THE ELLIPTIC LAW

HOI H. NGUYEN AND SEAN O’ROURKE

Abstract. We show that, under some general assumptions on the entries of a random complex $n \times n$ matrix $X_n$, the empirical spectral distribution of $\frac{1}{\sqrt{n}}X_n$ converges to the uniform law of an ellipsoid as $n$ tends to infinity. This generalizes the well-known circular law in random matrix theory.

1. Introduction

Let $X_n$ be a $n \times n$ matrix with complex eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. The empirical spectral measure $\mu_{X_n}$ of $X_n$ is defined as

$$\mu_{X_n} := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}$$

and the corresponding empirical spectral distribution (ESD) $F_{X_n}(x,y)$ is given by

$$F_{X_n}(x,y) := \frac{1}{n} \# \{1 \leq j \leq n : \text{Re}(\lambda_j) \leq x, \text{Im}(\lambda_j) \leq y\}.$$

Here $\#E$ denotes the cardinality of the set $E$. In the case when the eigenvalues of $X_n$ are real, we write the ESD $F_{X_n}$ as just a function of $x$,

$$F_{X_n}(x) := \frac{1}{n} \# \{1 \leq j \leq n : \lambda_j \leq x\}.$$ 

A fundamental problem in random matrix theory is to determine the limiting distribution of the ESD as the size of the matrix tends to infinity. In certain cases when the entries have special distribution, such as Gaussian, the joint distribution of the eigenvalues can be given explicitly, and so the limiting distribution can be derived directly. However, these explicit formulas are not available for many random matrix ensembles, and so the problem of finding the limiting distribution becomes much more difficult. On the other hand, the well-known universality phenomenon in random matrix theory predicts that the limiting distribution should not depend on the distribution of the entries. We give two famous examples below.

In the 1950s, Wigner studied the limiting ESD for a large class of random Hermitian matrices whose entries on or above the diagonal are independent [48]. In particular, Wigner showed that, under some additional moment and symmetry assumptions on the entries, the ESD of such a matrix converges to the semi-circular law $F_{sc}$ with density given by

$$F'_{sc}(x) := \left\{ \begin{array}{ll} \frac{1}{2\pi} \sqrt{4-x^2}, & -2 \leq x \leq 2 \\ 0, & \text{otherwise} \end{array} \right.$$
The most general form of the semi-circular law assumes only the first two moments of the entries [2].

**Theorem 1.1** (Semi-circular law for Wigner matrices). Assume that $X_n$ is a $n \times n$ Hermitian matrix whose entries on or above the diagonal are independent. Further assume that the diagonal entries are i.i.d. real random variables and those above the diagonal are i.i.d complex random variables with variance one. Then the ESD of the matrix $\frac{1}{\sqrt{n}} X_n$ converges almost surely to the semi-circular law as $n \to \infty$.

The ESD for non-Hermitian random matrices with i.i.d. entries was first studied by Mehta [27]. In particular, in the case where the entries of $X_n$ are i.i.d. complex normal random variables, Mehta showed that the ESD of $\frac{1}{\sqrt{n}} X_n$ converges, as $n \to \infty$, to the circular law $F_{\text{cir}}$ given by

$$F_{\text{cir}}(x,y) := \frac{1}{\pi} \text{mes} \left( \{ z : |z| \leq 1 \ : \ \text{Re}(z) \leq x, \ \text{Im}(z) \leq y \} \right).$$

In other words, $F_{\text{cir}}$ is the two-dimensional distribution function for the uniform probability measure on the unit disk in the complex plane.

Mehta used the joint density function of the eigenvalues of $\frac{1}{\sqrt{n}} X_n$ which was derived by Ginibre [10]. The real Gaussian case was studied by Edelman in [7]. For the general (non-Gaussian) case when there is no formula, the problem appears much more difficult. Important results were obtained by Girko [11, 12], Bai [1, 3], and more recently by Götze and Tikhomirov [18], Pan and Zhou [34], and Tao and Vu [41]. These results confirm the same limiting law under some moment or smoothness assumptions on the distribution of the entries. Recently, Tao and Vu (appendix by Krishnapur) were able to remove all these additional assumptions, establishing the law under the first two moments [42].

