

# Local Eigenvalue Density for General MANOVA Matrices

László Erdős\* and Brendan Farrell†

July 22, 2013

## Abstract

We consider random  $n \times n$  matrices of the form

$$(XX^* + YY^*)^{-\frac{1}{2}} YY^* (XX^* + YY^*)^{-\frac{1}{2}},$$

where  $X$  and  $Y$  have independent entries with zero mean and variance one. These matrices are the natural generalization of the Gaussian case, which are known as MANOVA matrices and which have joint eigenvalue density given by the third classical ensemble, the Jacobi ensemble. We show that, away from the spectral edge, the eigenvalue density converges to the limiting density of the Jacobi ensemble even on the shortest possible scales of order  $1/n$  (up to  $\log n$  factors). This result is the analogue of the local Wigner semicircle law and the local Marchenko-Pastur law for general MANOVA matrices.

**AMS Subject Classification:** 15B52, 62H86

**Keywords:** MANOVA random matrix, Jacobi ensemble, Local density of eigenvalues.

## 1 Introduction

The three classical families of eigenvalue distributions of Gaussian random matrices are the Hermite, Laguerre and Jacobi ensembles. Hermite ensembles correspond to Wigner matrices,  $X = X^*$ ; Laguerre ensembles describe sample covariance matrices,  $XX^*$ . The random  $n \times n$  matrices yielding the Jacobi ensembles have the form

$$(XX^* + YY^*)^{-\frac{1}{2}} YY^* (XX^* + YY^*)^{-\frac{1}{2}}, \quad (1)$$

---

\*Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg, A-3400, Austria. On leave from Institute of Mathematics, University of Munich, Theresienstr. 39, D-80333 Munich, Germany. Email: lerdos@ist.ac.at. Partially supported by SFB-TR12 Grant of the German Research Council.

†Computing and Mathematical Sciences, MC 305-16, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA.

Email: farrell@cms.caltech.edu. Partially supported by Joel A. Tropp under ONR awards N00014-08-1-0883 and N00014-11-1002 and a Sloan Research Fellowship.

where  $X$  and  $Y$  are  $n \times [bn]$  and  $n \times [an]$  matrices with independent standard Gaussian entries. Here  $a, b > 1$  are fixed parameters of the model,  $n$  is a large number, eventually tending to infinity, and  $[\cdot]$  denotes the integer part. The matrix entries can be real, complex or self-dual quaternions, corresponding to the three symmetry classes, commonly distinguished by the parameter  $\beta = 1, 2, 4$ , respectively. The results in this paper are insensitive to the symmetry class and for simplicity we will consider the complex case ( $\beta = 2$ ).

Matrices of the form (1) are used in statistics for multivariate analysis of variance to determine correlation coefficients (Section 3.3 of [13]). This analysis is called MANOVA, though it has been largely limited to the special case when the entries of (1) are Gaussian.

In this paper we address the case when the entries of  $X$  and  $Y$  in (1) are independent but have general distribution with zero mean and unit variance. In particular, the matrix entries are not required to be identically distributed. We will call such a matrix with general entries a *general MANOVA matrix*.

Similarly to the Wigner and sample covariance matrices, the joint eigenvalue density of (1) is explicitly known only for the Gaussian case. When the entries are standard complex Gaussians, it is given by

$$\text{density}(\lambda_1, \dots, \lambda_n) = C_{a,b,n} \prod_{j=1}^n \lambda_j^{(a-1)n} (1 - \lambda_j)^{(b-1)n} \prod_{1 \leq j < k \leq n} |\lambda_j - \lambda_k|^2, \quad \lambda_j \in [0, 1], \quad (2)$$

where  $C_{a,b,n}$  is a normalizing constant. The density has a similar form with different exponents when the matrix entries are real, or self-dual quaternions, see Section 3.6 of [11]. Equation (2) defines the Jacobi ensemble, where the name refers to the form of the polynomial term in front of the Vandermonde determinant in (2).

The empirical density of the eigenvalues of (1) – equivalently, the one-point correlation function of (2) – converges almost surely, as  $n \rightarrow \infty$ , to the distribution with density given by

$$f_M(x) = (a + b) \frac{\sqrt{(x - \lambda_-)(\lambda_+ - x)}}{2\pi x(1 - x)} \cdot I_{[\lambda_-, \lambda_+]}(x) \quad (3)$$

where

$$\lambda_{\pm} = \left( \sqrt{\frac{a}{a+b} \left(1 - \frac{1}{a+b}\right)} \pm \sqrt{\frac{1}{a+b} \left(1 - \frac{a}{a+b}\right)} \right)^2. \quad (4)$$

The density  $f_M$  was determined by Wachter [20] and is discussed in Section 3.6 of [11]. Note that  $\lambda_{\pm} \in (0, 1)$ , so that  $f_M$  is supported on a compact subinterval of  $(0, 1)$ . We will refer to  $f_M(x)$  as the *limiting distribution* of the eigenvalues of (1) or as the MANOVA distribution.

While the joint eigenvalue density (2) is valid only for the Gaussian case, the limiting empirical density is expected to be correct for general distributions as well, similarly to the universality of the Wigner semicircle law for Wigner matrices or the Marchenko-Pastur (MP) law for sample covariance matrices. Thus, general MANOVA matrices, the Jacobi ensemble and the distribution  $f_M$  constitute a triplet analogous to general Wigner matrices, the Hermite ensemble and the semicircle law or sample covariance matrices, the Laguerre ensemble and the Marchenko-Pastur law.

Universality results have been intensely pursued for the latter two types of matrices, starting from the fundamental work of Wigner [21] and Marchenko-Pastur [12] who identified the corresponding distributions. These first results were on the macroscopic scale; the empirical density on spectral scales containing  $O(n)$  eigenvalues were shown to converge in a weak sense to the limiting law. Recently local versions of these fundamental laws have also been established on the shortest possible scale, containing  $O(\varphi(n))$  eigenvalues, where  $\varphi(n)$  is a factor logarithmic in  $n$ . For Wigner matrices it was achieved first in the bulk [4, 5] then optimally up to the edges [9]. For sample covariance matrices the optimal scale in the bulk was reached in [7], followed by the optimal result up to the edge in [15]. Related results were also obtained in [18, 19, 14].

In this paper we prove the local convergence of the density on the optimal scale for the general MANOVA ensembles in the bulk spectrum. This establishes the analogue of the results [5, 7] for these ensembles. We remark that the convergence even on the largest scale, i.e. the analogue of [21, 12], has not been known before although it would essentially follow from [17] if combined with the recent result in [15]. The main novelty of the current paper is the effective stability analysis of the self-consistent equation for the Stieltjes transform of the density (18).

Precise results on the local density have opened up the route to establish the full universality of local eigenvalue statistics for Wigner and sample covariance matrices, including precise identification of the statistics of consecutive gaps. A new general method based on the Dyson Brownian motion (DBM) was first introduced in [6]. It is applicable to all symmetry classes [7], to very general distributions [9] and to sample covariance matrices [15]. The local semicircle law (or the local MP law) is a basic input in all these works. Local density results have also inspired an alternative route to universality [19, 18] that is applicable for the complex case,  $\beta = 2$ .

In light of these developments for the Wigner and sample covariance matrices, the current work is the first step towards establishing the full universality of eigenvalue statistics for the general MANOVA ensemble.

## 2 Statement of the Main Result

Given two positive constants  $\gamma = (\gamma_1, \gamma_2)$ , we say that a complex random variable  $Z$  is  $\gamma$ -subexponential if it satisfies the following conditions:

$$\begin{cases} \mathbb{E} Z = 0 \\ \mathbb{E} |Z|^2 = 1 \\ \mathbb{P}(|Z| \geq t^{\gamma_1}) \leq \gamma_2 e^{-t} \text{ for all } t > 0. \end{cases} \quad (5)$$

A set of random variables is uniformly  $\gamma$ -subexponential if each random variable is  $\gamma$ -subexponential for a common  $\gamma$ . Assuming that the matrix elements of  $X$  and  $Y$  are independent, uniformly  $\gamma$ -subexponential random variables, we will prove that the empirical distribution of the eigenvalues of (1) in the bulk converges on small scales to (3) as  $n \rightarrow \infty$ .

The main tool for this approach is the Stieltjes transform. The Stieltjes transform of a

real random variable with distribution function  $F$  is a function  $\mathbb{C}^+ \rightarrow \mathbb{C}^+$  defined by

$$m(z) = \int \frac{1}{t-z} dF(t). \quad (6)$$

If the random variable has a density, then we also refer to the Stieltjes transform of the density. The Stieltjes transform of  $f_M$  is

$$m_M(z) = \frac{(2-a-b)z + a - 1 + \sqrt{(a+b)^2 z^2 - (a+b)(2(a+1) - \frac{a}{a+b})z + (a-1)^2}}{2z(1-z)}. \quad (7)$$

This formula is derived in Appendix A.

For self-adjoint matrices, we misuse notation and refer to the function

$$m_A(z) = \frac{1}{n} \text{tr}(A - zI)^{-1}$$

as the Stieltjes transform of the self-adjoint,  $n \times n$  matrix  $A$ . If  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of  $A$ , then we equivalently have

$$m_A(z) = \frac{1}{n} \sum_{k=1}^n \frac{1}{\lambda_k - z},$$

which is the Stieltjes transform of the empirical measure.

Our main result shows that the eigenvalues of the general MANOVA matrix behave close to what is indicated by  $m_M(z)$  and  $f_M$  in the bulk with high probability. To state the result, we must formalize the term *bulk*. Following this definition we state the main theorem, Theorem 2.2.

**Definition 2.1** Let  $\lambda_+$  and  $\lambda_-$  be as given in (4). Define

$$\mathcal{E}_{\kappa, \eta}^{(\lambda)} := \{E + i\eta \in \mathbb{C}^+ : E \in (\lambda_-, \lambda_+) \text{ and } (\lambda_+ - E)(E - \lambda_-) \geq \kappa\}$$

and set  $\mathcal{E}_{\kappa}^{(\lambda)} = \mathcal{E}_{\kappa, 0}^{(\lambda)}$ .

**Theorem 2.2** Fix two real parameters  $a, b > 1$ . Let  $X$  be an  $n \times an$  random matrix and let  $Y$  be an  $n \times bn$  random matrix independent of  $X$ . We assume that both matrices have independent entries satisfying (5) for a common  $\gamma = (\gamma_1, \gamma_2)$ . Let  $m_{n,M}(z)$  be the Stieltjes transform of the general MANOVA matrix

$$(XX^* + YY^*)^{-\frac{1}{2}} YY^* (XX^* + YY^*)^{-\frac{1}{2}}. \quad (8)$$

i) Then for any  $\kappa, \eta > 0$  with  $\eta > \frac{1}{n\kappa^2} (\log n)^{2C \log \log n}$ , we have

$$\mathbb{P} \left( \sup_{z \in \mathcal{E}_{\kappa, \eta}^{(\lambda)}} |m_{n,M}(z) - m_M(z)| > \frac{(\log n)^{C \log \log n}}{\sqrt{\eta n \kappa}} \right) < n^{-c \log \log n} \quad (9)$$

for all  $n \geq n_0$  large enough and for constants  $C, c > 0$ . Here  $n_0, C$  and  $c$  depend only on  $\gamma$ .

ii) Let  $\mathcal{N}_\eta(E)$  denote the number of eigenvalues of (8) contained in  $[E - \frac{\eta}{2}, E + \frac{\eta}{2}]$  and assume  $\eta \geq \frac{1}{n\kappa^2}(\log n)^{3C \log \log n}$ . Then

$$\mathbb{P} \left( \sup_{E \in \mathcal{E}_\kappa^{(\lambda)}} \left| \frac{\mathcal{N}_\eta(E)}{n\eta} - f_M(E) \right| > \frac{(\log n)^{C \log \log n}}{(\eta n \kappa)^{1/4}} \right) \leq n^{-c \log \log n}. \quad (10)$$

We note that the entries of the matrices  $X$  and  $Y$  are not necessarily identically distributed.

Theorem 2.2 shows that the Stieltjes transform of the general MANOVA matrices is close to  $m_M(z)$  when the real part of  $z$ ,  $E = \Re z$ , is away from the spectral edge and the imaginary part  $\eta = \Im z$  is small. In particular,  $\eta$  may be as small as the shortest possible scale  $1/n$ , up to logarithmic corrections. The second part of Theorem 2.2 is an easy consequence of the first one and it asserts that the local density on scale  $1/n$  (modulo logarithmic corrections) is given by the Manova density  $f_M(E)$ . While our analysis is valid down to the optimal scale  $\eta \gtrsim 1/n$ , the error bound of the form  $(n\eta\kappa)^{-1/2}$  is not optimal. The best estimate should scale like  $(n\eta)^{-1}$  and should not blow up near the edge, similarly to the best estimate in the Wigner case [9] and in the sample covariance case [15]. Work to obtain the optimal error bounds is in preparation.

