthogonal projection  $\pi := \mathbf{e}\mathbf{e}^*$  in the space  $\mathbb{R}^{\mathcal{I}_N}$ , as well as its orthogonal complement  $\bar{\pi} := I_N - \pi$ . Now the upper-right block of  $\dot{G}$  is equal to

$$z^{-1} \dot{G}_M \dot{X} = z^{-1} G_M X \bar{\pi} + \frac{G_M X \mathbf{e} \mathbf{e}^* X^* G_M X \bar{\pi}}{z^2 \mathbf{e}^* G_N \mathbf{e}}.$$

From (4.3) we get  $X^* G_M X = z^2 G_N + z$ , so that, using the anisotropic local law for  $G$ , we conclude

$$\langle \mathbf{v}, z^{-1} \dot{G}_M \dot{X} \mathbf{w} \rangle = O_{\prec}(\Psi |\Sigma \mathbf{v}| |\mathbf{w}|)$$

for  $\mathbf{v} \in \mathbb{R}^{\mathcal{I}_M}$  and  $\mathbf{w} \in \mathbb{R}^{\mathcal{I}_N}$ . The lower-left block is dealt with analogously.

Finally, using (4.3) for  $\dot{G}$  as well as (11.4), we get

$$\dot{G}_N = z^{-2} \dot{X}^* \dot{G}_M \dot{X} - z^{-1} = \bar{\pi} G_N \bar{\pi} - \pi z^{-1} + \frac{\bar{\pi} G_N \pi G_N \bar{\pi}}{\mathbf{e}^* G_N \mathbf{e}}.$$

Using the anisotropic law for  $G$  and Lemma 4.4, we therefore get

$$\langle \mathbf{v}, \dot{G}_N \mathbf{w} \rangle = \langle \mathbf{v}, (\bar{\pi} G_N \bar{\pi} - \pi z^{-1}) \mathbf{w} \rangle + O_{\prec}(\Psi^2 |\mathbf{v}| |\mathbf{w}|) \quad (11.5)$$

for  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{\mathcal{I}_N}$ . This concludes the proof of the anisotropic local law.

Moreover, the averaged local law for  $\dot{G}$  is easy to deduce from (11.5) by setting  $\mathbf{v} = \mathbf{w} = \mathbf{e}_\mu$  and summing over  $\mu \in \mathcal{I}_N$ . In this way, the averaged local law for  $\dot{G}$  follows from the averaged and anisotropic local laws for  $G$ .  $\square$

We also obtain eigenvalue rigidity for the eigenvalues  $\dot{\lambda}_1 \geq \dot{\lambda}_2 \geq \dots \geq \dot{\lambda}_M$  of  $\dot{Q}$ . For instance, Theorem 3.12 has the following counterpart.

**THEOREM 11.3 (EIGENVALUE RIGIDITY FOR  $\dot{Q}$ ).** *Theorem 3.12 remains valid if  $\lambda_k$  is replaced with  $\dot{\lambda}_k$ .*

**PROOF.** The proof follows that of Theorem 3.12 to the letter, using Theorem 11.2 as input. In fact, as explained around (10.4), we need a stronger bound than (11.3) outside of the spectrum; this stronger bound follows easily from (11.5) and the analogous stronger bound for  $G_N$  established in (10.4).  $\square$

Finally, we obtain edge universality for  $\dot{Q}$ . The following result is proved exactly like Theorem 3.18, using Theorems 11.2 and 11.3 as input.

**THEOREM 11.4 (EDGE UNIVERSALITY FOR  $\dot{Q}$ ).** *Theorem 3.18 remains valid if  $\lambda_i$  is replaced with  $\dot{\lambda}_i$ .*

## 12. Deformed Wigner matrices

In this section we apply our method to deformed Wigner matrices as a further illustration of its applicability. Since the statements and arguments are similar to those of the previous sections, we keep the presentation concise.

**12.1. Model and results.** Let  $W = W^*$  be an  $N \times N$  Wigner matrix whose upper-triangular entries  $(W_{ij} : 1 \leq i \leq j \leq N)$  are independent and satisfy the same conditions (2.4) and (2.5) as  $X_{i\mu}$ . Let  $A = A^*$  be a deterministic  $N \times N$  matrix satisfying  $\|A\| \leq \tau^{-1}$ . For definiteness, we suppose that  $W$  and  $A$  are real symmetric matrices, remarking that similar results also hold for complex Hermitian matrices.

The main result of this section is the anisotropic local law for the *deformed Wigner matrix*  $W + A$ , analogous to Theorem 3.21. As an application, we establish the edge universality of  $W + A$ . We remark that the entrywise local law and edge universality were previously established in [27] under the assumption that  $A$  is diagonal. Before that, edge universality was established in [9, 23, 33] under the assumption that  $W$  is a GUE matrix. A somewhat different direction was pursued in [4, 6, 24], where local laws for the sum of two matrices that are invariant under unitary or orthogonal conjugations were analysed.

In order to avoid confusion with similar quantities defined previously for sample covariance matrices, we use the superscript  $W$  to distinguish quantities defined in terms of the deformed Wigner matrix  $W + A$ . The Stieltjes transform  $m^W$  of the asymptotic eigenvalue density of  $W + A$  is defined as the unique solution of the equation

$$m^W(z) = \frac{1}{N} \text{Tr}(-m^W(z) + A - z)^{-1}$$

satisfying  $\text{Im } m^W(z) > 0$  for  $\text{Im } z > 0$ . (See e.g. [31] for details.) Note that  $m^W$  only depends on the spectrum of  $A$  and not on its eigenvectors. For simplicity, following [27] we assume that support of the asymptotic eigenvalue density  $\varrho^W(E) := \lim_{\eta \downarrow 0} \pi^{-1} \text{Im } m^W(E + i\eta)$  is an interval, which we denote by  $[L_-, L_+]$ . This condition is however not necessary for our method, which may in

particular easily be extended to the multi-cut case, using an argument similar to the one developed in the context of sample covariance matrices in Section 10 and Appendix A.

We denote the eigenvalues of  $W + A$  by  $\lambda_1(W + A) \geq \lambda_2(W + A) \geq \dots \geq \lambda_N(W + A)$ . Moreover, we define the resolvent  $G^W(z) := (W + A - z)^{-1}$ , as well as

$$\Pi^W := \frac{1}{-m^W + A - z}, \quad \Psi^W := \sqrt{\frac{\operatorname{Im} m^W}{N\eta}} + \frac{1}{N\eta}.$$

The following definition is the analogue of Definition 3.20 for deformed Wigner matrices.

**DEFINITION 12.1 (LOCAL LAWS).** Define  $\mathbf{D}^W := \{z : |E| \leq \tau^{-1}, N^{-1+\tau} \leq \operatorname{Im} z \leq \tau^{-1}\}$ , and let  $\mathbf{S} \subset \mathbf{D}^W$  be a spectral domain, i.e. for each  $z \in \mathbf{S}$  we have  $\{w \in \mathbf{D}^W : \operatorname{Re} w = \operatorname{Re} z, \operatorname{Im} w \geq \operatorname{Im} z\} \subset \mathbf{S}$ .

(i) We say that the entrywise local law holds with parameters  $(W, A, \mathbf{S})$  if

$$(G^W(z) - \Pi^W(z))_{st} = O_{\prec}(\Psi^W(z))$$

uniformly in  $z \in \mathbf{S}$  and  $1 \leq s, t \leq N$ .

(ii) We say that the anisotropic local law holds with parameters  $(W, A, \mathbf{S})$  if

$$G^W(z) - \Pi^W(z) = O_{\prec}(\Psi^W(z))$$

uniformly in  $z \in \mathbf{S}$ .

(iii) We say that the averaged local law holds with parameters  $(W, A, \mathbf{S})$  if

$$\frac{1}{N} \operatorname{Tr} G^W(z) - m^W(z) = O_{\prec}\left(\frac{1}{N\eta}\right)$$

uniformly in  $z \in \mathbf{S}$ .

In analogy to (3.20), we always assume that  $|-m^W(z) + a_i - z| \geq \tau$  for all  $z \in \mathbf{S}$  and  $a_i \in \operatorname{spec}(A)$ ; this assumption has been verified under general hypotheses on the spectrum of  $A$  in [27]. In particular, it implies that  $\|\Pi\| \leq \tau^{-1}$ .

The following result is the analogue of Theorem 3.21. Throughout the following we denote by  $W^{\text{Gauss}}$  a GOE matrix and by  $D \equiv D_A$  the diagonalization of  $A$ .

**THEOREM 12.2 (GENERAL LOCAL LAWS).** Let  $W$  and  $A$  be as above. Fix  $\tau > 0$  and let  $\mathbf{S} \subset \mathbf{D}^W(\tau, N)$  be a spectral domain. Suppose that either (a)  $\mathbb{E}W_{ij}^3 = 0$  for all  $i, j$  or (b) there exists a constant  $c_0 > 0$  such that  $\Psi^W(z) \leq N^{-1/4-c_0}$  for all  $z \in \mathbf{S}$ .

(i) If the entrywise local law holds with parameters  $(W^{\text{Gauss}}, D, \mathbf{S})$ , then the anisotropic local law holds with parameters  $(W, A, \mathbf{S})$ .