**Theorem 1.2** (Circular law for non-Hermitian i.i.d. matrices). Assume that the entries of the $n \times n$ matrix $X_n$ are i.i.d. copies of a complex random variable with mean zero and variance one. Then the ESD of the matrix $\frac{1}{\sqrt{n}} X_n$ converges almost surely to the circular law as $n \to \infty$.

The two celebrated results above provide a somewhat complete picture of the limiting law for the ESD of Hermitian and non-Hermitian i.i.d matrices. In the 1980s, Girko initiated a study of the limiting distribution for more general matrices which interpolate between Hermitian and non-Hermitian models.

**Definition 1.3** (Condition C0). Let $(\xi_1, \xi_2)$ be a random vector in $\mathbb{C}^2$ where both $\xi_1$ and $\xi_2$ have mean zero and unit variance. Let $\{x_{ij}\}$ be an infinite double array of random variables on $\mathbb{C}$. For each $n \geq 1$ we define the random $n \times n$ matrix $X_n = (x_{ij})_{1 \leq i,j \leq n}$. We say that the sequence of random matrices $\{X_n\}_{n \geq 1}$ satisfies condition C0 with atom variables $(\xi_1, \xi_2)$ if the following conditions hold:

(i) (Independence) $\{x_{ii} : i \geq 1\} \cup \{(x_{ij}, x_{ji}) : 1 \leq i < j\}$ is a collection of independent random elements,

(ii) (Common distribution) each pair $(x_{ij}, x_{ji})$, $1 \leq i < j$ is an i.i.d. copy of $(\xi_1, \xi_2)$,

(iii) (Flexibility of the main diagonal) the diagonal elements, $\{x_{ii} : i \geq 1\}$, are i.i.d. with mean zero and finite variance.
It is clear that many Hermitian and non-Hermitian i.i.d matrix ensembles belong to the above class. In fact, it also consists of linear combinations of independent Hermitian and non-Hermitian i.i.d. matrices.

Over the past thirty years, Girko has established a number of results for the limiting law of random matrices satisfying condition \( C_0 \). We refer the reader to [13, 14, 15, 16, 17] and the references therein. To our best understanding, Girko’s proofs are incomplete and lack rigor. The familiar reader may also relate this to Girko’s controversial works on the circular law (see the discussions in [1, 7]).

When \( \{X_n\}_{n \geq 1} \) is a sequence of random matrices that satisfy condition \( C_0 \) with jointly Gaussian atom variables \( (\xi_1, \xi_2) \), the joint eigenvalue density can be derived explicitly and the limiting ESD can be computed directly; see [21, 22, 24] and references therein. Recently, Naumov [29] has been able to verify the same limiting law for a much more general class of real random matrices whose entries have finite fourth moment.

For \(-1 < \rho < 1\), denote by \( \mathcal{E}_\rho \) the ellipsoid

\[
\mathcal{E}_\rho := \left\{ z \in \mathbb{C} : \frac{\text{Re}(z)^2}{(1-\rho)^2} + \frac{\text{Im}(z)^2}{(1+\rho)^2} \leq 1 \right\}.
\]

**Theorem 1.4** ([29]). Let \( \{X_n\}_{n \geq 1} \) be a sequence of real random matrices that satisfy condition \( C_0 \) with real atom variables \( (\xi_1, \xi_2) \) where \( E[\xi_1\xi_2] = \rho \), \(-1 < \rho < 1\). Also, assume that \( \max(\{E|\xi_1|^4, E|\xi_2|^4\}) < \infty \). Then the ESD of the matrix \( \frac{1}{\sqrt{n}}X_n \) converges in probability as \( n \to \infty \) to the elliptic law \( F_\rho \) with parameter \( \rho \) given by

\[
F_\rho(x,y) := \frac{1}{\pi(1-\rho^2)} \text{mes} \left\{ z \in \mathcal{E}_\rho : \text{Re}(z) \leq x, \text{Im}(z) \leq y \right\}.
\]

In conjunction with Theorems 1.1, 1.2 and with the universality phenomenon, it is tempting to conjecture that Theorem 1.4 should hold without any further moment assumption. One of the main goals of this paper is to resolve this conjecture for the real case.