## 2.1 General Approach

The inside and outer matrices of the MANOVA matrix (1) are not independent, which is a requirement for our approach. However, the eigenvalues of (1) are the same as those of

$$(YY^*)^{\frac{1}{2}}(XX^* + YY^*)^{-1}(YY^*)^{\frac{1}{2}} = (I + (YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}})^{-1}.$$

Thus, we begin our work with the matrix  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$ , which we will call the *product matrix* and for which we can use the approach developed in [3, 4, 7, 15] and related papers. After determining the behavior of the product matrix we will return to the matrix (1). Our approach determines an implicit equation for the Stieltjes transform of the limiting empirical eigenvalue distribution of the product matrix and shows that the solution is stable. The implicit equation we obtain is analogous to the quadratic equations that the Stieltjes transforms of the Wigner semicircle law and the MP law satisfy. Then it is shown that with high probability the Stieltjes transform of the empirical distribution nearly satisfies this implicit equation. From stability we conclude how close the empirical distribution is to its limit.

After obtaining results for the matrices  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$ , we return the matrices of our original interest. Note that if  $\mu$  is an eigenvalue of  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$ , then  $\lambda = (1 + \mu)^{-1}$  is an eigenvalue of (1). The eigenvalues of  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$  are non-negative so that the map just given is regular. This allows us to show that if the eigenvalues of  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$  are distributed close to their limiting distribution, then the eigenvalues of (1) are also close to their limiting distribution.

## 2.2 Conventions

We make the following conventions, which will be used without referring to them. The letters  $C$  and  $c$  will denote positive absolute constants that may change from appearance to

appearance. We use the complex number  $z = E + i\eta$  for the spectral parameter, where  $\eta$  will always be positive. The edges of the limiting spectrum  $\lambda_-$  and  $\lambda_+$  are given in (4) and we set

$$\mu_- = \frac{1}{\lambda_+} - 1 \quad \text{and} \quad \mu_+ = \frac{1}{\lambda_-} - 1, \quad (11)$$

to be limiting spectral edges of the product matrix.

Rather than writing  $[an]$  and  $[bn]$  we will write  $an$  and  $bn$ , and it is implicit that we are using the integer part. Most quantities in this paper depend on the parameters  $a$  and  $b$ , but we usually omit this fact in the notation.

### 3 A Product Matrix

We begin by addressing the eigenvalues of the product matrix  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$ . Denote by  $f_p$  the limiting empirical spectral distribution for matrices of the form  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$  with Gaussian entries. This distribution, which will be derived below, is explicitly given by

$$f_{p,a,b}(x) = f_p(x) = \frac{C_{a,b}}{2\pi x} \sqrt{\left(\frac{1}{1+x} - \lambda_-\right) \left(\lambda_+ - \frac{1}{1+x}\right)} \cdot I_{[\lambda_-, \lambda_+]} \left(\frac{1}{1+x}\right), \quad (12)$$

where  $I_{[u,v]}$  is the characteristic function of the interval  $[u, v]$ . We denote this distribution's Stieltjes transform  $m_p$ , and we will use the subscript  $p$  for functions associated with the product matrix. The main theorem of this section, Theorem 3.2, relates the eigenvalues of  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}}$  to the functions  $f_p$  and  $m_p$ . It is followed by a delocalization result, Theorem 3.3. Our main result, Theorem 2.2, will also follow from Theorem 3.2.

The following definition formalizes the bulk for the product matrices; it is the analogous region to  $\mathcal{E}_{\kappa,\eta}^{(\lambda)}$ .

**Definition 3.1** For  $\lambda_{\pm}$  and  $\mu_{\pm}$  as defined in (4) and (11) and for  $\kappa > 0$ , define

$$\mathcal{E}_{\kappa,\eta}^{(\mu)} := \left\{ E + i\eta \in \mathbb{C}^+ : E \in (\mu_-, \mu_+) \quad \text{and} \quad \left(\lambda_+ - \frac{1}{E+1}\right) \left(\frac{1}{E+1} - \lambda_-\right) \geq \kappa \right\}$$

and set  $\mathcal{E}_{\kappa}^{(\mu)} = \mathcal{E}_{\kappa,0}$ .

The following is the main result on product matrices.

**Theorem 3.2** Let  $X$  be an  $n \times bn$ ,  $b > 1$ , random matrix with independent entries satisfying (5) for a uniform  $\gamma$ . Let  $Y$  be an  $n \times an$ ,  $a > 1$ , random matrix independent of  $X$  with independent entries also satisfying (5) for the same  $\gamma$ . Let  $m_{n,p}(z)$  be the Stieltjes transform of  $(YY^*)^{-1/2}XX^*(YY^*)^{-1/2}$ . Fix  $\kappa > 0$  to be a small positive constant.

i) Then for  $\eta > \frac{1}{n\kappa^2}(\log n)^{2C \log \log n}$

$$\mathbb{P} \left( \sup_{z \in \mathcal{E}_{\kappa,\eta}^{(\mu)}} |m_{n,p}(z) - m_p(z)| > \frac{(\log n)^{C \log \log n}}{\sqrt{\eta n \kappa}} \right) < n^{-c \log \log n} \quad (13)$$

for all  $n$  large enough and for constants  $C, c > 0$  depending only on  $\gamma$ .

ii) Let  $\mathcal{N}_\eta(E)$  denote the number of eigenvalues of  $(YY^*)^{-1/2}XX^*(YY^*)^{-1/2}$  in  $[E - \frac{\eta}{2}, E + \frac{\eta}{2}]$ , and assume  $\eta \geq \frac{1}{n\kappa^2}(\log n)^{3C \log \log n}$ . Then

$$\mathbb{P} \left( \sup_{E \in \mathcal{E}_\kappa^{(\mu)}} \left| \frac{\mathcal{N}_\eta(E)}{n\eta} - f_p(E) \right| \geq \frac{(\log n)^{C \log \log n}}{(\eta n \kappa)^{1/4}} \right) \leq n^{-c \log \log n}. \quad (14)$$

**Theorem 3.3** Set  $A := (YY^*)^{-1/2}XX^*(YY^*)^{-1/2}$  and assume that  $X, Y$  satisfy the same conditions as in Theorem 3.2. Then

$$\mathbb{P} \left( \exists v \in \mathbb{C}^n, \|v\|_2 = 1, Av = \mu v, \mu \in \mathcal{E}_\kappa^{(\mu)}, \text{ and } \|v\|_\infty > \frac{(\log n)^{C \log \log n}}{\sqrt{n}} \right) \leq n^{-c \log \log n}.$$

To prove Theorem 3.2 we first determine the implicit equation for  $m_p$ . We recall that the corresponding implicit equation for both the Wigner semicircle law and the Marchenko-Pastur law is a simple algebraic equation. In the case of the product matrix, it turns out that  $m_p$  satisfies an implicit integral equation with good stability properties. In Section 3.1 we show that the Stieltjes transform of the empirical density,  $m_{n,p}$ , approximately satisfies the implicit equation for  $m_p$  and we identify the error term. The error will be controlled in Section 3.4 after several preliminary lemmas in Section 3.2.

### 3.1 Stieltjes Transform of the Product Matrix

The first step of the proof is determining an implicit equation for the Stieltjes transform of the product matrix. While determining this equation, we initially view  $(YY^*)^{-1}$  as a fixed matrix and, adjusting the scaling, set  $T := (\frac{1}{n}YY^*)^{-1}$ .

Random covariance matrices of the form  $\frac{1}{an}YY^*$  have limiting distribution

$$f_{MP,a}(x) = f_{MP}(x) = \frac{a}{2\pi} \sqrt{\frac{\left[ \left[ \left( (1 + a^{-\frac{1}{2}})^2 - x \right) \left( x - (1 - a^{-\frac{1}{2}})^2 \right) \right]_+ \right]}{x^2}}, \quad (15)$$

which is called the Marchenko-Pastur distribution [12]. The local Marchenko-Pastur law was obtained in [7], and the optimal result for the edge was obtained in [15]. The random matrix  $T$  then has limiting distribution

$$\begin{aligned} f_{\text{Inv},a}(x) = f_{\text{Inv}}(x) &= \frac{1}{ax^2} f_{MP} \left( \frac{1}{ax} \right) \\ &= \frac{1}{2\pi x} \sqrt{\left[ \left[ \left( (1 + a^{-\frac{1}{2}})^2 - \frac{1}{ax} \right) \left( \frac{1}{ax} - (1 - a^{-\frac{1}{2}})^2 \right) \right]_+ \right]}. \end{aligned} \quad (16)$$

Note that  $f_{\text{Inv}}$  is supported on  $[a_-, a_+]$ , where  $a_\pm = (a(1 \mp \frac{1}{\sqrt{a}})^2)^{-1}$ .

Here we also address the distribution for  $f_p$ . When the entries of  $X$  and  $Y$  are Gaussian, the limiting eigenvalue distribution of  $(YY^*)^{-1/2}XX^*(YY^*)^{-1/2}$  can be obtained from the MANOVA distribution  $f_M$  using the transformation described in Section 2.1. It is

$$f_{p,a,b}(x) = f_p(x) = \frac{C_{a,b}}{(1+x)^2} \frac{\sqrt{\left( \frac{1}{1+x} - \lambda_- \right) \left( \lambda_+ - \frac{1}{1+x} \right)}}{2\pi \frac{1}{1+x} \left( 1 - \frac{1}{1+x} \right)} \cdot I_{[\lambda_-, \lambda_+]} \left( \frac{1}{1+x} \right)$$

$$= \frac{C_{a,b}}{2\pi x} \sqrt{\left(\frac{1}{1+x} - \lambda_-\right) \left(\lambda_+ - \frac{1}{1+x}\right)} \cdot I_{[\lambda_-, \lambda_+]} \left(\frac{1}{1+x}\right). \quad (17)$$

We use  $m_p(z)$  to denote the Stieltjes transform of  $f_p$ . We remark that no explicit formula is available for  $m_p$ , unlike the case of the semicircle or the Marchenko-Pastur law but it satisfies an integral equation. By Lemma 5.1 of [17] the function  $m_p$  is the unique solution taking values in  $\mathbb{C}^+$  to the implicit equation

$$m(z) = \int \frac{1}{\lambda(1 - \frac{1}{a} - \frac{1}{a}zm(z)) - z} f_{\text{Inv}}(\lambda) d\lambda. \quad (18)$$

For the remainder of Section 3 we will shorten the notation  $m_{n,p}$  to  $m_n$ . Thus,  $m_n(z)$  denotes the Stieltjes transform of the product matrix  $(YY^*)^{-\frac{1}{2}}XX^*(YY^*)^{-\frac{1}{2}} = \frac{1}{n}T^{\frac{1}{2}}XX^*T^{\frac{1}{2}}$ . The goal is to derive a self consistent equation for  $m_n(z)$  that is close to (18). We will then establish the stability of (18) and conclude that  $m_n(z)$  is close to  $m_p(z)$ .

The derivation of the equation for  $m_n(z)$  follows Silverstein and Bai in [17, 16]. Here we set  $N = bn$ . Let  $X$  be an  $n \times N$  matrix, and  $T$  an  $n \times n$  positive definite matrix. Then  $m_n$  is the Stieltjes transform of  $\frac{1}{n}T^{\frac{1}{2}}XX^*T^{\frac{1}{2}}$  and let  $\underline{m}_n$  be the Stieltjes transform of  $\frac{1}{n}X^*TX$ . Since the nonzero spectrum of  $AA^*$  and  $A^*A$  coincide for any matrix  $A$ , we easily get that

$$\underline{m}_n(z) = -\frac{1}{z} \left(1 - \frac{1}{a}\right) + \frac{1}{a}m_n(z). \quad (19)$$

We let  $x_j$  denote the  $j^{\text{th}}$  column of  $X$ . Set

$$r_j := \frac{1}{\sqrt{N}}T^{1/2}x_j, \quad B_n := \frac{1}{N}T^{1/2}XX^*T^{1/2} = \sum_{j=1}^N r_j r_j^*, \quad B_{(j)} := \sum_{i=1, i \neq j}^N r_i r_i^*.$$

We let  $m_{(j)}(z)$  be the Stieltjes transform of  $B_{(j)}$  and set

$$\underline{m}_{(j)}(z) = -\frac{1}{z} \left(1 - \frac{1}{a}\right) + \frac{1}{a}m_{(j)}(z). \quad (20)$$

We will use the following simple formula: if  $A$  is an  $n \times n$  matrix,  $q \in \mathbb{C}^n$  and both  $A$  and  $A + qq^*$  are invertible, then

$$q^*(A + qq^*)^{-1} = \frac{1}{1 + q^*A^{-1}q} q^*A^{-1}. \quad (21)$$

Using the definition of  $B_n$  we obtain the identity

$$I + z(B_n - zI)^{-1} = \sum_{j=1}^N r_j r_j^* (B_n - zI)^{-1}.$$

Using (21), for each  $j$

$$r_j^*(B_n - zI)^{-1} = r_j^*(B_{(j)} - zI + r_j r_j^*)^{-1} = \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} r_j^*(B_{(j)} - zI)^{-1}, \quad (22)$$

so that

$$I + z(B_n - zI)^{-1} = \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} r_j r_j^*(B_{(j)} - zI)^{-1}.$$