(ii) If the entrywise local law and the averaged local law hold with parameters  $(W^{\text{Gauss}}, D, \mathbf{S})$ , then the averaged local law holds with parameters  $(W, A, \mathbf{S})$ .

Note that the assumptions of Theorem 12.2 depend on both the deterministic matrix  $A$  and the spectral domain  $\mathbf{S}$ .

The following result guarantees the rigidity of the extreme eigenvalues of  $H + A$ . As explained below, the rigidity of all eigenvalues will be a simple consequence of Theorems 12.2 (ii) and 12.4.

**DEFINITION 12.3.** *We say the extreme eigenvalues of  $W + A$  are rigid if  $[\lambda_1(W + A) - L_+]_+ \prec N^{-2/3}$  and  $[\lambda_N(W + A) - L_-]_- \prec N^{-2/3}$ .*

**THEOREM 12.4.** *Suppose that, for any fixed and small enough  $\tau > 0$ , the entrywise local law and the averaged local law hold with parameters  $(W^{\text{Gauss}}, D, \mathbf{D}^W(\tau, N))$ . Moreover, suppose that the extreme eigenvalues of  $W^{\text{Gauss}} + D$  are rigid. Then the extreme eigenvalues of  $W + A$  are rigid.*

Together, Theorems 12.2 (ii) and 12.4 easily yield the rigidity of the eigenvalues, as explained in the proof of [20, Theorem 2.2]. We may apply Theorem 12.2 to extend the results of [27] to arbitrary non-diagonal  $A$ . For instance, we obtain the following edge universality result.

**THEOREM 12.5 (EDGE UNIVERSALITY).** *Let  $W$  and  $A$  be as above. Suppose that the spectrum  $D$  of the matrix  $A$  satisfies the assumptions of Theorem 12.4. Then for any fixed  $l$  and smooth  $f : \mathbb{R}^l \rightarrow \mathbb{R}$  there is a constant  $c_0 > 0$  such that*

$$f\left[\left(N^{2/3}(\lambda_i(W + A) - L_+)\right)_{i=1}^l\right] - f\left[\left(N^{2/3}(\lambda_i(W^{\text{Gauss}} + D) - L_+)\right)_{i=1}^l\right] = O(N^{-c_0}). \quad (12.1)$$

*A similar result holds for the extreme eigenvalues near the left edge.*

In [27], the assumptions of Theorem 12.4 were verified for a large class of diagonal matrices  $D$ . Moreover, for such matrices  $D$ , it was proved that the limit of the second term on the left-hand side of (12.1) is governed by the Tracy-Widom-Airy statistics. Theorem 12.5 therefore provides an extension of [27, Theorem 2.8] to non-diagonal matrices  $A$ . We refer to [27] for the detailed statements about the distribution of the eigenvalues of  $W^{\text{Gauss}} + D$ .

Further applications of Theorem 12.2 include a study of the eigenvectors of  $W + A$  and the outliers of finite-rank perturbations of  $W + A$ . We do not pursue these questions here.

The rest of this section is devoted to the proof of Theorems 12.2, 12.4, and 12.5. To unburden notation, from now on we omit the superscripts  $W$ , with the understanding that all quantities in this section are defined in terms of deformed Wigner matrices.

**12.2. Proof of Theorem 12.2 (i).** Exactly as in Section 6, we first prove that the entrywise local law with parameters  $(W^{\text{Gauss}}, D, \mathbf{S})$  implies the anisotropic local law with parameters  $(W^{\text{Gauss}}, A, \mathbf{S})$ . The proof is very similar to the one from Section 6. As explained there, the details may be taken over with trivial modifications from [7, Sections 5 and 7], where the proof is given for  $A = 0$ .

What remains, therefore, is the proof of the anisotropic local law with parameters  $(W, A, \mathbf{S})$ , starting from the anisotropic local law with parameters  $(W^{\text{Gauss}}, A, \mathbf{S})$ . The basic strategy is similar to the one in section 7. Define  $\hat{\mathbf{S}}$  and  $\hat{\mathbf{S}}_m$  as in Section 7, and denote by  $\mathbb{S} \subset \mathbb{R}^N$  the unit sphere of  $\mathbb{R}^N$ . The following definitions are analogous to (7.13) and (7.14).

**(A<sub>m</sub>)** For all  $z \in \hat{\mathbf{S}}_m$  and all deterministic  $\mathbf{v} \in \mathbb{S}$  we have

$$\text{Im } G_{\mathbf{v}\mathbf{v}}(z) \prec \text{Im } m(z) + N^{C_0\delta} \Psi(z). \quad (12.2)$$

( $\mathbf{C}_m$ ) For all  $z \in \widehat{\mathbf{S}}_m$  and all deterministic  $\mathbf{v} \in \mathbb{S}$  we have

$$|G_{\mathbf{v}\mathbf{v}}(z) - \Pi_{\mathbf{v}\mathbf{v}}(z)| \prec N^{C_0\delta}\Psi(z). \quad (12.3)$$

Here  $C_0$  is a constant that may depend only on  $\tau$ , and  $\delta$  satisfies (7.9).

Clearly, ( $\mathbf{A}_0$ ) holds, and ( $\mathbf{C}_m$ ) implies ( $\mathbf{A}_m$ ). As in Lemma 7.4, we only need to prove that ( $\mathbf{A}_{m-1}$ ) implies ( $\mathbf{C}_m$ ). The necessary a priori bound is summarized in the following result, which may be proved like Lemma 7.13.

LEMMA 12.6. *If ( $\mathbf{A}_{m-1}$ ) holds then*

$$G(z) = O_{\prec}(N^{2\delta}), \quad \frac{\operatorname{Im} G_{\mathbf{v}\mathbf{v}}(z)}{N\eta} \prec N^{(C_0+1)\delta}\Psi^2 \quad (12.4)$$

for all  $z \in \widehat{\mathbf{S}}_m$  and deterministic  $\mathbf{v} \in \mathbb{S}$ .

Making minor adjustments to the argument of Sections 7.3 and 7.4, we find that it suffices to prove the following result, which is analogous to Lemma 7.30.

LEMMA 12.7. *For  $\mathbf{v} \in \mathbb{S}$  define*

$$F_{\mathbf{v}}(W, z) := |G_{\mathbf{v}\mathbf{v}}(z) - \Pi_{\mathbf{v}\mathbf{v}}(z)|.$$

Let  $z \in \widehat{\mathbf{S}}$  and suppose that (12.4) holds. Then we have

$$N^{-n/2} \sum_{i \leq j} \mathbb{E} \left| \left( \frac{\partial}{\partial W_{ij}} \right)^n F_{\mathbf{v}}^p(W, z) \right| = O \left( (N^{C_0\delta}\Psi(z))^p + \mathbb{E} F_{\mathbf{v}}^p(W, z) \right) \quad (12.5)$$

for any  $n = 4, 5, \dots, 4p$ . Moreover, if in addition  $\Psi(z) \leq N^{-1/4-c_0}$  then (12.5) holds also for  $n = 3$ .

Here we also used that the function  $\eta \mapsto \Psi(E + i\eta)$  is decreasing, so that if the assumption  $\Psi \leq N^{-1/4-c_0}$  holds at  $z \in \mathbf{S}$  then it also holds for all  $w \in \mathbf{S}$  satisfying  $\operatorname{Re} w = \operatorname{Re} z$  and  $\operatorname{Im} w \geq \operatorname{Im} z$ .

The rest of this subsection is devoted to the proof of Lemma 12.7. We note that the word structure describing the derivatives on the left-hand side of (12.5) is very similar to that of (7.30) (given in Definition 7.19). Hence the proof of Lemma 12.7 is similar to that of Lemma 7.17.

The proof of (12.5) for  $n \geq 4$  is a trivial modification of the argument given in Section 7.5, whose details we omit. What remains therefore is the proof of (12.5) for  $n = 3$  under the assumption  $\Psi \leq N^{-1/4-c_0}$ . From now on we omit the arguments  $z$ .

The main new ingredient of the proof for deformed Wigner matrices is a further iteration step at fixed  $z$ . Suppose that

$$G - \Pi = O_{\prec}(N^{2\delta}\Phi) \quad (12.6)$$

for some  $\Phi \leq 1$ . Since  $\|\Pi\| \leq C$ , the estimate (12.6) is stronger than the first estimate of (12.4). Note that, by assumption (12.4), the estimate (12.6) holds for  $\Phi = 1$ . Assuming (12.6), we shall prove a self-improving bound of the form

$$N^{-3/2} \sum_{i \leq j} \mathbb{E} \left| \left( \frac{\partial}{\partial W_{ij}} \right)^3 F_{\mathbf{v}}^p(W) \right| = O \left( (N^{C_0\delta}\Psi)^p + (N^{-c_0/2}\Phi)^p + \mathbb{E} F_{\mathbf{v}}^p(W, z) \right). \quad (12.7)$$



Once (12.7) is proved, we may use it iteratively to obtain increasingly accurate bounds for the left-hand side of (12.3). After each step, we obtain an improved a priori bound (12.6), whereby  $\Phi$  is reduced by powers of  $N^{-c_0/2}$ . After  $O(1/c_0)$  iterations, (12.5) for  $n = 3$  follows.