For any matrix \( M \), we define the Hilbert-Schmidt norm \( \|M\|_2 \) by the formula

\[
\|M\|_2 := \sqrt{\text{tr}(M^*M)} = \sqrt{\text{tr}(MM^*)}.
\]

**Theorem 1.5** (Elliptic law for real random matrices). Let \( \{X_n\}_{n \geq 1} \) be a sequence of real random matrices that satisfy condition \( C_0 \) with real atom variables \( (\xi_1, \xi_2) \) where \( E[\xi_1\xi_2] = \rho \), \(-1 < \rho < 1\). Assume that \( \{F_n\}_{n \geq 1} \) is a sequence of deterministic matrices such that \( \text{rank}(F_n) = o(n) \) and \( \sup_n \frac{1}{n^2} \|F_n\|_2^2 < \infty \). Then the ESD of \( \frac{1}{\sqrt{n}}(X_n + F_n) \) converges almost surely to the elliptic law with parameter \( \rho \) as \( n \to \infty \).

In fact, we are able to extend Theorem 1.5 to the following more general setting.

**Definition 1.6** (\( (\mu, \rho) \)-family). Given parameters \( 0 \leq \mu \leq 1 \) and \(-1 < \rho < 1\), we say that the complex random variable pair \( (\xi_1, \xi_2) \) belongs to the \( (\mu, \rho) \)-family if the following holds.

(i) Both \( \xi_1 \) and \( \xi_2 \) have mean zero and unit variance;
In this optimal setting, the ESD of \(1\)

Overview and Outline.

1.8. As customary, we use \(a\) to denote the singular values of \(\theta\) when \(n \to \infty\). For a given \(0 \leq a \leq b\), we write \(a \ll b\) or \(b \gg a\). If \(a = \Omega(b)\) and \(b = \Omega(a)\), we write \(a \asymp b\).

Notice that if \((\xi_1, \xi_2)\) belongs to a \((\mu, \rho)\)-family then \(E[|\xi_1|^2] = E[|\xi_2|^2] = 1\) and \(E[\xi_1 \xi_2] = \rho\). More importantly, we do not require the imaginary and real parts of \(\xi_1, \xi_2\) to be independent.

**Theorem 1.7** (Elliptic law for complex random matrices). Let \(0 \leq \mu \leq 1 \) and \(-1 < \rho < 1\) be given. Let \(\{\mathbf{X}_n\}_{n \geq 1}\) be a sequence of complex matrices such that \(\mathbf{X}_n\) satisfies condition C0 with atom variables \((\xi_1, \xi_2)\) from the \((\mu, \rho)\)-family. Assume furthermore that \(\{F_n\}_{n \geq 1}\) is a sequence of deterministic matrices such that \(\text{rank}(F_n) = o(n)\) and \(\sup_n \frac{1}{n^2} \|F_n\|_2^2 < \infty\). Then the ESD of \(\frac{1}{\sqrt{n}}(\mathbf{X}_n + F_n)\) converges almost surely to the elliptic law with parameter \(\rho\) as \(n \to \infty\).

In light of the universality phenomenon, we conjecture that Theorem 1.7 continues to hold when \(E[|\xi_1|^2] = E[|\xi_2|^2] = 1\) and \(E[\xi_1 \xi_2] = \rho\), where \(\rho\) is a complex number satisfying \(|\rho| < 1\). In this optimal setting, the ESD of \(\frac{1}{\sqrt{n}}\mathbf{X}_n\) is conjectured to converge to the elliptic law associated with the rotated ellipsoid \(\mathcal{E}_\rho\) given by

\[
\mathcal{E}_\rho := \left\{ z \in \mathbb{C} : \frac{(\text{Re}(z) \cos \frac{\theta}{2} - \text{Im}(z) \sin \frac{\theta}{2})^2}{(1 - |\rho|)^2} + \frac{(\text{Re}(z) \sin \frac{\theta}{2} + \text{Im}(z) \cos \frac{\theta}{2})^2}{(1 + |\rho|)^2} \leq 1 \right\},
\]

where \(\theta = \text{Arg}(\rho)\).