We take the trace on each side and divide by  $N$

$$\begin{aligned} \frac{1}{a} + z\frac{1}{a}m_n(z) &= \frac{1}{N} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} \text{tr}[r_j r_j^*(B_{(j)} - zI)^{-1}] \\ &= \frac{1}{N} \sum_{j=1}^N \frac{r_j^*(B_{(j)} - zI)^{-1}r_j}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} \\ &= 1 - \frac{1}{N} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j}, \end{aligned}$$

so that

$$-\frac{1}{z} \left(1 - \frac{1}{a}\right) + \frac{1}{a}m_n(z) = -\frac{1}{z} \frac{1}{N} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j}.$$

Using (19) we have

$$\underline{m}_n(z) = -\frac{1}{z} \frac{1}{N} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j}. \quad (23)$$

With the resolvent identity and identities (22) and (23),

$$\begin{aligned} &(-z\underline{m}_n(z)T - zI)^{-1} - (B_n - zI)^{-1} \\ &= (-z\underline{m}_n(z)T - zI)^{-1} \left( \sum_{j=1}^N r_j r_j^* - (-z\underline{m}_n(z)T) \right) (B_n - zI)^{-1} \\ &= (-z\underline{m}_n(z)T - zI)^{-1} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} r_j r_j^*(B_{(j)} - zI)^{-1} \\ &\quad - (-z\underline{m}_n(z)T - zI)^{-1} \frac{1}{N} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} T (B_n - zI)^{-1} \\ &= \frac{1}{z} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} \\ &\quad \times \left[ (-\underline{m}_n(z)T - I)^{-1} r_j r_j^*(B_{(j)} - zI)^{-1} - \frac{1}{N} (-\underline{m}_n(z)T - I)^{-1} T (B_n - zI)^{-1} \right]. \end{aligned} \quad (24)$$

Taking the trace of (24) and dividing by  $n$  we have

$$\frac{1}{n} \text{tr}(-z\underline{m}_n(z)T - zI)^{-1} - m_n(z) = \frac{a}{zN} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j} d_j(z), \quad (25)$$

where

$$d_j(z) = r_j^*(B_{(j)} - zI)^{-1}(-\underline{m}_n(z)T - I)^{-1}r_j - \frac{1}{an}\text{tr}(B_n - zI)^{-1}(-\underline{m}_n(z)T - I)^{-1}T.$$

We break this into four terms

$$|d_j(z)| \leq |r_j^*(B_{(j)} - zI)^{-1}(\underline{m}_n(z)T + I)^{-1}r_j - r_j^*(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T + I)^{-1}r_j| \quad (26)$$

$$+ |r_j^*(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T + I)^{-1}r_j - \frac{1}{an}\text{tr}(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T + I)^{-1}T| \quad (27)$$

$$+ \left| \frac{1}{an}\text{tr}(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T + I)^{-1}T - \frac{1}{an}\text{tr}(B_{(j)} - zI)^{-1}(\underline{m}_n(z)T + I)^{-1}T \right| \quad (28)$$

$$+ \left| \frac{1}{an}\text{tr}(B_{(j)} - zI)^{-1}(\underline{m}_n(z)T + I)^{-1}T - \frac{1}{an}\text{tr}(B_n - zI)^{-1}(\underline{m}_n(z)T + I)^{-1}T \right|. \quad (29)$$

We will bound  $|d_j(z)|$  in Section 3.4 using the Lemmas developed in Sections 3.2 and 3.3.

### 3.2 Large deviation estimates for exceptional events

In this section we will define certain typical events, denoted by  $\mathcal{X}$ ,  $\mathcal{T}_1$ ,  $\mathcal{T}_2$  and  $\mathcal{B}$ , that will be needed to estimate  $|d_j(z)|$ . Recall that  $x_j$  denotes the  $j^{\text{th}}$  column of  $X$  and let  $X_{(j)}$  denote the matrix obtained by removing the  $j^{\text{th}}$  column of  $X$ . Define the event

$$\mathcal{X} := \left\{ \frac{1}{2} \leq \frac{\|x_j\|_2^2}{n} \leq 2 \text{ and } \left(1 - \frac{1}{\sqrt{b}}\right)^2 \leq \left\| \frac{1}{n} X_{(j)} X_{(j)}^* \right\| < 4b \text{ for all } 1 \leq j \leq bn \right\}. \quad (30)$$

For some small  $c > 0$ , let  $\mathcal{T}_1$  denote the event

$$\mathcal{T}_1 := \left\{ T = \left( \frac{1}{n} Y Y^* \right)^{-1} \text{ is well-defined and } (1-c)a_- \leq T \leq (1+c)a_+ \right\}. \quad (31)$$

Let  $\{t_1, \dots, t_n\}$  denote the eigenvalues of  $T$ . Let  $\{\tau_1, \dots, \tau_n\}$  denote their classical locations given through the limiting density  $f_{\text{Inv}}$  from (16), i.e. they are defined through the formula

$$\int_{-\infty}^{\tau_k} f_{\text{Inv}}(x) dx = \frac{k}{n}, \quad k = 1, \dots, n. \quad (32)$$

Let

$$\mathcal{T}_2 := \left\{ \frac{1}{n} \sum_{k=1}^n |t_k - \tau_k| \leq \frac{1}{n} (\log n)^{C_T \log \log n} \right\}, \quad (33)$$

denote the event that the actual eigenvalues are close to their classical location, where  $C_T > 0$  is a constant independent of  $n$ .

Let  $\mu_1^{(j)}, \dots, \mu_{n-1}^{(j)}$  denote the eigenvalues of  $B_{(j)}$  for  $j = 1, \dots, N$ . For a fixed constant  $K$ , let  $\mathcal{B}$  be the event that

$$\mathcal{B} := \left\{ \sup_{z \in \mathcal{E}(\mu)} \max_{1 \leq j \leq N} \frac{1}{n} \sum_{k=1}^{n-1} \frac{1}{|\mu_k^{(j)} - z|} \leq K(\log n)^2 \right\} \cap \left\{ \sup_{z \in \mathcal{E}(\mu)} \max_{1 \leq j \leq N} \frac{1}{n} \sum_{k=1}^{n-1} \frac{1}{|\mu_k^{(j)} - z|^2} \leq \frac{K}{\eta} \right\}, \quad (34)$$

where

$$\mathcal{E}(\mu) := \left\{ z = E + i\eta \in \mathbb{C} : E \in (\mu_-, \mu_+), \left( \lambda_+ - \frac{1}{1+E} \right) \left( \frac{1}{1+E} - \lambda_- \right) \geq \kappa, \frac{(\log n)^C}{n} \leq \eta \leq 1 \right\}.$$

**Proposition 3.4** *With the notations above, the following estimate holds*

$$\mathbb{P}(\mathcal{X} \cap \mathcal{T}_1 \cap \mathcal{T}_2 \cap \mathcal{B}) \geq 1 - n^{-c \log \log n}. \quad (35)$$

We first prove several lemmas.

**Lemma 3.5** *Let  $\{\bar{\tau}_1, \dots, \bar{\tau}_n\}$  be the classical locations w.r.t. the Marchenko-Pastur law, i.e.*

$$\int_{-\infty}^{\bar{\tau}_k} f_{MP}(x) dx = \frac{k}{n}, \quad k = 1, 2, \dots, n. \quad (36)$$

*Assume that the points  $\{\bar{t}_1, \dots, \bar{t}_n\}$  satisfy*

$$|\bar{t}_k - \bar{\tau}_k| \leq \delta, \quad \min(\bar{t}_1, \dots, \bar{t}_n) \geq \alpha > 0 \quad (37)$$

*for some  $\delta, \alpha > 0$ . Set  $t_k := (a\bar{t}_{n+1-k})^{-1}$ . Then, recalling the definition of  $\tau_k$  from (32), we have*

$$|\tau_k - t_k| \leq \frac{\delta}{\alpha a \left(1 - a^{-\frac{1}{2}}\right)^2}, \quad k = 1, 2, \dots, n.$$

**Proof** The smallest point contained in the support of  $f_{MP}$  is  $(1 - a^{-\frac{1}{2}})^2$ , so given (36) we have

$$\int_{(a\bar{\tau}_k)^{-1}}^{a^{-1}(1-a^{-\frac{1}{2}})^{-2}} \frac{1}{ax^2} f_{MP}\left(\frac{1}{ax}\right) dx = \frac{k}{n}.$$

Considering the relation between  $f_{\text{Inv}}$  and  $f_{MP}$  (see (16)), it follows that  $\tau_k := (a\bar{\tau}_{n+1-k})^{-1}$ . By the assumption (37) we have

$$\left| \frac{1}{a\bar{\tau}_k} - \frac{1}{a\bar{t}_k} \right| \leq \frac{\delta}{\alpha a (1 - a^{-\frac{1}{2}})^2}$$

for all  $k = 1, 2, \dots, n$ . □

The following lemma is a variation of Lemma 4.7 of [5].

**Lemma 3.6** *Let the entries of  $b \in \mathbb{C}^n$  be uniformly  $\gamma$ -subexponential independent random variables satisfying  $\mathbb{E}b_i = 0$  and  $\mathbb{E}|b_i|^2 = 1$  for  $1 \leq i \leq n$ . Let  $S$  be an  $n \times n$  positive definite matrix satisfying  $s_{\min} \leq S \leq s_{\max}$  for  $0 < s_{\min} \leq 1$  and  $1 \leq s_{\max} < \infty$ . Let  $\{v_i\}_{i=1}^m$  be a set of orthonormal vectors in  $\mathbb{C}^n$  and set  $\xi_i = |b^* S v_i|^2$  for  $i = 1, \dots, m$ . If  $\epsilon > 0$  and  $m$  satisfy  $\frac{1}{2}(1 - \epsilon)s_{\min}^2 \sqrt{m} \geq s_{\max}^2 (\log n)^C$  for a constant  $C$  depending only on  $\gamma$ , then*

$$\mathbb{P}\left(\sum_{i=1}^m \xi_i \leq \epsilon s_{\min}^2 m\right) \leq C n^{-\log \log n}.$$

**Proof** We set  $w_i = S v_i$  for  $1 \leq i \leq m$ . As in the proof of Lemma 4.7 in [5], we set

$$Z = \sum_{k,l=1}^n a_{k,l} [b_k \bar{b}_l - \mathbb{E}b_k \bar{b}_l] \quad \text{for} \quad a_{k,l} = \sum_{i=1}^m \bar{w}_i(k) w_i(l).$$

We note that  $\sum_{i=1}^m \xi_i = Z + \sum_{i=1}^m \|w_i\|^2$ . Also,  $\sum_{k,l=1}^n |a_{k,l}|^2 = \text{tr}[\sum_{i=1}^m (Sv_i)(v_i^*S)]^2 \leq s_{\max}^4 m$ , where in the last step we used that  $M = \sum_{i=1}^m (Sv_i)(v_i^*S)$  is a matrix of rank at most  $m$  and  $M \leq S^2 \leq s_{\max}^2$ . Our assumptions on  $\epsilon$  and  $m$  give  $\frac{1}{2}(1-\epsilon)s_{\min}^2 m \geq (\log n)^C s_{\max}^2 \sqrt{m}$ . Therefore, using Lemma B.2 of [8],

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^m \xi_i \leq \epsilon s_{\min}^2 m\right) &\leq \mathbb{P}\left(|Z| > \sum_{i=1}^m \|w_i\|^2 - \epsilon s_{\min}^2 m\right) \\
&\leq \mathbb{P}\left(|Z| > (1-\epsilon)s_{\min}^2 m\right) \\
&\leq \mathbb{P}\left(\left|\sum_{k \neq l} a_{k,l}[b_k \bar{b}_l - \mathbb{E}b_k \bar{b}_l]\right| > (\log n)^C s_{\max}^2 \sqrt{m}\right) \\
&\quad + \mathbb{P}\left(\left|\sum_{k=1}^n a_{k,k}[b_k \bar{b}_k - \mathbb{E}b_k \bar{b}_k]\right| > (\log n)^C s_{\max}^2 \sqrt{m}\right) \\
&\leq \mathbb{P}\left(\left|\sum_{k \neq l} a_{k,l}[b_k \bar{b}_l - \mathbb{E}b_k \bar{b}_l]\right| > (\log n)^C \left(\sum_{k \neq l} |a_{k,l}|^2\right)^{\frac{1}{2}}\right) \\
&\quad + \mathbb{P}\left(\left|\sum_{k=1}^n a_{k,k}[b_k \bar{b}_k - \mathbb{E}b_k \bar{b}_k]\right| > (\log n)^C \left(\sum_{k=1}^n |a_{k,k}|^2\right)^{\frac{1}{2}}\right) \\
&\leq Cn^{-\log \log n}.
\end{aligned}$$

□

**Lemma 3.7** Assume  $E \in \mathcal{E}_\kappa^{(\mu)}$ , set  $I_\eta = [E - \frac{\eta}{2}, E + \frac{\eta}{2}]$  and let  $\mathcal{N}_\eta$  denote the number of eigenvalues of  $(YY^*)^{-1/2}XX^*(YY^*)^{-1/2}$  in  $I_\eta$ . If  $\frac{(\log n)^C}{n} < \eta < E/2$  then there exist constants  $c, K > 0$ , depending only on  $\gamma$  such that

$$\mathbb{P}\left(\mathcal{N}_\eta \geq \frac{Kn\eta}{\sqrt{E}}\right) \leq n^{-c \log \log n}$$

for all large  $n$ .