It therefore suffices to prove (12.7) under the assumptions (12.4) and (12.6). As in Section 7.5, it suffices to prove

$$N^{-3/2} \left| \sum_{i \leq j} A_{\mathbf{v},i,j}(w_0)^{p-q} \prod_{r=1}^q A_{\mathbf{v},i,j}(w_r) \right| \prec F_{\mathbf{v}}^{p-q}(W) (N^{(C_0-1)\delta} \Psi + N^{-c_0} \Phi)^q, \quad (12.8)$$

where the words  $w$  and their values  $A_{\mathbf{v},i,j}(\cdot)$  are defined similarly to Definition 7.19. More precisely, abbreviate  $\tilde{G} := G - \Pi$ . Then for  $n(w) = 0$  we have  $A_{\mathbf{v},i,j}(w) = \tilde{G}_{\mathbf{v}\mathbf{v}}$ , and for  $n(w) \geq 1$  the random variable  $A_{\mathbf{v},i,j}(w)$  is a product of entries  $G_{\mathbf{v}i}$ ,  $G_{\mathbf{v}j}$ ,  $G_{ij}$ ,  $G_{ji}$ ,  $G_{i\mathbf{v}}$ , and  $G_{j\mathbf{v}}$ , as in (7.32). Clearly, to prove (12.8) it suffices to prove

$$N^{-3/2} \left| \sum_{i \leq j} \prod_{r=1}^q A_{\mathbf{v},i,j}(w_r) \right| \prec (N^{(C_0-1)\delta} \Psi + N^{-c_0} \Phi)^q. \quad (12.9)$$

We discuss the three cases  $q = 1, 2, 3$  separately.

*The case  $q = 1$ .* The single factor  $A_{\mathbf{v},i,j}(w_r)$  is of the form

$$G_{\mathbf{v}i} G_{ab} G_{cd} G_{j\mathbf{v}} \quad \text{or} \quad G_{\mathbf{v}i} G_{ab} G_{cd} G_{i\mathbf{v}} \quad (12.10)$$

or a term obtained from one of these two by exchanging  $i$  and  $j$ ; here  $a, b, c, d \in \{i, j\}$ . We deal first with the first expression; the others are dealt with analogously. We split it according to

$$G_{\mathbf{v}i} G_{ab} G_{cd} G_{j\mathbf{v}} = G_{\mathbf{v}i} \Pi_{ab} \Pi_{cd} G_{j\mathbf{v}} + G_{\mathbf{v}i} \tilde{G}_{ab} \Pi_{cd} G_{j\mathbf{v}} + G_{\mathbf{v}i} \Pi_{ab} \tilde{G}_{cd} G_{j\mathbf{v}} + G_{\mathbf{v}i} \tilde{G}_{ab} \tilde{G}_{cd} G_{j\mathbf{v}}. \quad (12.11)$$

We deal with the summation over  $i$  using the estimate

$$\frac{1}{N} \sum_i |G_{\mathbf{v}i}|^2 = \frac{\text{Im } G_{\mathbf{v}\mathbf{v}}}{N\eta} \prec N^{(C_0+1)\delta} \Psi^2, \quad (12.12)$$

where in the last step we used (12.4); a similar estimate holds for the summation over  $j$ . Using (12.12), (12.6), and  $\|\Pi\| \leq \tau^{-1}$ , we may estimate the second term on the right-hand side of (12.11) as

$$N^{-3/2} \sum_{i,j} |G_{\mathbf{v}i} \tilde{G}_{ab} \Pi_{cd} G_{j\mathbf{v}}| \prec N^{-3/2} \sum_{i,j} N^{2\delta} \Phi |G_{\mathbf{v}i} G_{j\mathbf{v}}| \prec N^{1/2} N^{(C_0+3)\delta} \Psi^2 \Phi \leq N^{-c_0} \Phi,$$

provided  $\delta$  is chosen small enough, depending on  $\tau$  and  $c_0$ . The third and fourth terms on the right-hand side of (12.11) are estimated in exactly the same way.

What remains is the first term on the right-hand side of (12.11). Here taking the absolute value inside the sum is not affordable. Instead, we use the first A priori bound of (12.4) to estimate  $|\sum_{i:i \leq j} G_{si} \Pi_{ab} \Pi_{cd}| \prec N^{1/2+2\delta}$ , where we used that  $\Pi$  is deterministic and satisfies the bound  $\sum_i |\Pi_{ab} \Pi_{cd}| \leq \tau^{-2} N$ . Combining this estimate with  $\sum_j |G_{j\mathbf{v}}| \prec N^{1+(C_0/2+1)\delta} \Psi$  from (12.12), we find

$$N^{-3/2} \left| \sum_{i \leq j} G_{\mathbf{v}i} \Pi_{ab} \Pi_{cd} G_{j\mathbf{v}} \right| \prec N^{(C_0/2+3)\delta} \Psi \leq N^{(C_0-1)\delta} \Psi,$$



provided  $C_0 \geq 8$ .

Finally, if  $A_{\mathbf{v},i,j}(w_r)$  is the second expression of (12.10), the argument is analogous. In this case at least one of the terms  $G_{ab}$  and  $G_{cd}$  is of the form  $G_{ij}$  or  $G_{ji}$ , so that, in the analogue of the first term on the right-hand side of (12.11), we may use the improved estimate  $\sum_i |\Pi_{ab}\Pi_{cd}| \leq \tau^{-1} \sum_i |\Pi_{ij}| \leq \tau^{-2} N^{1/2}$ , as follows from  $\|\Pi\| \leq \tau^{-1}$ .

*The case  $q = 2$ .* In this case the product  $\prod_{r=1}^q A_{\mathbf{v},i,j}(w_r)$  is of the form

$$G_{\mathbf{v}i}(GG_{j\mathbf{v}}^*)_{jt}G_{\mathbf{v}i}G_{ji}G_{j\mathbf{v}} \quad \text{or} \quad G_{\mathbf{v}i}G_{j\mathbf{v}}G_{\mathbf{v}i}G_{jj}G_{i\mathbf{v}}, \quad (12.13)$$

or an expression obtained from one of these two by exchanging  $i$  and  $j$ . The contribution of the first expression of (12.13) is estimated using (12.4) and (12.12) by

$$N^{-3/2} \sum_{i,j} |G_{\mathbf{v}i}G_{j\mathbf{v}}G_{\mathbf{v}i}G_{ji}G_{j\mathbf{v}}| \prec N^{1/2+(2C_0+4)\delta} \Psi^4 \leq N^{(2C_0+4)\delta-2c_0} \Psi^2 \leq \Psi^2,$$

provided  $\delta$  is chosen small enough, depending on  $\tau$  and  $c_0$ .

Next, in order to estimate the contribution of the second expression of (12.13), we split

$$\sum_j G_{j\mathbf{v}}G_{jj} = \sum_j G_{j\mathbf{v}}\Pi_{jj} + \sum_j G_{j\mathbf{v}}\tilde{G}_{jj} = O_{\prec}(N^{1/2+2\delta}) + O_{\prec}(N^{1+(C_0/2+1)\delta}\Psi\Phi), \quad (12.14)$$

where we used (12.4), (12.6), and (12.12). Similarly, using (12.4) and (12.12) we get

$$\left| \sum_i G_{\mathbf{v}i}G_{\mathbf{v}i}G_{i\mathbf{v}} \right| \prec N^{1+(C_0+3)\delta} \Psi^2 \leq N^{1/2-c_0} \quad (12.15)$$

for small enough  $\delta$ . Putting (12.14) and (12.15) together, it is easy to deduce (12.9).

*The case  $q = 3$ .* Now  $\prod_{r=1}^q A_{\mathbf{v},i,j}(w_r)$  is of the form  $(G_{\mathbf{v}i}G_{j\mathbf{v}})^3$ , or an expression obtained by exchanging  $i$  and  $j$  in some of the three factors. We estimate, using (12.6) and (12.12),

$$\begin{aligned} \left| \sum_i G_{\mathbf{v}i}^3 \right| &\leq 4 \sum_i |\tilde{G}_{\mathbf{v}i}|^3 + 4 \sum_i |\Pi_{\mathbf{v}i}|^3 \\ &\prec \sum_i (|G_{\mathbf{v}i}|^2 + |\Pi_{\mathbf{v}i}|^2) N^{2\delta} \Phi + 1 \prec N^{1+(C_0+1)\delta} \Psi^2 \Phi + N^{2\delta} \Phi + 1, \end{aligned}$$

where we used that  $\sum_i |\Pi_{\mathbf{v}i}|^2 \leq \tau^{-2}$ . Now (12.9) is easy to conclude using  $\Psi \geq N^{-1/2+\tau/2}$ .

**12.3. Proof of Theorem 12.2 (ii).** The proof is similar to that from Section 9. As in Section 12.2, it suffices to prove the following result.