1.8. Overview and Outline. For a \(n \times n\) matrix \(M\) we let

\[\sigma_1(M) \geq \sigma_2(M) \geq \cdots \geq \sigma_n(M) \geq 0\]

denote the singular values of \(M\). In particular, \(\sigma_n(M)\) is the least singular value of \(M\).

As in the proof of the circular law \([11]\), the main difficulty in proving Theorems 1.5 and 1.7 is controlling the least singular value of \(\mathbf{X}_n + F_n\). Theorem 2.1 below gives a lower bound for the least singular value motivated by the work of Tao and Vu \([11]\). Because of its importance, we prove Theorem 2.1 first; the proof is contained in Sections 2–6. Section 7 is dedicated to proving our main results, Theorems 1.5 and 1.7.

1.9. Notation. For a matrix \(M\) we use the notations \(r_i(M)\) and \(c_j(M)\) to denote its \(i\)-th row vector and its \(j\)-th column vector respectively; we use the notation \((M)_{ij}\) and \(M_{ij}\) to denote its \((i, j)\) entry. We let \(\|M\|_2\) denote the Hilbert-Schmidt norm of \(M\) (defined in \([11]\)) and let \(\|M\|\) denote the spectral norm of \(M\).

Here and later, asymptotic notations such as \(O, \Omega, \Theta, \omega, \) and so forth, are used under the assumption that \(n \to \infty\). A notation such as \(O_C(\cdot)\) emphasizes that the hidden constant depends on \(C\). If \(a = \Omega(b)\), we write \(b \ll a\) or \(a \gg b\). If \(a = \Omega(b)\) and \(b = \Omega(a)\), we write \(a \asymp b\).

As customary, we use \(\eta\) to denote random Bernoulli variables (thus \(\eta\) takes values \(\pm 1\) with probability \(1/2\)). For a given \(0 \leq \mu \leq 1\), we use \(\eta^{(\mu)}\) to denote random Bernoulli variables
of parameter $\mu$ (thus $\eta(\mu)$ takes values $\pm 1$ with probability $\mu/2$ and zero with probability $1 - \mu$).

We write a.s., a.a., and a.e. for almost surely, Lebesgue almost all, and Lebesgue almost everywhere respectively.

We use $\sqrt{-1}$ to denote the imaginary unit and reserve $i$ as an index.

2. The least singular value problem

One of the main ingredients to prove Theorems 1.5 and 1.7 is the following polynomial bound on the least singular value.

**Theorem 2.1** (Bound on the least singular value for perturbed random matrices). Assume that $M_n = F_n + X_n$, where the entries of the given complex matrix $F_n$ are bounded by $n^\alpha$ in absolute value, and $X_n$ is a random matrix from Theorem 1.7 for given $0 \leq \mu \leq 1$ and $-1 < \rho < 1$. Then for any $B > 0$, there exists $A > 0$ depending on $B, \alpha, \mu, \rho$ such that

$$
P(\sigma_n(M_n) \leq n^{-A}) \leq n^{-B}.
$$

We emphasize that our polynomial bound here is motivated by [41, Lemma 4.1] of Tao and Vu, which plays a fundamental role in their establishment of the circular law. We also refer the reader to the work [36] of Rudelson and Vershynin for an almost complete treatment for the least singular values of random non-Hermitian matrices. Recently, a similar study for random real symmetric matrices has been carried out independently by Vershynin in [47] and by the first author in [32]. In this paper we choose to follow [32] simply because our goal is to proving the universality for a broad range of random matrices. Nevertheless, because the matrix $X_n$ under consideration is much more complicated than symmetric or Hermitian ones, it is of great necessity to generalize and string a series of previous results [33], [31] and [32] altogether here. As a result, our ideas will not be fully original but a highly non-trivial generalization of existing ones. The rest of this section is devoted to sketching the approach, more details of the proofs will be presented subsequently.