**Proof** First, we use that  $(YY^*)^{-1/2}XX^*(YY^*)^{-1/2}$  and  $X^*(YY^*)^{-1}X = \frac{1}{n}X^*TX$  have the same nonzero eigenvalues to justify working with the latter matrix. Now, following the proof of Lemma 8.1 in [7], we need to bound the absolute value of the diagonal entries of  $(\frac{1}{n}X^*TX - zI)^{-1}$ . We consider the  $(1, 1)$  entry and let  $x_1$  denote the first column of  $X$  and  $X_{(1)}$  the matrix obtained by removing this column from  $X$ . We use the following identity for an arbitrary matrix  $A$ , which can be seen by using a singular value decomposition,

$$A(A^*A - zI)^{-1}A^* = I + z(AA^* - zI)^{-1}.$$

We will use this identity for the matrix  $A = \frac{1}{\sqrt{n}}T\frac{1}{2}X_{(1)}$ . By the matrix inversion formula,

$$\left(\frac{1}{n}X^*TX - zI\right)^{-1} (1, 1) = \frac{1}{\frac{1}{n}x_1^*Tx_1 - z - \frac{1}{n^2}x_1^*TX_{(1)}\left(\frac{1}{n}X_{(1)}^*TX_{(1)} - zI\right)^{-1}X_{(1)}^*Tx_1}$$

$$\begin{aligned}
&= \frac{1}{\frac{1}{n}x_1^*Tx_1 - z - \frac{1}{n}x_1^*T^{\frac{1}{2}}(I + z(\frac{1}{n}T^{\frac{1}{2}}X_{(1)}X_{(1)}^*T^{\frac{1}{2}} - zI)^{-1})T^{\frac{1}{2}}x_1} \\
&= \frac{1}{-z - \frac{1}{n}zx_1^*T^{\frac{1}{2}}(\frac{1}{n}T^{\frac{1}{2}}X_{(1)}X_{(1)}^*T^{\frac{1}{2}} - zI)^{-1}T^{\frac{1}{2}}x_1}.
\end{aligned}$$

Let  $\mu_1, \dots, \mu_{n-1}$  and  $u_1, \dots, u_{n-1}$  denote the eigenvalues and normalized eigenvectors of  $\frac{1}{n}T^{\frac{1}{2}}X_{(1)}X_{(1)}^*T^{\frac{1}{2}}$ . Then setting  $\xi_k = |u_k^*T^{\frac{1}{2}}x_1|^2$  for  $1 \leq k \leq n-1$ , we have

$$\begin{aligned}
\left| \left( \frac{1}{n}X^*T^{-1}X - zI \right)^{-1} (1, 1) \right| &= \frac{1}{|z(1 + \frac{1}{n} \sum_{k=1}^{n-1} \frac{\xi_k}{\mu_k - z})|} \\
&\leq \frac{n}{|z| \Im \sum_{k=1}^{n-1} \frac{\xi_k}{\mu_k - z}} \\
&\leq \frac{n}{E \Im \sum_{k: |\mu_k - E| \leq \frac{\eta}{2}} \frac{\xi_k}{\mu_k - z}} \\
&\leq \frac{Cn\eta}{E \sum_{k: |\mu_k - E| \leq \frac{\eta}{2}} \xi_k}.
\end{aligned}$$

Continuing to follow [7], we have

$$\mathbb{P} \left( \mathcal{N}_\eta \geq \frac{Knn\eta}{\sqrt{E}} \right) \leq n\mathbb{P} \left( \sum_{k: \mu_k \in I_\eta} \xi_k \leq \frac{Cn\eta}{K\sqrt{E}} \ \& \ \mathcal{N}_\eta \geq \frac{Knn\eta}{\sqrt{E}} \right). \quad (38)$$

By Theorem 3.1 in [15], inequality (31) occurs with probability at least  $1 - Cn^{-\log \log n}$ , i.e.  $\frac{1}{2}a_- \leq T \leq 2a_+$ . By setting  $\epsilon = \frac{1}{2}$  in Lemma 3.6 and choosing  $K$  large enough, we satisfy

$$\frac{1}{2} \left( \frac{1}{2}a_- \right)^2 \sqrt{\mathcal{N}_\eta} \geq \frac{1}{2} \left( \frac{1}{2}a_- \right)^2 \sqrt{\frac{Knn\eta}{\sqrt{E}}} \geq \frac{1}{2} \left( \frac{1}{2}a_- \right)^2 \sqrt{\frac{K}{\sqrt{E}}} (\log n)^C \geq 2a_+^2 (\log n)^C.$$

We now apply Lemma 3.6 to obtain the claim.  $\square$

**Lemma 3.8** *Let  $\mu_k^{(j)}$ ,  $k = 1, \dots, n-1$  denote the eigenvalues of  $B_{(j)}$  for an arbitrary  $j = 1, 2, \dots, N$ , and assume that  $\eta > n^{-1}(\log n)^C$  and that  $E$  satisfies  $(E - \lambda_-)(\lambda_+ - E) \geq \kappa$ . Then, when  $\mathcal{T}_1$  holds, there exists a positive constant  $K$  such that with probability at least  $1 - n^{-c \log \log n}$*

$$\sup_{z \in \mathcal{E}^{(\mu)}} \max_{1 \leq j \leq N} \left\{ \frac{1}{n} \sum_k \frac{1}{|\mu_k^{(j)} - z|} \right\} \leq K(\log n)^2 \quad (39)$$

and

$$\sup_{z \in \mathcal{E}^{(\mu)}} \max_{1 \leq j \leq N} \left\{ \frac{1}{n} \sum_k \frac{1}{|\mu_k^{(j)} - z|^2} \right\} \leq \frac{K}{\eta}. \quad (40)$$

**Proof** For a fixed  $z \in \mathcal{E}^{(\mu)}$  and index  $j$ , given the bound on  $\mathbb{P} \left( \mathcal{N}_\eta \geq \frac{Knn\eta}{\sqrt{E}} \right)$  from Lemma 3.7, the claim follows from the same calculation as is done in the proof of Proposition 4.3 in [5].

This proves the estimate for each fixed  $z$ . To obtain the result simultaneously for all  $z \in \mathcal{E}^{(\mu)}$ , we notice that the derivatives of the functions to be bounded in (39) and (40) are bounded by  $Cn\eta^{-4}$  on  $\mathcal{E}^{(\mu)}$ . Thus, we may discretize  $\mathcal{E}^{(\mu)}$  to  $Cn^2\eta^{-8}$  points and take the union bound with respect to the discrete grid and the indices  $1 \leq j \leq N$ .  $\square$

**Proof of Proposition 3.4** We will prove the following four inequalities.

$$\mathbb{P}(\mathcal{X}) \geq 1 - n^{-c \log \log n} \quad (41)$$

$$\mathbb{P}(\mathcal{T}_1) \geq 1 - n^{-c \log \log n} \quad (42)$$

$$\mathbb{P}(\mathcal{T}_2 | \mathcal{T}_1) \geq 1 - n^{-c \log \log n} \quad (43)$$

$$\mathbb{P}(\mathcal{B} | \mathcal{T}_1) \geq 1 - n^{-c \log \log n}. \quad (44)$$

Let  $x$  denote an arbitrary column of  $X$ . We assume that  $n$  is large enough so that  $(\log n)^C < n^{\frac{1}{2}}$  and apply Lemma B.2 of [8] to the identity matrix. Then (41) follows by summing the  $N = bn$  probabilities. Inequality (42) follows from the much stronger statement in Theorem 3.1 of [15].

Recall the definition of  $\bar{\tau}_1, \dots, \bar{\tau}_n$  from (36) and let  $\bar{t}_1, \dots, \bar{t}_n$  denote the eigenvalues of  $\frac{1}{an}XX^*$ . Then Theorem 3.3 of [15] gives

$$|\bar{\tau}_k - \bar{t}_k| \leq (\log n)^{C_T \log \log n} n^{-2/3} (\min(n+1-k, k))^{-1/3} \text{ for all } k = 1, 2, \dots, n,$$

with probability at least  $1 - C \exp(-(\log n)^{c \log \log n})$ . Lemma 3.5 then implies that

$$|\tau_k - t_k| \leq (\log n)^{C_T \log \log n} n^{-2/3} (\min(n+1-k, k))^{-1/3} \text{ for all } k$$

with the same probability. Thus, assuming  $\mathcal{T}_1$  holds to address  $t_1$  and  $t_n$ , we have shown (43). Lastly, Lemma 3.8 yields (44). Together these give the bound

$$\mathbb{P}(\mathcal{X} \cap \mathcal{T}_1 \cap \mathcal{T}_2 \cap \mathcal{B}) \geq 1 - n^{-c \log \log n}.$$

$\square$

### 3.3 Integral Equation

In this section we show that the integral equation (18) that  $m_p(z)$  satisfies is stable. This means that if  $m_\delta$  satisfies

$$m_\delta(z) = \int \frac{1}{\lambda(1 - \frac{1}{a} - \frac{1}{a}zm_\delta(z)) - z} f_{\text{Inv}}(\lambda) d\lambda + \delta(z), \quad (45)$$

for some small  $\delta(z)$ , then  $m_\delta$  is close to  $m_p$ .

**Lemma 3.9** *Assume  $m_\delta(z) \in \mathbb{C}^+$  is analytic on the upper half plane and is a solution to the perturbed equation (45). Let  $0 < \eta' < 1$ ,  $\kappa > 0$  and assume  $\mathcal{T}_1$  holds. Let  $E$  be chosen so that  $(\lambda_+ - \frac{1}{E+1})(\frac{1}{E+1} - \lambda_-) \geq \kappa$ . There exists a small universal constant  $c_1 > 0$  such that if*

$$|\underline{m}_\delta(E + i\eta) - \underline{m}_p(E + i\eta)| \leq c_1 \sqrt{\kappa + \eta} \quad (46)$$

for all  $\eta \in [\eta', 1]$  and

$$\sup_{\eta \in [\eta', 1]} |\delta(E + i\eta)| < \delta_1 \quad (47)$$

for some  $\delta_1 \leq c_1 \sqrt{\kappa}$ , then

$$|\underline{m}_\delta(E + i\eta) - \underline{m}_p(E + i\eta)| \leq C_2 \frac{\delta_1}{\sqrt{\kappa + \eta}} \quad (48)$$

and

$$|m_\delta(E + i\eta) - m_p(E + i\eta)| \leq C_2 \frac{\delta_1}{\sqrt{\kappa + \eta}}, \quad (49)$$

for all  $\eta \in [\eta', 1]$ , where  $C_2$  depends only on  $c_1$ .

We define

$$\underline{m}_p(z) = -\frac{1}{z} \left(1 - \frac{1}{a}\right) + \frac{1}{a} m_p(z) \quad \text{and} \quad \underline{m}_\delta(z) = -\frac{1}{z} \left(1 - \frac{1}{a}\right) + \frac{1}{a} m_\delta(z). \quad (50)$$

Notice that for  $\delta = 0$ ,  $m_\delta$  and  $\underline{m}_\delta$  are the same as  $m_p$  and  $\underline{m}_p$ . Simple algebra shows that from (45) we have the following equation for  $\underline{m}_\delta$ :

$$\underline{m}_\delta(z) \left( z - \frac{1}{a} \int \frac{\lambda}{1 + \lambda \underline{m}_\delta(z)} f_{\text{Inv}}(\lambda) d\lambda \right) = -1 + \frac{z\delta}{a}. \quad (51)$$

We will work mostly with  $\underline{m}_\delta$  instead of  $m_\delta$ . We introduce the notation

$$K_0 = \frac{1}{a} \int A^2(\lambda) f_{\text{Inv}}(\lambda) d\lambda \quad \text{and} \quad A(\lambda) = \frac{\lambda \underline{m}_p(z)}{1 + \lambda \underline{m}_p(z)}.$$

Note that both  $K_0$  and  $A$  depend on  $z$ , which we omit writing here but will include in the proof of Lemma 3.9. The following lemma bounds  $K_0$  away from 1, which will be used in the proof of Lemma 3.9. We set

$$\mathcal{G} := \left\{ z = E + i\eta \in \mathbb{C} : E \in (\mu_-, \mu_+), \left( \lambda_+ - \frac{1}{E+1} \right) \left( \frac{1}{E+1} - \lambda_- \right) \geq \kappa, 0 < \eta \leq 1 \right\}.$$

**Lemma 3.10** *There exist two positive constants  $C$  and  $c$  such that the following statements hold uniformly for all  $z = E + i\eta \in \mathcal{G}$ .*

1.

$$c \leq |\underline{m}_p(z)| \leq C, \quad \text{and} \quad c \leq |m_p(z)| \leq C, \quad (52)$$

and  $\underline{m}_p(z)$ ,  $m_p(z)$  extend continuously to the interval  $[\mu_-, \mu_+] \subset \mathbb{R}$ .