**LEMMA 12.8.** *Let  $z \in \mathbf{S}$  and suppose that the anisotropic local law holds at  $z$ . Define  $\tilde{F}(W) := \left| \frac{1}{N} \sum_i G_{ii} - m \right|$ . Then we have*

$$N^{-n/2} \sum_{i \leq j} \left| \mathbb{E} \left( \frac{\partial}{\partial W_{ij}} \right)^n \tilde{F}^p(W) \right| = O \left( (N^\delta \Psi^2)^p + \left( \frac{N^{-c_0/2}}{N\eta} \right)^p + \mathbb{E} \tilde{F}^p(W) \right) \quad (12.16)$$

for any  $n = 4, 5, \dots, 4p$ . Moreover, if in addition  $\Psi(z) \leq N^{-1/4-c_0}$  then (12.16) holds also for  $n = 3$ .

The case  $n \geq 4$  can be easily proved as in covariance case. We therefore focus on the case  $n = 3$  in Lemma 12.8. The proof is similar to the discussion below (12.8). The main difference is that for each  $q$  we have some extra averaging  $N^{-q} \sum_{s_1, \dots, s_q}(\cdot)$ , and we need to extract an extra factor  $\Psi^q$  (or, alternatively,  $(N^{-1-c_0/2}\eta^{-1}\Psi^{-1})^q$ ) from this average. We take over the notations from Sections 7.5 and 12.2 without further comment. We consider the three cases  $q = 1, 2, 3$  separately, and tacitly use the anisotropic local law from Theorem 12.2 (i).

*The case  $q = 1$ .* Consider first the case  $A_{s,s,i,j}(w_1) = G_{si}G_{jj}G_{ij}G_{is}$ . We estimate

$$\left| \sum_j G_{ij}G_{jj} \right| \leq \left| \sum_j G_{ij}\Pi_{jj} \right| + O_{\prec}(N\Psi^2) \prec N^{1/2}, \quad \sum_i |G_{si}G_{is}| \prec N\Psi^2.$$

This gives

$$\left| \frac{1}{N} \sum_s \sum_{i,j} G_{si}G_{jj}G_{ij}G_{is} \right| \prec N^{3/2}\Psi^2,$$

as desired. Next, in the case  $A_{s,s,i,j}(w_1) = G_{si}G_{ji}G_{ji}G_{js}$  we estimate

$$\left| \sum_{i,j} G_{si}G_{ji}G_{ji}G_{js} \right| \leq \sum_{i,j} (|G_{si}|^2 + |G_{sj}|^2)|G_{ji}|^2 \prec N^2\Psi^4 \prec N^{3/2}\Psi^2.$$

Finally, in the case  $A_{s,s,i,j}(w_1) = G_{si}G_{jj}G_{ii}G_{js}$  we estimate

$$\left| \sum_i G_{si}G_{ii} \right| \leq \left| \sum_i G_{si}\tilde{G}_{ii} \right| + \left| \sum_i G_{si}\Pi_{ii} \right| \prec N\Psi^2 + N^{1/2}\Psi \leq CN\Psi^2.$$

Using a similar bound for the sum over  $j$ , we find  $\frac{1}{N} \sum_s \sum_{i,j} G_{si}G_{jj}G_{ii}G_{js} = O_{\prec}(N^2\Psi^4) = O_{\prec}(N^{3/2}\Psi^2)$ . All other terms are obtained from these three by exchanging  $i$  and  $j$ .

*The case  $q = 2$ .* In this case  $N^{-1} \sum_{s_1, s_2} \prod_{r=1}^2 A_{s_r, s_r, i, j}(w_r)$  is of the form

$$\frac{1}{N^2} \sum_{s_1, s_2} G_{s_1 i} G_{j s_1} G_{s_2 i} G_{j i} G_{j s_2} \quad \text{or} \quad \frac{1}{N^2} \sum_{s_1, s_2} G_{s_1 i} G_{j s_1} G_{s_2 i} G_{j j} G_{i s_2}, \quad (12.17)$$

or an expression obtained from one of these two by exchanging  $i$  and  $j$ . These may be written as

$$\frac{1}{N^2} (G^2)_{ji}^2 G_{ji} \quad \text{or} \quad \frac{1}{N^2} (G^2)_{ji} (G^2)_{ii} G_{jj}. \quad (12.18)$$

We estimate the contribution of the first expression by

$$\begin{aligned} \left| \sum_{i,j} \frac{1}{N^2} (G^2)_{ji}^2 G_{ji} \right| &\prec \frac{1}{N^2} \text{Tr} |G|^4 = \frac{1}{N^2} \sum_k \frac{1}{((\lambda_k - E)^2 + \eta^2)^2} \leq \frac{1}{N^2 \eta^3} \sum_k \frac{\eta}{(\lambda_k - E)^2 + \eta^2} \\ &\prec N^2 \frac{\text{Im } m + \Psi}{(N\eta)^3} \leq 2N^2 \frac{1}{(N\eta)^2} \Psi^2 \leq 2N^{3/2} \left( \frac{N^{-c_0}}{N\eta} \right)^2. \end{aligned}$$

Next, we split the contribution of the second expression of (12.18) as

$$\sum_{i,j} \frac{1}{N^2} (G^2)_{ji} (G^2)_{ii} G_{jj} = \sum_{i,j} \frac{1}{N^2} (G^2)_{ji} (G^2)_{ii} \tilde{G}_{jj} + \sum_{i,j} \frac{1}{N^2} \Pi_{jj} (G^2)_{ji} (G^2)_{ii}.$$

Using the anisotropic local law, it is easy to prove that  $|(G^2)_{ij}| \prec N\Psi^2$  and  $|\sum_j \Pi_{jj}(G^2)_{ji}| \prec N^{3/2}\Psi^2$ . Therefore

$$\begin{aligned} \left| \sum_{i,j} \frac{1}{N^2} (G^2)_{ji} (G^2)_{ii} G_{jj} \right| &\prec \Psi^3 (\text{Tr} |G|^4)^{1/2} + N^{3/2} \Psi^4 \\ &\prec N^2 \frac{1}{N\eta} \Psi^4 + N^{3/2} \Psi^4 \leq N^{3/2} \Psi^2 \frac{N^{-c_0}}{N\eta} + N^{3/2} \Psi^4, \end{aligned}$$

where we estimate  $\text{Tr} |G|^4$  as above. This concludes the proof in the case  $q = 2$ .

*The case  $q = 3$ .* In this case  $\sum_{s_1, s_2, s_3} \prod_{r=1}^3 A_{s_r, s_r, i, j}(w_r)$  is of the form  $(G^2)_{ij}^3$ , or an expression obtained by exchanging  $i$  and  $j$  in some of the three factors. We estimate its contribution by

$$\left| \frac{1}{N^3} \sum_{i,j} (G^2)_{ij}^3 \right| \prec N^{-2} \text{Tr} |G|^4 \Psi^2 \leq N^2 \frac{1}{(N\eta)^2} \Psi^4 \prec N^{3/2} \left( \frac{N^{-c_0}}{N\eta} \right)^2 \Psi^2,$$

which concludes the proof.

**12.4. Proof of Theorem 12.4.** The proof is analogous to that of Theorem 3.12 in Section 10. Define the domain

$$\mathbf{S}_\tau := \left\{ z : N^{-2/3+\tau} \leq \text{dist}(E, [L_-, L_+]) \leq \tau^{-1}, N^{-\tau} \kappa(E) \leq \eta \leq N^{-\delta/2} \kappa(E) \right\}.$$

From the assumptions on  $W^{\text{Gauss}} + D$ , it is not hard to deduce the estimate

$$\frac{1}{N} \text{Tr} (W^{\text{Gauss}} + D - z)^{-1} - m(z) = O_{\prec} \left( \frac{N^{-c}}{N\eta} \right)$$

for all  $z \in \mathbf{S}_\tau$ , where  $c > 0$  is a positive constant that depends only on  $\tau$ .

Next, from Theorem 12.2 (ii), we have  $N^{-1} \text{Tr} G(z) - m(z) = O_{\prec}((N\eta)^{-1})$  for  $z \in \mathbf{S}_\tau$ , which is not enough to establish the rigidity of the extreme eigenvalues. However, analogously to Proposition 9.1, our proof of Theorem 12.2 (ii) in fact yields a stronger result. Indeed, Lemma 12.8 implies that

$$\left| \frac{1}{N} \text{Tr} G(z) - m(z) \right| \prec \frac{N^{-c}}{N\eta} + \Psi^2 \prec \frac{N^{-c}}{N\eta}$$

for  $z \in \mathbf{S}_\tau$ . We may now repeat the argument starting at (10.5), with trivial modifications, to deduce that  $[\lambda_1(W + A) - L_+]_+ \prec N^{-2/3}$ . This concludes the proof of Theorem 12.4.

**12.5. Proof of Theorem 12.5.** Analogously to the proof of Theorem 3.18, the proof is a routine application of the Green function comparison method near the edge [20, Section 6]. The key technical inputs are Theorems 12.2 and 12.4. Note that Theorem 12.2 is applicable since the Green function comparison argument only involves  $z$  satisfying  $|\Psi(z)| \leq N^{-1/3-c}$  with some small constant  $c > 0$ . Hence the assumption (b) from Theorem 12.2 is satisfied.