For the sake of simplicity, we will prove our result under the following condition.

**Condition 1.** With probability one, $|x_{ij}| \leq n^{B+1}$ for all $i, j$.

In fact, because all $x_{ij}$ have bounded variance, we have $\mathbb{P}(|x_{ij}| \geq n^{B+1}) = O(n^{-2B-2})$. Thus, we can assume that $|x_{ij}| \leq n^{B+1}$ at the cost of an additional negligible term $o(n^{-B})$ in probability.

We next assume that $\sigma_n(M_n) \leq n^{-A}$. Thus $M_nx = y$ for some $\|x\|_2 = 1$ and $\|y\|_2 \leq n^{-A}$. There are two cases to consider.

2.2. **Case 1.** $M_n$ has full rank. This is the main case to consider as most of random matrices are non-singular with very high probability.
Let $C(M_n) = (c_{ij}(M_n))$, $1 \leq i, j \leq n$, be the matrix of the cofactors of $M_n$. By definition, $C(M_n)y = \det(M_n) \cdot x$, and thus we have $\|C(M_n)y\|_2 = |\det(M_n)|$.

By paying a factor of $n$ in probability, without loss of generality we can assume that the first component of $C(M_n)y$ is greater than $\det(M_n)/n^{1/2}$,

$$|c_{11}(M_n)y_1 + \ldots + c_{1n}(M_n)y_n| \geq |\det(M_n)|/n^{1/2}. \quad (2)$$

Note that $\|y\|_2 \leq n^{-A}$, it thus follows

$$\sum_{j=1}^{n} |c_{ij}(M_n)|^2 \geq n^{2A-1} |\det(M_n)|^2. \quad (3)$$

For $j \geq 2$, we write

$$c_{1j}(M_n) = \sum_{i=2}^{n} m_{i1} c_{ij}(M_{n-1}),$$

where $M_{n-1}$ is the matrix obtained from $M_n$ by removing its first row and first column, and $c_{ij}(M_{n-1})$ are the corresponding cofactors of $M_{n-1}$, and $m_{ij}$ are the entries of $M_n$.

Hence, by the Cauchy-Schwarz inequality, by Condition 1, and by the bounds $f_{ij} \leq n^\alpha$ for the entries of $F_n$, we have

$$|c_{1j}(M_n)|^2 \leq \sum_{i=2}^{n} |m_{i1}|^2 \sum_{i=2}^{n} |c_{ij}(M_{n-1})|^2 \leq n^{2B+2\alpha+3} \sum_{i=2}^{n} |c_{ij}(M_{n-1})|^2. \quad (4)$$

Similarly, for $j = 1$ we write $c_{11}(M_n) = \sum_{i=2}^{n} m_{i2} c_{i2}(M_{n-1})$, and thus,

$$|c_{11}(M_n)|^2 \leq n^{2B+2\alpha+3} \sum_{i=2}^{n} |c_{i2}(M_{n-1})|^2. \quad (5)$$

It follows from (3), (4) and (5) that

$$\sum_{2 \leq i,j \leq n} |c_{ij}(M_{n-1})|^2 \geq n^{2A-2B-2\alpha-4} |\det(M_n)|^2.$$

Hence, for proving Theorem 2.1 it suffices to justify the following result.
**Theorem 2.3.** For any $B > 0$, there exists $A > 0$ such that

$$\mathbf{P}\left(\left(\sum_{2 \leq i,j \leq n} |c_{ij}(M_{n-1})|^2\right)^{1/2} \geq n^A |\det(M_n)| \right) \leq n^{-B}.$$ 