2.

$$\Im \underline{m}_p(z) \geq \frac{1}{a} \Im m_p(z) \geq c \sqrt{\kappa + \eta}. \quad (53)$$

3. Let  $\underline{m}(z)$  be a function  $\mathbb{C}^+ \rightarrow \mathbb{C}^+$ . Suppose for some  $G \subset \mathcal{G}$

$$\sup_{z \in G} |\underline{m}_p(z) - \underline{m}(z)| < c. \quad (54)$$

When  $\mathcal{T}_1$  holds with a sufficiently small  $c$  in its definition (31), then

$$\sup_{z \in G} \|(\underline{m}(z)T + I)^{-1}\| < C. \quad (55)$$

4.

$$|1 - K_0(z)| > c\sqrt{\kappa + \eta}. \quad (56)$$

**Proof** 1.) From (51) with  $\delta = 0$  we have the following equation for  $\underline{m}_p(z)$

$$-\frac{1}{\underline{m}_p(z)} = z - \frac{1}{a} \int \frac{\lambda}{1 + \lambda \underline{m}_p(z)} f_{\text{Inv}}(\lambda) d\lambda. \quad (57)$$

If  $\{z_k\}_{k=1}^\infty$  is sequence in  $\{E + i\eta : E \in (\mu_-, \mu_+), 0 < \eta \leq 1\}$  such that either  $\lim_{k \rightarrow \infty} |\underline{m}_p(z_k)| = 0$  or  $\lim_{k \rightarrow \infty} |\underline{m}_p(z_k)| = \infty$ , then since  $0 < \mu_- \leq |z_k| \leq \mu_+ + 1$  for all  $k$ , (57) would be violated. Thus, no such sequence exists, which proves (52). The continuous extension follows from  $\underline{m}_p$  being analytic on  $\mathbb{C}^+$ . The statements for  $m_p$  follow from (50) and the fact that  $E = \Re z$  is separated away from zero.

2.) We define the interval  $J := [a_- + c(\kappa + \eta), a_+ - c(\kappa + \eta)]$ ; by setting the constant  $c$  sufficiently small and using  $\eta \leq 1$  we can ensure that  $|J| \geq \frac{1}{2}(a_+ - a_-)$ . Since  $f_p$  has a square root singularity near the edge of its support, we have  $f_p(\lambda) \geq c\sqrt{\kappa + \eta}$  for all  $\lambda \in J$ . Thus

$$\begin{aligned} \Im m_p(z) &= \int \frac{\eta}{(E - \lambda)^2 + \eta^2} f_p(\lambda) d\lambda \\ &\geq c\sqrt{\kappa + \eta} \int_J \frac{\eta}{(E - \lambda)^2 + \eta^2} d\lambda \\ &\geq c\sqrt{\kappa + \eta}, \end{aligned} \quad (58)$$

where the last inequality follows from the fact that the length of  $J$  and the distance from  $E$  to  $J$  are both  $O(1)$  and  $\eta \leq 1$ . We then also have  $\Im \underline{m}_p(z) \geq \frac{1}{a} \Im m_p(z) \geq c_1 \sqrt{\kappa + \eta}$  maybe with a smaller constant  $c_1$ .

3.) We take the imaginary part of (57) to obtain

$$\frac{\Im \underline{m}_p}{|\underline{m}_p|^2} = \eta + \frac{1}{a} \int \frac{\lambda^2 \Im \underline{m}_p}{|1 + \lambda \underline{m}_p|^2} f_{\text{Inv}}(\lambda) d\lambda, \quad (59)$$

so that

$$1 = \frac{\eta |\underline{m}_p|^2}{\Im \underline{m}_p} + \frac{1}{a} \int |A(\lambda)|^2 f_{\text{Inv}}(\lambda) d\lambda \quad (60)$$

and

$$\frac{1}{a} \int |A(\lambda)|^2 f_{\text{Inv}}(\lambda) d\lambda \leq 1. \quad (61)$$

From (61), the uniform bound on  $|\underline{m}_p(z)|$  and the lower bound on the support of  $f_{\text{Inv}}$  we have

$$\int \frac{1}{|1 + \lambda \underline{m}_p(z)|^2} f_{\text{Inv}}(\lambda) d\lambda < C' \quad (62)$$

with some constant  $C'$  uniformly for all  $z \in \mathcal{G}$ . Suppose that

$$w(z) := \inf_{\lambda \in [a_-, a_+]} |1 + \lambda \underline{m}_p(z)|$$

is attained at  $\lambda = \lambda_0(z)$  for any  $z \in \mathcal{G}$ . Note that  $w(z) \geq a_- \Im \underline{m}_p(z) \geq c\sqrt{\kappa + \eta}$  with some positive constant  $c$ , where we used (53). Since the derivative of  $\lambda \mapsto |1 + \lambda \underline{m}_p(z)|$  is uniformly bounded by (52), there exists a subinterval  $J = J(z) \subset [a_-, a_+]$  of length at least  $cw(z)$  such that  $|1 + \lambda \underline{m}_p(z)| \leq 2w(z)$  for all  $\lambda \in J$ . Since  $f_{\text{Inv}}(\lambda)$  has a square-root singularity at its edges and is bounded away from zero between the edges, we have

$$C' \geq \int \frac{1}{|1 + \lambda \underline{m}_p(z)|^2} f_{\text{Inv}}(\lambda) d\lambda \geq \frac{1}{4w(z)^2} \int_J f_{\text{Inv}}(\lambda) d\lambda \geq \frac{c}{\sqrt{w(z)}}. \quad (63)$$

Therefore we have a uniform lower bound

$$|1 + \lambda \underline{m}_p(z)| > c', \quad (64)$$

for  $c'$  depending only on the constant  $C'$  in (62), for all  $\lambda \in (a_-, a_+)$  and for all  $z \in \mathcal{G}$ . Using continuity in both  $\lambda$  and  $\underline{m}_p(z)$  and using that (54) holds with a sufficiently small  $c$ , we have  $|1 + t \underline{m}_p(z)| > c$  for all  $t \in [(1-c)a_-, (1+c)a_+]$  and  $z \in \mathcal{G}$  if  $c$  is chosen sufficiently small. By applying the spectral theorem for  $T$  and (31) with a small  $c$ , we obtain (55).

4.) The variable  $z$  plays no role in the remainder of the proof and so we omit it from the notation. By the assumption on  $E$ , we have  $\Im \underline{m}_p \geq c\sqrt{\kappa + \eta}$ . The property  $|\underline{m}_p(z)| < C$  and  $\lambda \leq a_+$  give

$$\frac{\Im A(\lambda)}{|A(\lambda)|} = \frac{\Im \underline{m}_p}{|1 + \lambda \underline{m}_p| |\underline{m}_p|} \geq c \Im \underline{m}_p(z) \geq c\sqrt{\kappa + \eta}. \quad (65)$$

We set

$$B := \frac{1}{a} \int |A(\lambda)|^2 f_{\text{Inv}}(\lambda) d\lambda,$$

so that by (61) we have  $|K_0| \leq B < 1$ . We claim that

$$|B - K_0| \geq cB\sqrt{\kappa + \eta} \quad (66)$$

for a positive constant  $c$ .

By the lower bound on  $|\underline{m}_p(z)|$  in (52) and on  $|1 + \lambda \underline{m}_p(z)|$  in (64), there exists a constant  $C$  such that

$$\frac{\Im A(\lambda)}{|A(\lambda)|} = \frac{\Im \underline{m}_p(z)}{|1 + \lambda \underline{m}_p(z)| |\underline{m}_p(z)|} \leq C \Im \underline{m}_p(z) \quad (67)$$

for all  $z \in \mathcal{G}$  and  $\lambda \in (a_-, a_+)$ . Let  $\epsilon > 0$  be a sufficiently small constant so that  $1 - 2C\epsilon \geq \frac{1}{2}$  with the constant  $C$  from (67). If  $\Im \underline{m}_p(z) > \epsilon$ , then, using (65),

$$|B - K_0| = \frac{1}{a} \left| \int (|A(\lambda)|^2 - A^2(\lambda)) f_{\text{Inv}}(\lambda) d\lambda \right|$$

$$\begin{aligned}
&= \frac{1}{a} \left| \int 2(\Im A(\lambda))^2 f_{\text{Inv}}(\lambda) d\lambda - 2i \int (\Re A(\lambda))(\Im A(\lambda)) f_{\text{Inv}}(\lambda) d\lambda \right| \\
&\geq \frac{2}{a} \int (\Im A(\lambda))^2 f_{\text{Inv}}(\lambda) d\lambda \\
&\geq c\epsilon \Im \underline{m}_p(z) \int |A(\lambda)|^2 f_{\text{Inv}}(\lambda) d\lambda \\
&\geq c\epsilon B \sqrt{\kappa + \eta}.
\end{aligned}$$

If  $\Im \underline{m}_p(z) \leq \epsilon$ , then we set  $A(\lambda) = e^{i\phi(\lambda)} |A(\lambda)|$ . We note that  $\phi \in (0, \pi)$  since  $\Im A(\lambda) > 0$ , and that  $\phi$  is well-defined since  $|A| \neq 0$ . By (67) we have  $\sin \phi \leq C\epsilon$ , and by continuity, either  $0 \leq \phi \leq C\epsilon$  or  $\pi - C\epsilon \leq \phi \leq \pi$ . In both cases we have  $\cos \phi \geq \frac{1}{2}$  if  $\epsilon$  is small. So we have

$$\begin{aligned}
|B - K_0| &= \frac{1}{a} \left| \int (|A(\lambda)|^2 - A^2(\lambda)) f_{\text{Inv}}(\lambda) d\lambda \right| \\
&= \frac{2}{a} \left| \int |A(\lambda)|^2 e^{i\phi} \sin \phi f_{\text{Inv}}(\lambda) d\lambda \right| \\
&\geq \frac{2}{a} \left| \int |A(\lambda)|^2 \cos \phi \sin \phi f_{\text{Inv}}(\lambda) d\lambda \right| - \frac{2}{a} \int |A(\lambda)|^2 \sin^2 \phi f_{\text{Inv}}(\lambda) d\lambda \\
&\geq \frac{1}{a} \int |A(\lambda)|^2 (\sin \phi - 2 \sin^2 \phi) f_{\text{Inv}}(\lambda) d\lambda \\
&\geq \frac{1}{2a} \int |A(\lambda)|^2 \sin \phi f_{\text{Inv}}(\lambda) d\lambda \\
&\geq c \Im \underline{m}_p(z) \int |A(\lambda)|^2 f_{\text{Inv}}(\lambda) d\lambda,
\end{aligned}$$

where for the last inequality we use  $\sin \phi \sim \Im A/|A|$  and (67). Since  $\epsilon$  depends only on  $C$  from (67), we use (53) to obtain

$$|B - K_0| \geq cB \sqrt{\kappa + \eta}.$$

We trivially have  $|1 - K_0| \geq |B - K_0|$  for any positive number  $B$  with  $B \leq 1$  and complex number  $K_0$  with  $|K_0| \leq B$ . Therefore (56) follows from (66) and from a bound  $B \geq c > 0$ . The bound  $B \geq c > 0$ , follows from the fact that  $|A(\lambda)| \geq c > 0$  on the support of  $f_{\text{Inv}}$  and that  $f_{\text{Inv}}$  is a probability measure whose support is separated from zero.  $\square$

**Proof of Lemma 3.9** Throughout the proof,  $\delta$  is a function of  $z$ , but we will omit writing this dependence. We also fix  $E = \Re z$  and we vary only  $\eta = \Im z$ . Note that from definition (50) we have

$$m_p(z) - m_\delta(z) = a(\underline{m}_p(z) - \underline{m}_\delta(z)), \quad (68)$$

so that (49) follows from (48).