## A. Properties of $\varrho$ and Stability of (2.11)

This appendix is devoted the proofs of the basic properties of  $\varrho$  and the stability of (2.11) in the sense of Definition 5.4. In this appendix we abbreviate  $r_i := \phi\pi(\{s_i\})$ .

**A.1. Properties of  $\varrho$  and proof of Lemmas 2.4–2.6.** In this subsection we establish the basic properties of  $\varrho$  and prove Lemmas 2.4–2.6.

PROOF OF LEMMA 2.4. We have

$$f'(x) = \frac{1}{x^2} - \sum_{i=1}^n \frac{r_i}{(x + s_i^{-1})^2}, \quad f''(x) = \frac{-2}{x^3} + 2 \sum_{i=1}^n \frac{r_i}{(x + s_i^{-1})^3}. \quad (\text{A.1})$$

From (A.1) we find that  $(x^3 f''(x))' > 0$  on  $I_0 \cup \dots \cup I_n$ . Therefore for  $i = 2, \dots, n$  there is at most one point  $x \in I_i$  such that  $f''(x) = 0$ . We conclude that  $I_i$  has at most two critical points of  $f$ . Using the boundary conditions of  $f$  on  $\partial I_i$ , we conclude the proof of  $|\mathcal{C} \cap I_i| \in \{0, 2\}$  for  $i = 2, \dots, n$ .

From (A.1) we also find that  $(x^2 f'(x))' < 0$  for  $x \in I_1$ . We conclude that there exists at most one point  $x \in I_1$  such that  $f'(x) = 0$ . Using the boundary conditions of  $f'$  on  $\partial I_1$ , we deduce that  $|\mathcal{C} \cap I_1| = 1$ .

Finally, if  $\phi \neq 1$  we have  $f(x) = (\phi - 1)x^{-1} + O(x^{-2})$  as  $x \rightarrow \infty$ . From the boundary conditions of  $f$  on  $\partial I_0$  we therefore deduce that  $|\mathcal{C} \cap I_0| = 1$ . Moreover, if  $\phi = 1$  we find from (A.1) that  $f'(x) \neq 0$  for all  $x \in I_0 \setminus \{\infty\}$ . This concludes the proof.  $\square$

By multiplying both sides of the equation  $z = f(m)$  in (2.11) with the product of all denominators on the right-hand side of (2.12), we find that  $z = f(m)$  may be also written as  $P_z(m) = 0$ , where  $P_z$  is a polynomial of degree  $n + 1$ , whose coefficients are affine linear functions of  $z$ . (Here we used that all  $s_i$  are distinct and that  $m + s_i^{-1} \neq 0$ .) This polynomial characterization of  $m$  is useful for the proofs of Lemmas 2.5 and 2.6.

From Lemma 4.10 we find that the measure  $\varrho$  has a bounded density in  $[\tau, \infty)$  for any fixed  $\tau > 0$ . Hence, we may extend the definition of  $m$  down to the real axis by setting  $m(E) := \lim_{\eta \downarrow 0} m(E + i\eta)$ .

PROOF OF LEMMA 2.5. For  $i = 0, \dots, n$  define the subset  $J_i := \{x \in I_i : f'(x) > 0\}$ . The graph of  $f$  restricted to  $J_0 \cup \dots \cup J_n$  is depicted in red in Figures 2.1 and 2.2. From Lemma 2.4 we deduce that if  $i = 2, \dots, n$  then  $J_i \neq \emptyset$  if and only if  $I_i$  contains two distinct critical points of  $f$ , in which case  $J_i$  is an interval. Moreover, we always have  $J_1 \neq \emptyset \neq J_0$ .

Next, we observe that for any  $0 \leq i < j \leq n$  we have  $f(J_i) \cap f(J_j) = \emptyset$ . Indeed, if there were  $E \in f(J_i) \cap f(J_j)$  then we would have  $|\{x : f(x) = E\}| > n + 1$  (see Figure 2.1). Since  $f(x) = E$  is equivalent to  $P_E(x) = 0$  and  $P_E$  has degree  $n + 1$ , we arrive at the desired contradiction. We conclude that the sets  $f(J_i)$ ,  $0 \leq i \leq n$ , may be strictly ordered.

The claim  $a_1 \geq a_2 \geq \dots \geq a_{2p}$  may now be reformulated as

$$f(J_i) > f(J_j) > f(J_0) \quad \text{if } 1 \leq i < j \leq n. \quad (\text{A.2})$$

In order to show (A.2), we use a continuity argument in  $\phi$ . Let  $t \in (0, 1]$  and denote by  $f^t$  the right-hand side of (2.12) with  $\phi$  is replaced by  $t\phi$ . We use the superscript  $t$  to denote quantities defined using  $f^t$  instead of  $f = f^1$ . Using directly the definition of  $f^t$ , it is easy to check that (A.2) holds for small enough  $t > 0$ .

We focus on the first inequality of (A.2); the proof of the second one is similar. We have to show that  $f(J_i) > f(J_j)$  whenever  $1 \leq i < j \leq n$  and  $J_i \neq \emptyset \neq J_j$ . First, we claim that for  $1 \leq i \leq n$  we have

$$J_i \neq \emptyset \implies J_i^t \neq \emptyset \text{ for all } t \in (0, 1]. \quad (\text{A.3})$$

To prove (A.3), we remark that if  $2 \leq i \leq n$  then  $J_i^t \neq \emptyset$  is equivalent to  $I_i$  containing two distinct critical points. Moreover,  $\partial_t \partial_x f^t(x) < 0$  in  $I_2 \cup \dots \cup I_n$ , from which we deduce that the number of distinct critical points in each  $I_i$ ,  $i = 2, \dots, n$ , does not decrease as  $t$  decreases. Recalling that  $I_1^t \neq \emptyset$ , we deduce (A.3).

Next, suppose that there exist  $1 \leq i < j \leq n$  satisfying  $J_i \neq \emptyset \neq J_j$  and  $f(J_i) < f(J_j)$ . From (A.3) we deduce that  $J_i^t \neq \emptyset \neq J_j^t$  for all  $t \in (0, 1]$ . Moreover, by a simple continuity argument and the fact that for each  $t$  we have either  $f^t(J_i^t) < f^t(J_j^t)$  or  $f^t(J_i^t) > f^t(J_j^t)$ , we conclude that  $f^t(J_i^t) < f^t(J_j^t)$  for all  $t \in (0, 1]$ . As explained after (A.2), this is impossible for small enough  $t > 0$ . This concludes the proof of (A.2), and hence also of the first sentence of Lemma 2.5.

Next, we prove the second sentence of Lemma 2.5. Suppose that  $\phi \neq 1$  and that  $E \neq 0$  lies in the (open) set  $\bigcup_{i=0}^n f(J_i) = f(\bigcup_{i=0}^n J_i)$ . Then the set  $\{x : f(x) = E\} = \{x : P_E(x) = 0\} \subset \mathbb{R}$  has  $n+1$  points. Recall that  $m(E)$  is a solution of  $P_E(m(E)) = 0$ . Since  $n+1 = \deg P_E$ , we deduce that  $m(E)$  is real. Since  $m$  is the Stieltjes transform of  $\varrho$ , we conclude that  $E \notin \text{supp } \varrho$  and that  $m'(E) > 0$ . This means that  $m|_{f(J_i)}$  is increasing, and is therefore the inverse function of  $f|_{J_i}$ . We conclude that  $f$  is a bijection from  $\bigcup_{i=0}^n J_i$  onto  $\bigcup_{i=0}^n f(J_i)$  with inverse function  $m$  (here we use the convention that for  $\phi \leq 1$  we have  $m(0) = \infty$  and  $f(\infty) = 0$ ). By definition, we have  $a_k = f(x_k)$  and the inverse relation  $x_k = m(a_k)$  follows by a simple limiting argument from  $\mathbb{C}_+$  to  $\mathbb{R}$ . The case  $\phi = 1$  is handled similarly; we omit the minor modifications.

Finally, in order to prove the third sentence of Lemma 2.5 we have to show that  $a_{2p} \geq 0$  and  $a_1 \leq C$ . It is easy to see that if  $a_{2p} < 0$  then there is an  $E < 0$  with  $|\{x : f(x) = E\}| > n+1$ , which is impossible since  $\{x : f(x) = E\} = \{x : P_E(x) = 0\}$  and  $\deg P_E = n+1$ . Moreover, the estimate  $a_1 \leq C$  is easy to deduce from the definition of  $f$  and the estimates  $\phi \leq \tau^{-1}$  and  $s_1^{-1} \geq \tau$ .  $\square$

PROOF OF LEMMA 2.6. In the proof of Lemma 2.5 we showed that if  $E \in (0, \infty) \cap \bigcup_{i=0}^n f(J_i)$  then  $E \notin \text{supp } \varrho$ . It therefore suffices to show that if  $E \in \bigcup_{k=1}^p (a_{2k}, a_{2k-1})$  then  $\text{Im } m(E) > 0$ . In that case the set of real preimages  $\{x : f(x) = E\} = \{x : P_E(x) = 0\}$  has  $n-1$  points (see Figure 2.1). Since  $P_E$  has degree  $n+1$  and real coefficients, we conclude that  $P_E$  has a root with positive imaginary part. By the uniqueness of the root of  $P_{E+i\eta}$  in  $\mathbb{C}_+$  (see (2.11)) and the continuity of the roots of  $P_{E+i\eta}$  in  $\eta$ , we therefore conclude by taking  $\eta \downarrow 0$  that  $\text{Im } m(E) > 0$ .  $\square$

For the following counting argument, in order to avoid extraneous complications we assume that  $\sigma_i > 0$  for all  $i$ . Recall the definition of  $N_k$  from (3.16).