To see why the assumption $(\sum_{2 \leq i,j \leq n} |c_{ij}(M_{n-1})|^2)^{1/2} \geq n^A |\det(M_n)|$ is useful, we next express $\det(M_n)$ as a bilinear form of its first row and column,

$$\det(M_n) = c_{11}(M_n)m_{11} + \sum_{2 \leq i,j \leq n} c_{ij}(M_{n-1})m_{1i}m_{j1}.$$ 

In other words, with $c := (\sum_{2 \leq i,j \leq n} |c_{ij}(M_{n-1})|^2)^{1/2}$ (which is nonzero as $M_n$ has full rank) and with $a_{ij} := c_{ij}(M_{n-1})/c$ we have

$$\frac{1}{c} \det(M_n) = \frac{1}{c} m_{11}c_{11}(M_n) + \sum_{2 \leq i,j \leq n} a_{ij}m_{1i}m_{j1}.$$ 

Intuitively, if we condition on $M_{n-1}$ and $m_{11}$, then the right hand side of (6), as a bilinear form of the random variables $x_{1i}, x_{i1}, 2 \leq i$, is comparable to 1 in absolute value with probability extremely close to one. Thus the assumption $\mathbf{P}(|\det(M_n)|/c \leq n^{-A}) \geq n^{-B}$ of Theorem 2.3 with appropriately large $A$, must yield a high cancelation of the bilinear form.

Basing on this intuition, our rough approach will consist of two main steps below.

- **Step 1** (Inverse step). Assume that for appropriately large $A$ we have

$$\mathbf{P}_{x_{11},\ldots,x_{1n},x_{21},\ldots,x_{n1}}\left(\left(c_{11}(M_n)/c\right)m_{11} + \sum_{2 \leq i,j \leq n} a_{ij}m_{1i}m_{j1} \leq n^{-A}|M_{n-1}|\right) \geq n^{-B}.$$ 

Then there must be a strong structure among the cofactors $c_{ij}$ of $M_{n-1}$.

- **Step 2** (Counting step). The probability, with respect to $M_{n-1}$, that there is a strong structure among the $c_{ij}$ is negligible.

Before stating the steps above in greater detail, we pause to introduce the structure appearing in our analysis.

A set $Q \subset \mathbf{C}$ is a *generalized arithmetic progression* (GAP) of rank $r$ if it can be expressed as in the form

$$Q = \{g_0 + k_1g_1 + \cdots + k_rg_r | k_i \in \mathbf{Z}, K_i \leq k_i \leq K'_i \text{ for all } 1 \leq i \leq r\}$$

for some $\{g_0, \ldots, g_r\}, \{K_1, \ldots, K_r\}$ and $\{K'_1, \ldots, K'_r\}$. 
It is convenient to think of $Q$ as the image of an integer box $B := \{(k_1, \ldots, k_r) \in \mathbb{Z}^r | K_i \leq k_i \leq K'_i\}$ under the linear map

$$\Phi : (k_1, \ldots, k_r) \mapsto g_0 + k_1g_1 + \cdots + k_r g_r.$$ 

The numbers $g_i$ are the generators of $Q$, the numbers $K'_i$ and $K_i$ are the dimensions of $Q$. We say that $Q$ is proper if this map is one to one, or equivalently if $|Q| = |B|$. For non-proper GAPs, we of course have $|Q| < |B|$. If $-K_i = K'_i$ for all $i \geq 1$ and $g_0 = 0$, we say that $Q$ is symmetric.

We refer the reader to Sections 3 and 4 for further explanation as to why GAPs is the right object to study here. In the sequel we state our main steps rigorously with the help of GAPs.

**Theorem 2.4** (Step 1). Let $0 < \epsilon < 1$ be given constant. Assume that

$$\sup_a P_{x_2, \ldots, x_n, x'_2, \ldots, x'_n} \left( \left| \sum_{2 \leq i, j \leq n} a_{ij} (x_i + f_i)(x'_j + f'_j) - a \right| \leq n^{-A} \right) \geq n^{-B}$$

for some sufficiently large integer $A$, where

- $a_{ij} = c_{ij}(M_{n-1})/c$,
- $f_i = f_{i1}, f'_i = f_{i1}$ are the entries of $F_n$, and thus fixed,
- $(x_i, x'_i)$ are i.i.d copies of $(\xi_1, \xi_2)$ of a given $(\mu, \rho)$-family with $0 \leq \mu \leq 1$ and $-1 < \rho < 1$.
- any collection $x_{i_1}, \ldots, x_{i_k}, x'_{j_1}, \ldots, x'_{j_l}$ of random variables are mutually independent as long as the indices $i_1, \ldots, i_k, j_1, \ldots, j_l$ are distinct.