To prove (48) we observe that

$$\underline{m}_p(z) - \underline{m}_\delta(z) = (\underline{m}_p(z) - \underline{m}_\delta(z))K(z) - \frac{z}{a} \left( z - \frac{1}{a} \int \frac{\lambda}{1 + \lambda \underline{m}_\delta(z)} f_{\text{Inv}}(\lambda) d\lambda \right)^{-1} \delta, \quad (69)$$

where

$$K(z) := \frac{\frac{1}{a} \int \frac{\lambda}{1+\lambda \underline{m}_\delta(z)} \frac{\lambda}{1+\lambda \underline{m}_p(z)} f_{\text{Inv}}(\lambda) d\lambda}{\left(z - \frac{1}{a} \int \frac{\lambda}{1+\lambda \underline{m}_\delta(z)} f_{\text{Inv}}(\lambda) d\lambda\right) \left(z - \frac{1}{a} \int \frac{\lambda}{1+\lambda \underline{m}_p(z)} f_{\text{Inv}}(\lambda) d\lambda\right)} \quad (70)$$

is obtained by taking the difference of the expressions for  $\underline{m}_p(z)$  and  $\underline{m}_\delta(z)$  given in (51) and (57). Since  $|\underline{m}_\delta(z)| \leq C$  and, for  $\delta$  small enough,  $|z\delta| \leq \frac{a}{2}$ , using (51) we have

$$\left| \left( z - \frac{1}{a} \int \frac{\lambda}{1+\lambda \underline{m}_\delta(z)} f_{\text{Inv}}(\lambda) d\lambda \right)^{-1} \right| \leq C,$$

and therefore (69) yields

$$|\underline{m}_p(z) - \underline{m}_\delta(z)| = \frac{C|\delta|}{|1 - K(z)|}. \quad (71)$$

From (51) and (70) we have

$$K(z) = \frac{1}{a} \int \frac{\lambda \underline{m}_p(z)}{1+\lambda \underline{m}_p(z)} \frac{\lambda \underline{m}_\delta(z)}{1+\lambda \underline{m}_\delta(z)} f_{\text{Inv}}(\lambda) d\lambda - \frac{z\delta}{z\delta - a} \int \frac{\lambda \underline{m}_p(z)}{1+\lambda \underline{m}_p(z)} \frac{\lambda \underline{m}_\delta(z)}{1+\lambda \underline{m}_\delta(z)} f_{\text{Inv}}(\lambda) d\lambda. \quad (72)$$

Since  $|z\delta| \leq \frac{a}{2}$ ,  $|\delta| \leq \delta_1$ , the absolute value of the second term in (72) is bounded by  $C\delta_1$ . Here we used (64) and that a similar positive lower bound holds for  $|1 + \lambda \underline{m}_\delta(z)|$  as well, assuming that  $c_1$  in (46) is sufficiently small. Thus,

$$\begin{aligned} |K(z) - K_0(z)| &\leq \left| \frac{1}{a} \int \frac{\lambda \underline{m}_p(z)}{1+\lambda \underline{m}_p(z)} \left( \frac{\lambda \underline{m}_\delta(z)}{1+\lambda \underline{m}_\delta(z)} - \frac{\lambda \underline{m}_p(z)}{1+\lambda \underline{m}_p(z)} \right) f_{\text{Inv}}(\lambda) d\lambda \right| + C\delta_1 \\ &\leq \frac{1}{a} \int \left| \frac{\lambda \underline{m}_p(z)}{1+\lambda \underline{m}_p(z)} \frac{\lambda(\underline{m}_\delta(z) - \underline{m}_p(z))}{(1+\lambda \underline{m}_p(z))(1+\lambda \underline{m}_\delta(z))} \right| f_{\text{Inv}}(\lambda) d\lambda + C\delta_1 \\ &\leq C|\underline{m}_\delta(z) - \underline{m}_p(z)| + C\delta_1. \end{aligned}$$

By (46) and  $\delta_1 \leq c_1\sqrt{\kappa}$  we now have

$$|K(z) - K_0(z)| \leq Cc_1\sqrt{\kappa + \eta} \quad (73)$$

for all  $\eta \in [\eta', 1]$ .

Choosing  $c_1$  so small so that  $Cc_1 \leq \frac{1}{2}c$  where  $c$  is the constant obtained in the estimate (56) in Lemma 3.10, we have

$$\frac{1}{|1 - K(z)|} \leq \frac{C}{\sqrt{\kappa + \eta}}. \quad (74)$$

Using (71) we have

$$|\underline{m}_p(z) - \underline{m}_\delta(z)| \leq \frac{C\delta_1}{\sqrt{\kappa + \eta}} \quad (75)$$

for all  $\eta \in [\eta', 1]$  and with a sufficiently large constant  $C$ .  $\square$

**Lemma 3.11** *Let  $\mu$  be a probability measure supported on some interval  $[u, v] \subset \mathbb{R}$  and let  $\{\tau_k\}_{k=1}^n$  be real numbers such that*

$$\int_u^{\tau_k} d\mu(\tau) = \frac{k}{n}$$

for  $k = 1, \dots, n$  and assume  $m \in \mathbb{C}^+$ . Then

$$\left| \frac{1}{n} \sum_{k=1}^n \frac{1}{1 + m\tau_k} - \int \frac{1}{1 + m\tau} d\mu(\tau) \right| \leq \frac{1}{n} \frac{|v - u| \cdot |m|}{\inf_{t \in [u, v]} |1 + mt|^2}.$$

If  $\{t_k\}_{k=1}^n$  is another set of points in  $\mathbb{R}$ , then

$$\left| \frac{1}{n} \sum_{k=1}^n \frac{1}{1 + mt_k} - \frac{1}{n} \sum_{k=1}^n \frac{1}{1 + m\tau_k} \right| \leq \frac{1}{n} \frac{|m|}{\inf_{t \in [u, v]} |1 + mt|^2} \sum_{k=1}^n |t_k - \tau_k|.$$

**Proof** For an arbitrary differentiable  $f$  we have

$$\left| \frac{1}{n} \sum_{k=1}^n f(\tau_k) - \int f(\tau) d\mu(\tau) \right| \leq \frac{|b - a|}{n} \sup_{t \in [u, v]} |f'(t)|.$$

Here we have

$$\sup_{t \in [u, v]} \left| \frac{d}{dt} \frac{1}{1 + mt} \right| \leq \frac{|m|}{\inf_{t \in [u, v]} |1 + tm|^2}. \quad (76)$$

This proves the first claim. The second claim is proven similarly.  $\square$

### 3.4 A bound on $|d_j|$

For any  $z \in \mathcal{G}$  we set

$$M(z) := \max \left( \left\| (\underline{m}_n(z)T + I)^{-1} \right\|, \max_{1 \leq j \leq N} \left\| (\underline{m}_{(j)}(z)T + I)^{-1} \right\| \right). \quad (77)$$

**Lemma 3.12** *Suppose that for some  $z \in \mathcal{G}$  we have  $M(z) \leq C_0$  with some constant  $C_0$ . Then there exists a constant  $C$  such that*

$$\mathbb{P} \left( |d_j(z)| > \frac{M(\log n)^C}{\sqrt{n\eta}} \mid \mathcal{B}, \mathcal{T}_1, \mathcal{T}_2, \mathcal{X} \right) \leq Cn^{-\log \log n}$$

for  $j = 1, \dots, N$  whenever  $n\eta \geq M^2$ .

We start with a short lemma.

**Lemma 3.13** *For  $m_n$  and  $m_{(j)}$  as defined in (20) and  $z = E + i\eta$ ,*

$$|m_n(z) - m_{(j)}(z)| \leq \frac{\pi}{\eta n}$$

for all  $n \in \mathbb{N}$  and all  $E \in \mathbb{R}$ .

**Proof** Using  $F^n$  and  $F^{(j)}$  to denote the corresponding distribution functions, by Theorem A.44 of [1],

$$\sup_{t \in \mathbb{R}} |F^n(t) - F^{(j)}(t)| \leq \frac{1}{n}.$$

Then

$$|m_n(z) - m_{(j)}(z)| \leq \frac{1}{n} \int \frac{dx}{|x - z|^2} = \frac{\pi}{\eta n}.$$

□

**Proof of Lemma 3.12** We need to bound the terms (26)-(29) and will use Lemmas 3.8 and 3.10 repeatedly. For (26) we use Lemma 3.13 and obtain

$$\frac{1}{n} |\mathbb{E} \text{tr}(B_{(j)} - zI)^{-1} (\underline{m}_n(z)T + I)^{-1} (\underline{m}_n(z) - \underline{m}_{(j)}(z))T (\underline{m}_n(z)T + I)^{-1}| \leq \frac{CM^2}{n\eta} \log^2 n.$$

Similarly,

$$\begin{aligned} & \frac{1}{n} \|(B_{(j)} - zI)^{-1} (\underline{m}_n(z)T + I)^{-1} (\underline{m}_n(z) - \underline{m}_{(j)}(z))T (\underline{m}_n(z)T + I)^{-1}\|_{HS} \\ & \leq \frac{CM^2}{n\eta} \frac{1}{n} (\text{tr}|B_{(j)} - zI|^{-2})^{\frac{1}{2}} \\ & \leq \frac{CM^2}{(n\eta)^{3/2}}, \end{aligned}$$

where the last inequality follows from (40). Using Lemma B.2 of [8],

$$\mathbb{P} \left( |(26)| \geq \frac{M(\log n)^C}{\sqrt{n\eta}} \right) \leq n^{-\log \log n}.$$

For (27) we need to bound the absolute value of

$$\frac{b}{N} \sum_{k,l=1}^N \left[ T^{\frac{1}{2}} (B_{(j)} - zI)^{-1} (\underline{m}_{(j)}(z)T + I)^{-1} T^{\frac{1}{2}} \right]_{k,l} (b_k \bar{b}_l - \mathbb{E} b_k \bar{b}_l), \quad (78)$$

where  $b_k$  denotes the  $k^{\text{th}}$  entry of the vector  $x_j$ . On the set  $\mathcal{B}$ , the second inequality in (34) gives the following bound on the Hilbert-Schmidt norm

$$\begin{aligned} \frac{1}{n} \|T^{1/2} (B_{(j)} - zI)^{-1} (\underline{m}_{(j)}(z)T + I)^{-1} T^{1/2}\|_{HS} & \leq \frac{C}{n} \|(\underline{m}_{(j)}(z)T + I)^{-1}\| \| (B_{(j)} - zI)^{-1} \|_{HS} \\ & \leq \frac{CM}{n} \| (B_{(j)} - zI)^{-1} \|_{HS} \\ & \leq \frac{CM}{\sqrt{n\eta}}. \end{aligned}$$

Thus, by Lemma B.2 of [8],

$$\mathbb{P} \left( |(27)| > M \frac{(\log n)^C}{\sqrt{n\eta}} \right) \leq C n^{-\log \log n}. \quad (79)$$

For (28), using Lemma 3.13, (39) and the resolvent identity, we have

$$\begin{aligned}
|(28)| &\leq \frac{C}{n} \|(\underline{m}_n(z)T + I)^{-1} - (\underline{m}_{(j)}(z)T + I)^{-1}\| \operatorname{tr}|(B_{(j)} - zI)^{-1}| \\
&\leq \frac{CM^2}{n} \frac{1}{n\eta} \operatorname{tr}|(B_{(j)} - zI)^{-1}| \\
&\leq \frac{CM^2 \log^2 n}{n\eta}.
\end{aligned}$$

For (29), using the resolvent identity we see that

$$((B_{(j)} - zI)^{-1} - (B_n - zI)^{-1})(\underline{m}_n(z)T + I)^{-1}T = (B_{(j)} - zI)^{-1}r_j r_j^*(B_n - zI)^{-1}(\underline{m}_n(z)T + I)^{-1}T \quad (80)$$

has rank one, so that a bound on the norm of (80) gives a bound on its trace. We then return to the expression on the left of (80) and use  $\|(B_{(j)} - zI)^{-1}\|, \|(B_n - zI)^{-1}\| \leq \eta^{-1}$  to obtain

$$|(29)| \leq \frac{1}{n} \|((B_{(j)} - zI)^{-1} - (B_n - zI)^{-1})(\underline{m}_n(z)T + I)^{-1}T\| \leq \frac{CM}{n\eta}.$$

Combining these four bounds just obtained gives the statement of the lemma.  $\square$

### 3.5 Proof of Theorem 3.2

We isolate three parts of the proof of Theorem 3.2 in the following lemmas.

**Lemma 3.14** *Let  $z = E + i\eta$  with  $E \in (\mu_-, \mu_+)$  and  $\eta > n^{-1}\kappa^{-2}(\log n)^{4C_T \log \log n}$ , and assume that  $\mathcal{T}_1$  and  $\mathcal{X}$  hold. Then with probability at least  $1 - n^{-c \log \log n}$ ,*

$$\min_{1 \leq j \leq N} |1 + r_j^*(B_{(j)} - zI)^{-1}r_j| \geq \frac{1}{4\mu_+ |m_n(z)|}. \quad (81)$$

The proof of Lemma 3.14 is given following the proof of Theorem 3.2.

**Lemma 3.15** *Assume  $E \in (\mu_-, \mu_+)$  and  $(\lambda_+ - \frac{1}{E+1}) (\frac{1}{E+1} - \lambda_-) \geq \kappa$  and that  $\mathcal{T}_1$  and  $\mathcal{X}$  hold. Then for any sufficiently small constant  $c_1 > 0$  there exists  $c > 0$  depending only on  $c_1$  such that with probability at least  $1 - n^{-c \log \log n}$*

$$|m_p(E + i) - m_n(E + i)| \leq \frac{(\log n)^{2C_T \log \log n}}{\sqrt{n}}. \quad (82)$$

In particular, for any sufficiently small constant  $c_1 > 0$

$$|m_p(E + i) - m_n(E + i)| \leq c_1 \sqrt{\kappa + 1}. \quad (83)$$

Moreover, we also have

$$\min_{1 \leq j \leq N} |1 + r_j^*(B_{(j)} - (E + i))^{-1}r_j| \geq c_1 \quad (84)$$

holds for all  $n$  large enough.