LEMMA A.1. *Suppose that  $\sigma_j > 0$  for all  $j$ . Then  $N\varrho(\{0\}) = (N - M)_+$  and*

$$\sum_{k=1}^p N_k = N - (N - M)_+ = N \wedge M. \quad (\text{A.4})$$

Moreover, for  $k = 1, \dots, p-1$  we have

$$N_k = \sum_{i \in \mathcal{I}_M} \mathbf{1}(x_{2k} \leq -\sigma_i^{-1} \leq x_{2k-1}). \quad (\text{A.5})$$

PROOF. Suppose first that  $N > M$ . Then  $x_{2p} < 0$  and  $a_{2p} > 0$  (see Figure 2.1). Let  $1 \leq k \leq p$  and define  $\gamma$  as the positively oriented rectangular contour joining the points  $a_{2k-1} \pm i$  and  $a_{2k} \pm i$ . Then we have  $N_k = -\frac{N}{2\pi i} \oint_{\gamma} m(z) dz$ . We now do a change of coordinates by writing  $w = m(z)$ . Define

$\tilde{\gamma}$  as the positively oriented rectangular contour joining the points  $x_{2k-1} \pm i$  and  $x_{2k} \pm i$ . Since the imaginary parts of  $w$  and  $z$  have the same sign, we conclude using Cauchy's theorem that

$$\begin{aligned} N_k &= -\frac{N}{2\pi i} \oint_{\gamma} m(z) dz = -\frac{N}{2\pi i} \oint_{f(\tilde{\gamma})} m(z) dz \\ &= -\frac{N}{2\pi i} \oint_{\tilde{\gamma}} w f'(w) dw = \sum_{i=1}^n \mathbf{1}(x_{2k} \leq -s_i^{-1} \leq x_{2k-1}) \pi(\{s_i\}) M, \end{aligned}$$

where in the last step we used the definition of  $f$  from (2.12). Hence (A.5) follows. Note that we proved (A.5) also for  $k = p$ . Now (A.4) follows easily using  $\sigma_i > 0$  for all  $i$ . Since  $\text{supp } \varrho \subset \{0\} \cup [a_{2p}, a_{2p-1}] \cup \dots \cup [a_2, a_1]$  (recall Lemma 2.6), we get  $N = N_{\varrho}(\{0\}) + \sum_{k=1}^p N_k$ . Therefore summing over  $k = 1, \dots, p$  in (A.5) yields  $N_{\varrho}(\{0\}) = N - M$ , as desired. This concludes the proof in the case  $N > M$ .

Next, suppose that  $N \leq M$ . Then we deduce directly from (2.12) that  $\varrho(\{0\}) = 0$  (see Figure 2.2). Hence, (A.4) follows. Finally, exactly as above, we get (A.5) for  $i = 1, \dots, p-1$ . This concludes the proof.  $\square$

**A.2. Stability near a regular edge.** The rest of this appendix is devoted to the proofs of the assumption (3.20) and the stability of (2.11) in the sense of Definition 5.4. In fact, we shall prove the following stronger notation of stability, which will be needed to establish eigenvalue rigidity. Recall the definition of  $\kappa$  from (10.1).

**DEFINITION A.2 (STRONG STABILITY OF (2.11) ON  $\mathbf{S}$ ).** *We say that (2.11) is strongly stable on  $\mathbf{S}$  if Definition 5.4 holds with (5.20) replaced by*

$$|u - m| \leq \frac{C\delta}{\sqrt{\kappa + \eta} + \sqrt{\delta}}. \quad (\text{A.6})$$

That the notion of stability from Definition A.2 is stronger than that of Definition 5.4 follows from the estimate

$$\text{Im } m \asymp \begin{cases} \sqrt{\kappa + \eta} & \text{if } E \in \text{supp } \varrho \\ \frac{\eta}{\sqrt{\kappa + \eta}} & \text{if } E \notin \text{supp } \varrho \end{cases} \quad (\text{A.7})$$

for  $z \in \mathbf{S}$ , which we shall establish under the regularity assumptions of Definition 2.7. Under the same regularity assumptions, we shall also prove

$$\min_i |m + s_i^{-1}| \geq c \quad (\text{A.8})$$

for  $z \in \mathbf{S}$  and some constant  $c > 0$  depending only on  $\tau$ . Using Lemma 4.10, we find that (A.8) implies  $|1 + m\sigma_i| \geq c$ , which is the estimate from (3.20) (after a possible renaming of the constant  $\tau$ ).

In this subsection we deal with a regular edge; the case of a regular bulk component is dealt with in the next subsection. We begin with basic estimates on the behaviour of  $f$  near a critical point.

**LEMMA A.3.** *Fix  $\tau > 0$ . Suppose that the edge  $k$  is regular in the sense of Definition 2.7 (i). Then there exists a  $\tau' > 0$ , depending only on  $\tau$ , such that*

$$|x_k| \asymp 1, \quad |f''(x_k)| \asymp 1, \quad |f'''(\zeta)| \leq C \quad \text{for } |\zeta - x_k| \leq \tau'. \quad (\text{A.9})$$

PROOF. From Lemma 2.5 we get  $x_k = m(a_k)$ . Since by assumption  $a_k \in [\tau, C]$  for some constant  $C$ , we find from Lemma 4.10 that  $|x_k| \leq C$ . The lower bound  $|x_i| \geq c$  follows from  $f'(x_k) = 0$  and (A.1), using that  $|s_i| \leq C$ . This proves the first estimate of (A.9).

Next, from  $f'(x_k)$  and (A.1) we find

$$f''(x_k) = -\frac{2}{x_k} \sum_i \frac{s_i^{-1} r_i}{(x_k + s_i^{-1})^3}. \quad (\text{A.10})$$

Using the first estimate of (A.9) and (2.14) we find from (A.10) that  $|f''(x_k)| \leq C$ . Moreover for  $k \in \{1, 2p\}$  (i.e.  $x_k \in I_0 \cup I_1$ ) all terms in (A.10) have the same sign, and it is easy to deduce that  $|f''(x_k)| \geq c$ .

Next, we prove the lower bound  $|f''(x_k)| \geq c$  for  $k = 2, \dots, 2p-1$ . Suppose for definiteness that  $k$  is odd. (The case of even  $k$  is handled in exactly the same way.) For  $x \in [x_k, x_{k-1}]$  we have  $f'(x) \geq 0$  and

$$f'(x) = \int_{x_k}^x (-y^{-3})(-y^3 f''(y)) dy \leq \int_{x_k}^x (-y^{-3})(-x_k^3 f''(x_k)) dy \leq C f''(x_k)(x_{k-1} - x_k),$$

where in the second step we used that  $(-x^3 f''(x))' < 0$  as follows from (A.1), and in the last step the estimate  $|x_k| \asymp |x_{k-1}| \asymp 1$ . We therefore find

$$a_{k-1} - a_k = \int_{x_k}^{x_{k-1}} f'(x) dx \leq C f''(x_k)(x_{k-1} - x_k)^2.$$

Since  $a_{k-1} - a_k \geq \tau$  by the assumption in Definition 2.7 (i) and  $x_{k-1} - x_k \leq |x_{k-1}| + |x_k| \leq C$ , we deduce the lower bound  $f''(x_k) \geq c$ .

Finally, the last estimate of (A.9) follows easily from the assumption (2.14) and by differentiating (A.1).  $\square$

We record the following easy consequence of (A.9). Recalling that  $f(x_k) = a_k$  and  $f'(x_k) = 0$ , we find, using (A.9) and the fact that  $f$  is continuous in a neighbourhood of  $x_k$ , that after choosing  $\tau' > 0$  small enough (depending on  $\tau$ ) we have

$$z - a_k = \frac{f''(x_k)}{2}(m(z) - x_k)^2 + O(|m(z) - x_k|^3) \quad \text{for } |z - a_k| \leq \tau'. \quad (\text{A.11})$$

LEMMA A.4. Fix  $\tau > 0$ . Suppose that the edge  $k$  is regular in the sense of Definition 2.7 (i). Then there exists a constant  $\tau' > 0$ , depending only on  $\tau$ , such that for  $z \in \mathbf{D}_k^\epsilon$  we have (A.7) and (A.8).