Then there exists a complex vector $\mathbf{u} = (u_1, \ldots, u_{n-1})$ which satisfies the following properties.

- (orthogonality) $\|\mathbf{u}\|_2 \asymp 1$ and either $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1}) \rangle| \leq n^{-A/2 + O_{B,\epsilon}(1)}$ for $n - O_{B,\epsilon}(1)$ rows of $M_{n-1}$ or $|\langle \mathbf{u}, \mathbf{c}_i(M_{n-1}) \rangle| \leq n^{-A/2 + O_{B,\epsilon}(1)}$ for $n - O_{B,\epsilon}(1)$ columns of $M_{n-1}$;
- (additive structure) there exists a generalized arithmetic progression $Q$ of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ that contains at least $n - 2n^\epsilon$ components $u_i$;
- (controlled form) all the components $u_i$, and all the generators of the generalized arithmetic progression are rational complex numbers of the form $\frac{p}{q} + \sqrt{-1} \frac{p'}{q'}$, where $|p|, |q|, |p'|, |q'| \leq n^{A/2 + O_{B,\epsilon}(1)}$.

In the second step of the approach, we show that the probability for $M_{n-1}$ having the above properties is negligible.

**Theorem 2.5** (Step 2). With respect to $M_{n-1}$, the probability that there exists a vector $\mathbf{u}$ as in Theorem 2.4 is $\exp(-\Omega(n))$. 
2.6. **Case 2.** $M_n$ does not have full rank, which is the case to consider if $\xi_1, \xi_2$ have discrete distribution. We show that this event holds with probability at most $n^{-B}$.

First, instead of the entries $x_{ij}$ of $X_n$, consider $x'_{ij} := (1 - \epsilon^2)x_{ij} + \epsilon \xi_{ij}$, where $\xi_{ij}$ are independently uniform on the interval $[-1, 1]$ and $\epsilon$ is very small, say $n^{-1000A}$. It is clear that the continuous matrix $M'_n = X'_n + F_n$, where $X'_n$ is formed by the $x'_{ij}$ above, has full rank with probability one. By applying Theorem 2.1 obtained from Case 1 for the matrix $M'_n$, with probability at least $1 - n^{-B}$ one has

$$|\det(M'_n)| \geq n^{-A}, \quad (7)$$

Next, because $M'_n = M_n - \epsilon(e x_{ij} + \xi_{ij})$ and as $|x_{ij}| \leq n^B + 1$, by Brunn-Minkowski inequality and Hadamard’s bound we have

$$|\det(M'_n)| \leq (|\det(M_n)|^{1/n} + O(n^{-500A}))^n,$$

where we use the fact that $A$ is chosen sufficiently large compared to $B$.

Combining with (7), we then infer that $|\det(M_n)| \geq n^{-(1 + o(1))A}$, and thus $\det(M_n) \neq 0$ with probability at least $1 - n^{-B}$, concluding the treatment for this case.

The proof of Theorem 2.4 will be given in Section 5 thanks to useful tools from Sections 3 and 4. Our goal now is to focus on linear forms.

3. **Anti-concentration, a warm-up**

Recall that in the inverse step, Theorem 2.4, we assumed that

$$\sup_a P_{x_2, \ldots, x_n, x'_2, \ldots, x'_n} (| \sum_{2 \leq i, j \leq n} a_{ij}(x_i + f_i)(x'_j + f'_j) - a| \leq n^{-A}) \geq n^{-B}. \quad (8)$$

This can be considered as a high concentration of the bilinear form $\sum_{2 \leq i, j \leq n} a_{ij}(x_i + f_i)(x'_j + f'_j)$ on a small ball of radius $n^{-A}$, where $x_i$ and $x'_i$ are not necessarily jointly independent. The main idea to extract this bit of information is to relate it to a high concentration of an appropriate linear form. This step is postponed until Section 4. Our goal now is to focus on linear forms.