**Proof** Let  $\{\mu_k\}_{k=1}^{n-1}$  and  $\{v_k\}_{k=1}^{n-1}$  denote the eigenvalues and corresponding normalized eigenvectors of  $B_{(j)}$ . In the following we use that the eigenvalues of  $B_{(j)}$  and  $B_n$  interlace (see page 82 of [2]). Given that we are on the set  $\mathcal{T}_1 \cap \mathcal{X}$  (see (30) and (31)), we have

$$\begin{aligned} \Im r_j^*(B_{(j)} - (E + i)I)^{-1}r_j &\geq \sum_{k=1}^{n-1} \frac{|v_k^* r_j|^2}{(8ba_+ + \mu_+)^2 + 1} \geq c \sum_{k=1}^{n-1} |v_k^* r_j|^2 \\ &= c \|r_j\|^2 = c \|Tx\|^2 \geq c \|x\|^2 \geq c. \end{aligned} \quad (85)$$

A similar calculation together with (50) implies  $\Im \underline{m}_n(E + i), \Im \underline{m}_{(j)}(E + i) \geq c$ , hence, recalling the definition (77),

$$M = M(E + i) = \max \left( \|(\underline{m}_n(E + i)T + I)^{-1}\|, \max_{1 \leq j \leq N} \|(\underline{m}_{(j)}(z)T + I)^{-1}\| \right) \leq C. \quad (86)$$

Recalling the constant  $C_T$  from (33), we set

$$s_\eta := \frac{(\log n)^{\frac{3}{2}} C_T \log \log n}{\sqrt{n\eta}}. \quad (87)$$

Using  $M$  as just estimated, for  $n$  large enough the extra factors of  $\log n$  in  $s_1$  will exceed the constant necessary to satisfy the condition of Lemma 3.12. Thus, by Lemma 3.12 and (35),

$$|d_j(E + i)| < s_1, \quad j = 1, \dots, N,$$

with probability at least  $1 - Cn^{-\log \log n}$ . Returning to (25) and using the lower bound on the denominators given by (85),

$$\left| \frac{a}{(E + i)N} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - (E + i)I)^{-1}r_j} d_j(E + i) \right| \leq C s_1. \quad (88)$$

Using equation (19) and recalling that  $\{t_1, \dots, t_n\}$  are the eigenvalues of  $T$ ,

$$\left| m_n(E + i) - \int \frac{1}{\lambda(1 - \frac{1}{a} - \frac{1}{a}(E + i)m_n(E + i)) - (E + i)} f_{\text{Inv}}(\lambda) d\lambda \right| \quad (89)$$

$$\begin{aligned} &= \left| m_n(E + i) - \frac{1}{E + i} \int \frac{1}{1 + \lambda \underline{m}_n(E + i)} f_{\text{Inv}}(\lambda) d\lambda \right| \\ &\leq \left| m_n(E + i) - \frac{1}{E + i} \frac{1}{n} \sum_{k=1}^n \frac{1}{1 + t_k \underline{m}_n(E + i)} \right| \end{aligned} \quad (90)$$

$$+ \left| \frac{1}{E + i} \frac{1}{n} \sum_{k=1}^n \frac{1}{1 + t_k \underline{m}_n(E + i)} - \frac{1}{E + i} \int \frac{1}{1 + \lambda \underline{m}_n(E + i)} f_{\text{Inv}}(\lambda) d\lambda \right|. \quad (91)$$

For (90) we use (25) and (88) to obtain

$$\left| m_n(E + i) - \frac{1}{E + i} \frac{1}{n} \sum_{k=1}^n \frac{1}{1 + t_k \underline{m}_n(E + i)} \right| \leq C s_1.$$

For (91) we use (33), both parts of Lemma 3.11, and the bounds (52), (86) to obtain

$$\begin{aligned} & \left| \frac{1}{E+i} \frac{1}{n} \sum_{k=1}^n \frac{1}{1+t_k \underline{m}_n(E+i)} - \frac{1}{E+i} \int \frac{1}{1+\lambda \underline{m}_n(E+i)} f_{\text{Inv}}(\lambda) d\lambda \right| \\ & \leq \frac{C}{n} + \frac{(\log n)^{C_T \log \log n}}{n} \\ & \leq \frac{(\log n)^{C_T \log \log n}}{n}. \end{aligned}$$

Since  $s_1 > n^{-1}(\log n)^{C_T \log \log n}$  and  $s_1 \leq c_1 \sqrt{\kappa}$  if  $n$  is sufficiently large, we have

$$(89) \leq C s_1, \quad (92)$$

so that by inequality (48) of Lemma 3.9 we have

$$|m_p(E+i) - m_n(E+i)| \leq C s_1 \leq \frac{(\log n)^{2C_T \log \log n}}{\sqrt{n}}. \quad (93)$$

This yields (82) and for  $n$  large enough we also get (83). From (93), (52) and (81) we have the bound

$$\min_{1 \leq j \leq N} |1 + r_j^*(B_{(j)} - (E+i)I)^{-1} r_j| \geq \frac{1}{4C\mu_+}$$

with probability at least  $1 - n^{-c \log \log n}$ . Choosing  $c_1$  small enough, this yields (84).  $\square$

**Lemma 3.16** *Assume  $E \in (\mu_-, \mu_+)$  and  $(\lambda_+ - \frac{1}{E+1}) (\frac{1}{E+1} - \lambda_-) \geq \kappa$  and choose an  $\eta' \in [n^{-1}\kappa^{-2}(\log n)^{4C_T \log \log n}, 1]$ , where  $C_T$  is the constant appearing in (33). Set  $\eta'' := \eta' - n^{-2}$ . Assume that  $\mathcal{T}_1$  and  $\mathcal{X}$  hold and that with a sufficiently small constant  $c_1$*

$$|m_p(E+i\eta) - m_n(E+i\eta)| \leq c_1 \sqrt{\kappa + \eta} \quad (94)$$

*holds for all  $\eta \in [\eta', 1]$  with some probability at least  $1 - P(n)$ . Then with a probability at least  $1 - P(n) - n^{-c \log \log n}$ , we have*

$$|m_p(E+i\eta) - m_n(E+i\eta)| \leq \frac{(\log n)^{2C_T \log \log n}}{\sqrt{n\eta\kappa}} \quad (95)$$

*for all  $\eta \in [\eta'', 1]$ . In particular, provided  $\eta'' \geq n^{-1}\kappa^{-2}(\log n)^{4C_T \log \log n}$ , the bound (94) holds for all  $\eta \in [\eta'', 1]$ , with a probability at least  $1 - P(n) - n^{-c \log \log n}$ .*

**Proof** By a trivial continuity argument, first we prove that

$$|m_p(E+i\eta) - m_n(E+i\eta)| \leq 2c_1 \sqrt{\kappa + \eta} \quad (96)$$

holds for all  $\eta \in [\eta'', 1]$  with probability at least  $1 - P(n)$ . Indeed, the functions  $m_p(E+i\eta)$  and  $m_n(E+i\eta)$  are Lipschitz continuous for fixed  $E$  and  $\eta \in [\eta'', 1]$  with derivatives bounded by  $(1/\eta'')^2$ . In particular, the derivative of  $m_n$  with respect to  $\eta$  is uniformly bounded on

$[n^{-1}\kappa^{-2}(\log n)^{4C_T \log \log n}, 1]$  by  $Cn^2\kappa^4(\log n)^{-8C_T \log \log n}$ . By this continuity and the proximity of  $m_n(E + i\eta')$  and  $m_p(E + i\eta')$  given by (94), the inequality (96) holds for all  $\eta \in [\eta'', 1]$ .

Now we show that the stronger estimate (94) can be regained from the weaker estimate (96) for all  $\eta \in [\eta'', 1]$ . Assuming that  $c_1$  is sufficiently small, Lemma 3.13, the first inequality of (96) and inequality (55) of Lemma 3.10 along with the spectral theorem give the bound  $M(E + i\eta) < C$  for all  $\eta \in [\eta'', 1]$ . Recalling the definition of  $s_\eta$  from (87), we use the bound on  $M$  and Lemma 3.12 for large enough  $n$  to obtain

$$\max_{1 \leq j \leq N} |d_j(E + i\eta)| < s_\eta$$

with probability at least  $1 - P(n) - n^{-c \log \log n}$  for all  $\eta \in [\eta'', 1]$ .

Using (52) and (96), we get that  $|m_n(E + i\eta)| \geq \frac{1}{2}c$  for all  $\eta \in [\eta'', 1]$  if  $c_1$  is sufficiently small, where  $c$  is the constant from Lemma 3.10. Then from Lemma 3.14, we see that, with probability at least  $1 - P(n) - n^{-c \log \log n}$ ,

$$\min_{1 \leq j \leq N} |1 + r_j^*(B_{(j)} - (E + i\eta))^{-1}r_j| \geq c' \quad (97)$$

with some positive constant  $c'$ . Using (97) to bound the denominators and an argument just like the one used for the point  $E + i$  in Lemma 3.15,

$$\left| m_n(E + i\eta) - \int \frac{1}{\lambda(1 - \frac{1}{a} - \frac{1}{a}(E + i\eta)m_n(E + i\eta)) - (E + i\eta)} f_{\text{Inv}}(\lambda) d\lambda \right| \leq C s_\eta \quad (98)$$

for all  $\eta \in [\eta'', 1]$ . By Lemma 3.9 we obtain

$$|m_p(E + i\eta) - m_n(E + i\eta)| \leq \frac{C s_\eta}{\sqrt{\kappa + \eta}} \leq \frac{(\log n)^{2C_T \log \log n}}{\sqrt{n\eta\kappa}} \quad (99)$$

for all  $\eta \in [\eta'', 1]$  with probability at least  $1 - P(n) - n^{-c \log \log n}$ . Since

$$\eta \geq \eta'' > n^{-1}\kappa^{-2}(\log n)^{4C_T \log \log n},$$

we have

$$s_\eta < (\kappa + \eta)(\log n)^{-2C_T \log \log n},$$

which yields (94) with a very high probability for all  $\eta \in [\eta'', 1]$  and for sufficiently large  $n$ .  $\square$

**Proof of Theorem 3.2** The proof of *i*) follows directly from the previous two lemmas. Set  $\eta_p = 1 - pn^{-2}$ ,  $p = 0, 1, 2, \dots$ , then (13) is proved in Lemma 3.15 for  $\eta = 1$  with probability at least  $1 - n^{-c \log \log n}$ . Lemma 3.16 then implies that

$$|m_p(E + i\eta) - m_n(E + i\eta)| \leq \frac{(\log n)^{2C_T \log \log n}}{\sqrt{n\eta\kappa}} \quad (100)$$

holds for all  $\eta \in [\eta_1, \eta_0 = 1]$  with the possible exception of a set of probability  $2n^{-c \log \log n}$ . Iterating Lemma 3.16, we get that (100) holds for all  $\eta \in [\eta_p, 1]$  with probability at least  $1 - (p + 1)n^{-c \log \log n}$  as long as  $\eta_p \geq n^{-1}\kappa^{-2}(\log n)^{4C_T \log \log n}$ . This proves the first part of

Theorem 3.2 for a fixed energy  $E$ . To take care of all energies simultaneously, we use that the derivative of  $m_n$  is uniformly bounded by  $C\eta^{-2}$ . Thus we can discretize the energy range to  $Cn^2$  points and take the union bound to obtain  $i)$  of Theorem 3.2.

Now we prove  $ii)$ . Set

$$\rho_\eta(E) = \frac{1}{\pi} \Im m_n(E + i\eta) = \frac{1}{\pi n} \sum_{k=1}^n \frac{\eta}{(\mu_k - E)^2 + \eta^2}.$$

Given Lemma 3.7, the argument for Corollary 2.2 in [4] gives

$$\mathbb{P} \left( \sup_{E \in \mathcal{E}_\kappa^{(\mu)}} \rho_\eta(E) > K \right) \leq 1 - Cnn^{-\log \log n}$$

where  $K$  is the constant in Lemma 3.7. Using this inequality,  $ii)$  follows from the argument given to prove the analogous claim in Corollary 4.2 of [4].  $\square$

**Proof of Lemma 3.14** From (23) we have

$$m_n(z) = \frac{-1}{1 + \frac{1}{n} \text{tr} T(B_n - zI)^{-1}} \frac{1}{z} (1 - A - B) \quad (101)$$

where

$$A := \frac{1}{N} \sum_{j=1}^N \frac{\frac{1}{n} \text{tr} T(B_{(j)} - zI)^{-1} - \frac{1}{n} \text{tr} T(B_n - zI)^{-1}}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j}$$

and

$$B := \frac{1}{N} \sum_{j=1}^N \frac{r_j^*(B_{(j)} - zI)^{-1} r_j - \frac{1}{n} \text{tr} T(B_{(j)} - zI)^{-1}}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j}.$$

We first give a bound on  $A$ . Using (22) and the resolvent identity we have

$$\text{tr} T(B_n - zI)^{-1} - \text{tr} T(B_{(j)} - zI)^{-1} = \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j} r_j^*(B_{(j)} - zI)^{-1} T(B_{(j)} - zI)^{-1} r_j.$$