PROOF. We first prove (A.7). From (A.11) we find that there exists a  $\tau' > 0$  such that for  $|E - a_k| \leq \tau'$  we have

$$E - a_k = \frac{f''(x_k)}{2}(m(E) - m(a_k))^2(1 + O(|m(E) - m(a_k)|)). \quad (\text{A.12})$$

From (A.9) we find that  $|f''(x_k)| \geq c$ . For definiteness, suppose that  $k$  is odd. Then  $\text{Im } m(E) = 0$  for  $E \geq a_k$  (i.e.  $E \notin \text{supp } \varrho$ ). From (A.12) we therefore deduce that for  $E \leq a_k$  we have  $\text{Im } m(E) \asymp \sqrt{|E - a_k|}$ . We conclude that for  $|E - a_k| \leq \tau'$  we have the square root behaviour  $\text{Im } m(E) \asymp \sqrt{\kappa(E)} \mathbf{1}(E \in \text{supp } \varrho)$  for the density of  $\varrho$ . Now (A.7) easily follows from the definition of the Stieltjes transform.

Finally, (A.8) follows from the assumption (2.14) and (A.11).  $\square$

LEMMA A.5. Fix  $\tau > 0$ . Suppose that the edge  $k$  is regular in the sense of Definition 2.7 (i). Then there exists a constant  $\tau'$ , depending only on  $\tau$ , such that (2.11) is strongly stable on  $\mathbf{D}_k^e$  in the sense of Definition A.2.

PROOF. We take over the notation of Definition 5.4. Abbreviate  $w(z) := f(u(z))$ , so that  $|w(z) - z| \leq \delta(z)$ .

Suppose first that  $\text{Im } z \geq \tau'$  for some constant  $\tau' > 0$ . Then, from the assumption  $\delta(z) \leq (\log N)^{-1}$  we get  $\text{Im } w(z) \geq \tau'/2$ , and therefore from the uniqueness of (2.11) we find  $u(z) = m(w(z))$ . Hence,

$$|u(z) - m(z)| = |m(w(z)) - m(z)| \leq \int_{w(z)}^z |m'(\zeta)| |\mathrm{d}\zeta| \leq 2(\tau')^{-2} \delta(z),$$

where we used the trivial bound  $|m'(\zeta)| \leq (\text{Im } \zeta)^{-2}$ . This yields (A.6) at  $z$ .

What remains therefore is the case  $|z - a_k| \leq 2\tau'$ , which we assume for the rest of the proof. We drop the arguments  $z$  and write the equation  $f(u) - f(m) = w - z$  as

$$\alpha (u - m)^2 + \beta (u - m) = u m (w - z), \quad (\text{A.13})$$

where

$$\alpha := - \sum_i \frac{m s_i^{-1} r_i}{(m + s_i^{-1})^2 (u + s_i^{-1})}, \quad \beta := 1 - \sum_i \frac{m^2 r_i}{(m + s_i^{-1})^2}. \quad (\text{A.14})$$

Note that  $\alpha$  and  $\beta$  depend on  $z$ , and  $\alpha$  also depends on  $u$ . We suppose that

$$|z - a_k| \leq 2\tau', \quad |u - m| \leq (\log N)^{-1/3}. \quad (\text{A.15})$$

Then we claim that for small enough  $\tau'$  we have

$$|\alpha| \asymp 1, \quad |\beta| \asymp \sqrt{\kappa + \eta}. \quad (\text{A.16})$$

In order to prove (A.16), we note that, by Lemma A.3, the statement of (A.11) holds under the assumption (A.15) provided  $\tau' > 0$  is chosen small enough. Using (A.15), (A.8) (see Lemma A.3), and (A.10) we get

$$\alpha = - \sum_i \frac{x_k s_i^{-1} r_i}{(x_k + s_i^{-1})^3} + O(|m - x_k| + (\log N)^{-1}) = \frac{x_k^2}{2} f''(x_k) + O(|m - x_k| + (\log N)^{-1}).$$

Using (A.9) and (A.11) we conclude that for small enough  $\tau' > 0$  we have  $|\alpha| \asymp 1$ . This concludes the proof of the first estimate of (A.16).

In order to prove the second estimate of (A.16), we note that  $\beta = m^2 f'(m)$ , so that for small enough  $\tau' > 0$  we get from (A.11) and Lemma A.3 that

$$m^{-2} \beta = \int_{x_k}^m f''(\zeta) \mathrm{d}\zeta = \int_{x_k}^m (f''(x_k) + O(|m - x_k|)) \mathrm{d}\zeta = (m - x_k) f''(x_k) + O(|m - x_k|^2).$$

Using (A.11) and Lemmas A.3 and (4.10), we conclude for small enough  $\tau' > 0$  that

$$|\beta| \asymp |m - x_k| \asymp \sqrt{|z - a_k|} \asymp \sqrt{\kappa + \eta},$$



This concludes the proof of the second estimate of (A.16).

Using (4.21) and the estimate  $|w - z| \leq \delta$ , we write (A.13) as  $\alpha(u - m)^2 + \beta(u - m) = O(\delta)$ . We now proceed exactly as in the proof of [7, Lemma 4.5], by solving the quadratic equation for  $u - m$  explicitly. We select the correct solution by a continuity argument using that (A.6) holds by assumption at  $z + iN^{-5}$ . The second assumption of (A.15) is obtained by continuity from the estimate on  $|u - m|$  at the neighbouring value  $z + iN^{-5}$ . We refer to [7, Lemma 4.5] for the full details of the argument. (Note that the fact  $\alpha$  depends on  $u$  does not change this argument. This dependence may be easily dealt with using a simple fixed point argument whose details we omit.) This concludes the proof.  $\square$

**A.3. Stability in a regular bulk component.** In this subsection we establish the basic estimates (A.7) and (A.8) as well as the stability of (2.11) in a regular bulk component  $k$ , i.e. for  $z \in \mathbf{D}_k^b$ .

LEMMA A.6. *Fix  $\tau, \tau' > 0$  and suppose that the bulk component  $k = 1, \dots, p$  is regular in the sense of Definition 2.7 (ii). Then for  $z \in \mathbf{D}_k^b$  we have (A.7) and (A.8). Moreover, (2.11) is strongly stable on  $\mathbf{D}_k^b$  in the sense of Definition A.2.*

PROOF. The estimates (A.7) and (A.8) follow trivially from the assumption in Definition 2.7 (ii), using  $\kappa \leq C$  as follows from Lemma 2.5.

In order to show the strong stability of (2.11), we proceed as in the proof of Lemma A.5. Mimicking the argument there, we find that it suffices to prove that  $|\alpha| \leq C$  and  $|\beta| \asymp 1$  for  $z \in \mathbf{D}_k^b$ , where  $\alpha$  and  $\beta$  were defined in (A.14) and  $u$  satisfies  $|u - m| \leq (\log N)^{-1/3}$ .

Note that Definition 2.7 (ii) immediately implies that  $\text{Im } m \geq c$  for some constant  $c > 0$ . The upper bound  $|\alpha| + |\beta| \leq C$  easily follows from the definition (A.14) combined with Lemma 4.10. What remains is the proof of the lower bound  $|\beta| \geq c$ . To that end, we take the imaginary part of (2.11) to get

$$\sum_i \frac{r_i |m|^2}{|m + s_i^{-1}|^2} = 1 - \frac{|m|^2 \eta}{\text{Im } m}. \quad (\text{A.17})$$

Using that  $\text{Im } m \geq c$ , a simple analysis of the arguments of the expressions  $(m + s_i^{-1})^2$  on the left-hand side of (A.17) yields

$$\left| \sum_i \frac{r_i m^2}{(m + s_i^{-1})^2} \right| \leq \sum_i \frac{r_i |m|^2}{|m + s_i^{-1}|^2} - c \leq 1 - c,$$

where in the last step we used (A.17). Recalling the definition of  $\beta$  from (A.14), we conclude that  $|\beta| \geq c$ . This concludes the proof.  $\square$

REMARK A.7. The bulk regularity condition from Definition 2.7 (ii) is stable under perturbation of  $\pi$ . To see this, define the shifted empirical density  $\pi_t := M^{-1} \sum_{i=1}^M \delta_{\sigma_i + t}$ , and the associated Stieltjes transform  $m_t(E)$  and function  $f_t(x)$ . Differentiating  $f_t(m_t(E)) = E$  in  $t$  yields  $(\partial_t m)_t(E) = -(\partial_t f)_t(m_t(E)) / (\partial_x f)_t(m_t(E))$ . Using that

$$(\partial_t f)_0(m(E)) = \frac{1}{N} \sum_{i=1}^M \frac{\sigma_i^{-2}}{(m(E) + \sigma_i^{-1})^2} = O(1)$$

by (A.8), and  $(\partial_x f)_t(m_t(E)) = m(E)^{-2}\beta(E)$ , we find from  $|\beta(E)| \asymp |m(E)| \asymp 1$  that  $(\partial_t m)_0(E) = O(1)$  for  $E \in [a_{2k} + \tau', a_{2k-1} - \tau']$ . A simple extension of this argument shows that if Definition 2.7 (ii) holds for  $\pi$  then it holds for all  $\pi_t$  with  $t$  in some ball of fixed radius around zero.

**A.4. Stability outside of the spectrum.** In this subsection we establish the basic estimates (A.7) and (A.8) as well as the stability of (2.11) outside of the spectrum, i.e. for  $z \in \mathbf{D}^o$ .

LEMMA A.8. *Fix  $\tau, \tau' > 0$ . Then for  $z \in \mathbf{D}^o$  we have (A.7) and (A.8). Moreover, (2.11) is strongly stable on  $\mathbf{D}^o$  in the sense of Definition A.2.*

PROOF. Since  $m$  is the Stieltjes transform of a measure  $\varrho$  with bounded density (see Lemma 4.10), we find that

$$\operatorname{Im} m \asymp \eta \tag{A.18}$$

on  $\mathbf{D}^o$ . This proves (A.7).

We first establish (A.8) for  $z = E \in \mathbb{R}$  satisfying  $\operatorname{dist}(E, \operatorname{supp} \varrho) \geq \tau'$ . For definiteness, let  $z = E \in [a_{2l}, a_{2l+1}]$  for some  $l$ , and let  $i = 2, \dots, n$  satisfy  $\{a_{2l}, a_{2l+1}\} = f(\partial J_i)$ , where we recall the set  $J_i := \{x \in I_i : f'(x) > 0\}$  from the proof of Lemma 2.5. Then there exists an  $x \in J_i$  such that  $E = f(x)$ . Moreover, from (A.1) we get  $f'(y) \leq y^{-2}$  for  $y \in \mathbb{R}$ , so that for  $y \in J_i$  we have  $0 \leq f'(y) \leq C$  by (A.9). Since  $\operatorname{dist}(E, \{a_{2l}, a_{2l+1}\}) \geq \tau'$ , we conclude that  $\operatorname{dist}(x, \partial J_i) \geq c$ , which implies  $\operatorname{dist}(x, \partial I_i) \geq c$ . The cases  $E > a_1$  and  $E < a_{2p}$  are handled similarly. This concludes the proof of (A.8) for  $z = E \in \mathbb{R}$  satisfying  $\operatorname{dist}(E, \operatorname{supp} \varrho) \geq \tau'$ .

Next, we note that if  $\eta \geq \varepsilon$  for some fixed  $\varepsilon > 0$  then (A.8) is trivially true by (A.18). On the other hand, if  $\operatorname{Im} m \leq \varepsilon$  then we get

$$\min_i |m(E + i\eta) + s_i^{-1}| \geq \min_i |m(E) + s_i^{-1}| - C\varepsilon \geq c - C\varepsilon \geq c/2$$

for  $\varepsilon$  small enough, where we used the case  $z = E \in \mathbb{R}$  satisfying  $\operatorname{dist}(E, \operatorname{supp} \varrho) \geq \tau'$  established above and the estimate  $|m'(z)| = \left| \int \frac{1}{(x-z)^2} \varrho(dx) \right| \leq C$  (since  $|x-z| \geq \tau'$  for  $z \in \mathbf{D}^o$  and  $x \in \operatorname{supp} \varrho$ ). This concludes the proof of (A.8) for  $z \in \mathbf{D}^o$ .

As in the proof of Lemma A.6, to prove the strong stability of (2.11) on  $\mathbf{D}^o$ , it suffices to show that  $|\alpha| \leq C$  and  $|\beta| \asymp 1$  for  $z \in \mathbf{D}_k^e$ , where  $\alpha$  and  $\beta$  were defined in (A.14) and  $u$  satisfies  $|u - m| \leq (\log N)^{-1/3}$ . The upper bound  $|\alpha| + |\beta| \leq C$  follows immediately from (A.8), and the lower bound  $|\beta| \geq c$  follows from the definition (A.14) combined with (A.17), (A.18), and Lemma 4.10. This concludes the proof.  $\square$

## Acknowledgements

A.K. was partially supported by Swiss National Science Foundation grant 144662 and the SwissMAP NCCR grant. J.Y. was partially supported by NSF Grant DMS-1207961. We are very grateful to the Institute for Advanced Study, Thomas Spencer, and Horng-Tzer Yau for their kind hospitality during the academic year 2013-2014. We also thank the Institute for Mathematical Research (FIM) at ETH Zürich for its generous support of J.Y.'s visit in the summer of 2014. We are indebted to Jamal Najim for stimulating discussions.

## References

- [1] Z. Bai and J. W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Prob. **26** (1998), 316–345.
- [2] ———, *Exact separation of eigenvalues of large dimensional sample covariance matrices*, Ann. Prob. **27** (1999), 1536–1555.
- [3] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, Springer, 2010.
- [4] Z. Bao, L. Erdős, and K. Schnelli, *Local law of addition of random matrices on optimal scale*, Preprint arXiv:1509.07080.
- [5] Z. Bao, G. Pan, and W. Zhou, *Universality for the largest eigenvalue of a class of sample covariance matrices*, Preprint arXiv:1304.5690v4.
- [6] F. Benaych-Georges, *Local single ring theorem*, Preprint arXiv:1501.07840.
- [7] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Isotropic local laws for sample covariance and generalized Wigner matrices*, Electron. J. Probab **19** (2014), 1–53.
- [8] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin, *On the principal components of sample covariance matrices*, Prob. Theor. Rel. Fields **164** (2016), 459–552.
- [9] M. Capitaine and S. Péché, *Fluctuations at the edges of the spectrum of the full rank deformed GUE*, Prob. Theor. Rel. Fields **165** (2016), 117–161.
- [10] S. Chatterjee, *A generalization of the Lindeberg principle*, Ann. Prob. **34** (2006), 2061–2076.
- [11] N. El Karoui, *Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices*, Ann. Prob. **35** (2007), 663–714.
- [12] L. Erdős, A. Knowles, and H.-T. Yau, *Averaging fluctuations in resolvents of random band matrices*, Ann. H. Poincaré **14** (2013), 1837–1926.
- [13] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. **314** (2012), 587–640.
- [14] ———, *The local semicircle law for a general class of random matrices*, Electron. J. Probab **18** (2013), 1–58.
- [15] ———, *Spectral statistics of Erdős-Rényi graphs I: Local semicircle law*, Ann. Prob. **41** (2013), 2279–2375.
- [16] L. Erdős, B. Schlein, and H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. **287** (2009), 641–655.
- [17] ———, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119.
- [18] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics of random matrices*, Ann. Inst. . Poincaré (B) **48** (2012), 1–46.
- [19] L. Erdős, H.-T. Yau, and J. Yin, *Bulk universality for generalized Wigner matrices*, Prob. Theor. Rel. Fields **154** (2012), 341–407.
- [20] ———, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math **229** (2012), 1435–1515.
- [21] W. Hachem, A. Hardy, and J. Najim, *Large complex correlated Wishart matrices: Fluctuations and asymptotic independence at the edges*, Preprint arXiv:1409.7548.
- [22] W. Hachem, P. Loubaton, J. Najim, and P. Vallet, *On bilinear forms based on the resolvent of large random matrices*, Ann. Inst. H. Poincaré (B) **49** (2013), 36–63.

- [23] K. Johansson, *From Gumbel to Tracy-Widom*, Prob. Theor. Rel. Fields **138** (2007), 75–112.
- [24] V. Kargin, *Subordination for the sum of two random matrices*, Ann. Prob. **43** (2015), 2119–2150.
- [25] A. Knowles and J. Yin, *The isotropic semicircle law and deformation of Wigner matrices*, Comm. Pure Appl. Math. **66** (2013), 1663–1749.
- [26] J. O. Lee and K. Schnelli, *Tracy-Widom distribution for the largest eigenvalues of real sample covariance matrices with general population*, Preprint arXiv:1409.4979.
- [27] J. O. Lee and K. Schnelli, *Edge universality for deformed Wigner matrices*, Rev. Math. Phys. **27** (2015), 1550018.
- [28] A. Lytova and L. Pastur, *Central limit theorem for linear eigenvalue statistics of random matrices with independent entries*, Ann. Prob. **37** (2009), 1778–1840.
- [29] V. A. Marchenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Mat. Sbornik **72** (1967), 457–483.
- [30] A. Onatski, *The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices*, Ann. Appl. Prob. **18** (2008), 470–490.
- [31] L. A. Pastur, *On the spectrum of random matrices*, Teor. Math. Phys. **10** (1972), 67–74.
- [32] N. S. Pillai and J. Yin, *Universality of covariance matrices*, Ann. Appl. Probab. **24** (2014), 935–1001.
- [33] T. Shcherbina, *On universality of bulk local regime of the deformed Gaussian unitary ensemble*, Math. Phys. Anal. Geom. **5** (2009), 396–433.
- [34] J. W. Silverstein and Z. Bai, *On the empirical distribution of eigenvalues of a class of large dimensional random matrices*, J. Multivar. Anal. **54** (1995), 175–192.
- [35] J. W. Silverstein and S.-I. Choi, *Analysis of the limiting spectral distribution of large dimensional random matrices*, J. of Multivariate Anal. **54** (1995), 295–309.
- [36] T. Tao and V. Vu, *Random matrices: Universality of local eigenvalue statistics*, Acta Math. **206** (2011), 1–78.
- [37] C. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys. **159** (1994), 151–174.
- [38] ———, *On orthogonal and symplectic matrix ensembles*, Comm. Math. Phys. **177** (1996), 727–754.
- [39] E.P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. Math. **62** (1955), 548–564.