Using Lemma 3.8,

$$\left| \mathbb{E} \frac{1}{n} r_j^*(B_{(j)} - zI)^{-1} T(B_{(j)} - zI)^{-1} r_j \right| \leq \frac{C}{n^2} \text{tr} |B_{(j)} - zI|^{-2} \leq \frac{C}{n\eta}, \quad (102)$$

which for large  $n$  can be made smaller than any constant. Similarly,

$$\begin{aligned} \frac{1}{n^2} \sqrt{\sum_{k,l} |(T|B_{(j)} - zI|^{-2})_{k,l}|^2} &\leq \frac{C}{n^2} \sqrt{\text{tr} |B_{(j)} - zI|^{-4}} \\ &\leq \frac{C}{n^2} \sqrt{\frac{1}{\eta^2} \text{tr} |B_{(j)} - zI|^{-2}} \\ &\leq \frac{C}{(n\eta)^{3/2}}, \end{aligned}$$

which is also smaller than any constant for large  $n$ . Using the resolvent identity, Lemma B.2 of [8] and a union bound, for any constant  $\epsilon > 0$ ,

$$\begin{aligned} & \mathbb{P} \left( \max_{1 \leq j \leq N} \left| \frac{1}{n} \text{tr} T(B_n - zI)^{-1} - \frac{1}{n} \text{tr} T(B_{(j)} - zI)^{-1} \right| \geq \epsilon \right) \\ & \leq \mathbb{P} \left( \max_{1 \leq j \leq N} \left| \frac{1}{n} \text{tr} T(B_n - zI)^{-1} r_j r_j^* (B_{(j)} - zI)^{-1} - \mathbb{E} \frac{1}{n} r_j^* (B_{(j)} - zI)^{-1} T(B_{(j)} - zI)^{-1} r_j \right| \geq \frac{\epsilon}{2} \right) \\ & \leq n^{-c \log \log n}. \end{aligned} \quad (103)$$

Now we look at  $B$ . By Lemma 3.8,  $\|T(B_{(j)} - zI)^{-1}\|_{HS} \leq C \sqrt{\frac{n}{\eta}}$ . In this case, again for any constant  $\epsilon > 0$  and for  $n$  large enough,

$$\frac{(\log n)^C}{n} \|T(B_{(j)} - zI)^{-1}\|_{HS} \leq C \frac{(\log n)^C}{\sqrt{n\eta}} \leq \epsilon.$$

Therefore, by Lemma B.2 of [8]

$$\mathbb{P} \left( \max_{1 \leq j \leq n} |r_j^* (B_{(j)} - zI)^{-1} r_j - \frac{1}{n} \text{tr} T(B_{(j)} - zI)^{-1}| > \epsilon \right) \leq n^{-c \log \log n}. \quad (104)$$

If  $|1 + r_j^* (B_{(j)} - zI)^{-1} r_j| > c$ , then by setting  $\epsilon = c^2$  in (103) and (104), with probability at least  $1 - n^{-c \log \log n}$ , we have  $|z|^{-1}(1 - |A| - |B|) \geq (2\mu_+)^{-1}$ , so that, by (101),

$$\left| \frac{1}{1 + \frac{1}{n} \text{tr} T(B_n - zI)^{-1}} \right| \leq 2\mu_+ |m_n(z)|.$$

By (103) and (104), with probability at least  $1 - n^{-c \log \log n}$ ,

$$\max_{1 \leq j \leq N} \left| \frac{1}{1 + r_j^* (B_{(j)} - zI)^{-1} r_j} \right| \leq 4\mu_+ |m_n(z)|.$$

□

**Proof of Theorem 3.3** Recall that  $X_{(1)}$  denotes the  $n \times (N - 1)$  matrix obtained by removing the first column of  $X$ . Consider the component  $v_1$  of  $v$ . Following the proof of Theorem 1.2 in [3],

$$|v_1|^2 = (1 + x_1^* T^{-\frac{1}{2}} X_{(1)} (\mu - X_{(1)}^* T^{-1} X_{(1)})^{-2} X_{(1)}^* T^{-\frac{1}{2}} x_1)^{-1}. \quad (105)$$

Let  $T^{-\frac{1}{2}} X_{(1)}$  have the singular value decomposition

$$T^{-\frac{1}{2}} X_{(1)} = \sum_{k=1}^{n-1} \sqrt{\mu_k^{(1)}} u_k w_k^*.$$

When  $\mathcal{T}_1$  and  $\mathcal{X}$  occur,  $\min(\mu_1^{(1)}, \dots, \mu_{n-1}^{(1)}) \geq \frac{1}{2} a_- (1 - \frac{1}{\sqrt{b}})^2$ . As in [3], we now partition  $\mathcal{E}_\kappa^{(\mu)}$  into subintervals of length  $\eta$ . If  $\mu$  is in the interval  $I_a = [a - \frac{\eta}{2}, a + \frac{\eta}{2}]$ , then

$$(105) = \left( 1 + \sum_{k=1}^{n-1} \frac{\mu_k^{(1)} |u_k^* x_1|^2}{(\mu - \lambda_k)^2} \right)^{-1}$$

$$\begin{aligned}
&\leq \left( \sum_{k:\lambda_k \in I_a} \frac{\mu_i^{(1)} |u_k^* x_1|^2}{(\mu - \lambda_k)^2} \right)^{-1} \\
&\leq \left( \frac{1}{2} a_- \left( 1 - \frac{1}{\sqrt{b}} \right)^2 \right)^{-1} \left( \frac{1}{\eta^2} \sum_{k:\lambda_k \in I_a} |u_k^* x_1|^2 \right)^{-1}
\end{aligned}$$

so that, setting  $\xi_k = |\sqrt{n} x_1^* u_k|$ ,

$$\mathbb{P}(|v_1| > t) \leq \mathbb{P} \left( \sum_{k:\lambda_k \in I_a} \xi_k \leq \left( \frac{1}{2} a_- (1 - b^{-\frac{1}{2}})^2 \right) \frac{\eta^2 n}{t^2} \right). \quad (106)$$

The eigenvalues of  $T^{-\frac{1}{2}} X_{(1)} X_{(1)}^* T^{-\frac{1}{2}}$  and  $T^{-\frac{1}{2}} X X^* T^{-\frac{1}{2}}$  interlace (see page 82 of [2]), so that the number of eigenvalues of each matrix in an interval differs at most by two. We use Lemma 3.6 and an argument similar to the proof of Lemma 3.7 to obtain the optimal  $t$ ,  $t = \eta^{\frac{1}{2}}$ , and a bound for (106). We partition  $\mathcal{E}_\kappa^{(\mu)}$  into at most  $n$  subintervals of the form  $I_a$  and take the union bound for the terms of the form (106) corresponding to the subintervals and the  $bn$  indices.  $\square$

## 4 General MANOVA Matrices

Now we turn to the proof of Theorem 2.2, which follows easily from Theorem 3.2.

**Proof of Theorem 2.2** We recall the observations that

$$(XX^* + YY^*)^{-\frac{1}{2}} YY^* (XX^* + YY^*)^{-\frac{1}{2}} \quad (107)$$

and

$$(YY^*)^{\frac{1}{2}} (XX^* + YY^*)^{-1} (YY^*)^{\frac{1}{2}} \quad (108)$$

have the same eigenvalues and

$$(YY^*)^{\frac{1}{2}} (XX^* + YY^*)^{-1} (YY^*)^{\frac{1}{2}} = (I + (YY^*)^{-\frac{1}{2}} XX^* (YY^*)^{\frac{1}{2}})^{-1}.$$

Note that by factoring  $YY^*$  as two factors of  $(YY^*)^{\frac{1}{2}}$  rather than  $Y$  and  $Y^*$ , we avoid changing the number of zero eigenvalues. If  $(YY^*)^{-\frac{1}{2}} XX^* (YY^*)^{-\frac{1}{2}}$  has eigenvalues  $\mu_1, \mu_2, \dots, \mu_n$ , then (108), and hence (107), has eigenvalues  $\lambda_k = (1 + \mu_k)^{-1}$ ,  $k = 1, 2, \dots, n$ , i.e. the correspondence  $\lambda = (1 + \mu)^{-1}$  maps one set of eigenvalues into the other one. Since the eigenvalues of  $(YY^*)^{-\frac{1}{2}} XX^* (YY^*)^{-\frac{1}{2}}$  are positive, this correspondence is regular. The correspondence also maps the interval  $[\mu_-, \mu_+]$  to  $[\lambda_-, \lambda_+]$  and the regions  $\mathcal{E}_{\kappa, \eta}^{(\mu)}$  to  $\mathcal{E}_{\kappa, \eta}^{(\lambda)}$  and  $\mathcal{E}_\kappa^{(\mu)}$  to  $\mathcal{E}_\kappa^{(\lambda)}$ . The same transformation gives the correspondence between  $f_p$  and  $f_M$ :

$$f_p(\mu) = \frac{1}{(1 + \mu)^2} f_M \left( \frac{1}{1 + \mu} \right).$$

An easy calculation shows the relation between the Stieltjes transforms. Setting  $z' = z^{-1} - 1$ , after a change of variables, we have

$$m_p(z') = \int \frac{1}{\mu - z'} f_p(\mu) d\mu = - \int \frac{\lambda z}{\lambda - z} f_M(\lambda) d\lambda = -z - z^2 m_M(z).$$

Similarly

$$m_{n,p}(z') = \frac{1}{n} \sum_{k=1}^n \frac{1}{\mu_k - z'} = -\frac{z}{n} \sum_k \frac{\lambda_k}{\lambda_k - z} = -z - z^2 m_{n,M}(z).$$

so

$$m_{n,M}(z) - m_M(z) = -z^{-2}(m_{n,p}(z') - m_p(z')). \quad (109)$$

(Strictly speaking, we defined Stieltjes transforms  $m(z)$  for  $z \in \mathbb{C}^+$ , but the formula (6) clearly defines it for all  $z \in \mathbb{C} \setminus \mathbb{R}$  and we have  $\overline{m(z)} = m(\bar{z})$ .) Since the sets  $\mathcal{E}_\kappa^{(\mu)}$  and  $\mathcal{E}_\kappa^{(\lambda)}$  are separated away from zero, we always have a positive lower bound on  $z$  and  $z'$ . Notice that  $\Im z' = -|z|^{-2} \Im z$ , therefore  $\Im z'$  and  $\Im z$  are comparable. Thus part *i*) of Theorem 2.2 follows from (109) and part *i*) of Theorem 3.2. One can similarly conclude part *ii*) of Theorem 2.2 from part *ii*) of Theorem 3.2. This proves Theorem 2.2.  $\square$

## A Computing $m_M$

We refer to [10] for the calculation of  $m_M(z)$ . There the Stieltjes transform of

$$f(x) = \frac{a}{a+b} \delta(x) + \frac{\sqrt{(x-\lambda_-)(\lambda_+ - x)}}{2\pi x(1-x)} I_{[\lambda_-, \lambda_+]} + \left(1 - \frac{b-1}{b}\right) \delta(x-1)$$

is obtained, which we will denote  $m_f(z)$  here. The result is

$$m_f(z) = \frac{z - \frac{a+1}{a+b} + \sqrt{z^2 - (2\frac{a+1}{a+b} - \frac{a+1}{(a+b)^2})z + (\frac{a-1}{a+b})^2}}{2z(1-z)}. \quad (110)$$

This is seen by setting  $p = \frac{a}{a+b}$  and  $q = \frac{1}{a+b}$  in the notation of [10] (note, though, that the point mass at zero in  $f(x)$  has been removed in [10]). We then have that

$$m_M(z) = (a+b) \left( m_f(z) + \frac{a}{a+b} \frac{1}{z} - \frac{b-1}{a+b} \frac{1}{1-z} \right).$$

Inserting (110) into this formula, we obtain (7).

## References

- [1] Z. D. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, New York, 2010.
- [2] R. Bhatia. *Matrix analysis*. Springer-Verlag, New York, 1997.
- [3] L. Erdős, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.
- [4] L. Erdős, B. Schlein, and H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

- [5] L. Erdős, B. Schlein, and H.-T. Yau. Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not. IMRN*, (3):436–479, 2010.
- [6] L. Erdős, B. Schlein, and H.-T. Yau. Universality of random matrices and local relaxation flow. *Invent. Math.*, 185(1):75–119, 2011.
- [7] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. H. Poincaré (B), Probab. Statist.*, 48(1): 1-46, 2012.
- [8] L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Prob. Theor. Rel. Fields*, 154(1-2): 341-407, 2012.
- [9] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3): 1435-1515, 2012.
- [10] B. Farrell. Limiting empirical singular value distribution of restrictions of discrete Fourier matrices. *Journal of Fourier Analysis and Applications*, 17:733–753, 2011.
- [11] P. J. Forrester. *Log-gases and random matrices*. Princeton University Press, Princeton, NJ, 2010.
- [12] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.
- [13] R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York, 1982.
- [14] S. Péché. Universality in the bulk of the spectrum for complex sample covariance matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 48 (1):80–106, 2012.
- [15] N. S. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*. to appear.
- [16] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.
- [17] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, 55(2):175–192, 1995.
- [18] T. Tao and V. Vu. Random covariance matrices: universality of local statistics of eigenvalues. *Ann. Probab.*, 40(3):1285-1315, 2012.
- [19] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011.
- [20] K. W. Wachter. The limiting empirical measure of multiple discriminant ratios. *Ann. Stat.*, 8:937–957, 1980.
- [21] E. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.