A classical result of Erdős [9] and Littlewood-Offord [26] in the 1940s asserts that if $a_i$ are complex numbers of magnitude $|a_i| \geq 1$, then the probability that the linear form $\sum_{i=1}^n a_i x_i$ concentrates on a disk of radius one is of order $O(n^{-1/2})$, where $x_i$ are i.i.d. copies of a Bernoulli random variable. Recently, motivated by inverse theorems from additive combinatorics, Tao and Vu studied the underlying reason as to why the concentration probability of $\sum_{i=1}^n a_i x_i$ on a small ball is large. They call this the inverse Littlewood-Offord problem. A closer look at the definition of generalized arithmetic progressions defined in
Section 2 reveals that if \( a_i \) are very close to the elements of a GAP of rank \( O(1) \) and size \( n^{O(1)} \), then the probability that \( \sum_{i=1}^{n} a_i x_i \) concentrates on some small ball is of order \( n^{-O(1)} \), where \( x_i \) are i.i.d. copies of a Bernoulli random variable.

It was shown implicitly by Tao and Vu in \([38, 41, 44]\) that these are essentially the only examples that have high concentration probability. An explicit and somewhat optimal version has been proved in a recent paper by the first author and Vu in \([33]\). Before stating this result, we pause to introduce some terminology.

We say that a real random variable \( \xi \) is anti-concentrated if there exist positive constants \( \alpha_1, \alpha_2, \alpha_3 \) such that \( P(\alpha_1 < |\xi - \xi'| < \alpha_2) \geq \alpha_3 \), where \( \xi' \) is an i.i.d. copy of \( \xi \). (Note that the requirement of anti-concentration is somewhat weaker than having mean zero and unit variance.) We say that a complex number \( a \in C \) is \( \delta \)-close to a set \( Q \subset C \) if there exists \( q \in Q \) such that \( |a - q| \leq \delta \).

**Theorem 3.1** (Inverse Littlewood-Offord theorem for linear forms, \([33]\)). Let \( 0 < \epsilon < 1 \) and \( B > 0 \). Let \( \beta > 0 \) be an arbitrary real number that may depend on \( n \). Suppose that \( \sum_{i=1}^{n} |a_i|^2 = 1 \), and

\[
\sup_a \mathbb{P}_x \left( \left| \sum_{i=1}^{n} a_i (x_i + f_i) - a \right| \leq \beta \right) = \gamma \geq n^{-B},
\]

where \( x = (x_1, \ldots, x_n) \), and \( x_i \) are i.i.d. copies of a real random variable \( \xi \) satisfying the anti-concentration condition. Then, for any number \( n' \) between \( n \) and \( n \), there exists a proper symmetric GAP \( Q = \{ \sum_{i=1}^{r} k_i g_i : k_i \in \mathbb{Z}, |k_i| \leq L_i \} \) such that

(i) (control of rank and size) \( Q \) has small rank, \( r = O_{B,\epsilon}(1) \), and small cardinality \( |Q| \leq \max \left( O_{B,\epsilon}(\sqrt{\frac{\gamma^{-1}}{n'}}), 1 \right) \);

(ii) (control of the steps) there is a non-zero integer \( p = O_{B,\epsilon}(\sqrt{n'}) \) such that all steps \( g_i \) of \( Q \) have the form \( g_i = \beta \frac{p}{p_i} \), with \( p_i \in \mathbb{Z} \) and \( p_i = O_{B,\epsilon}(\beta^{-1}\sqrt{n'}) \).

(iii) (good approximation) at least \( n - n' \) elements of \( a_i \) are \( \beta \)-close to \( Q \);

Here the implied constants are allowed to depend on \( \alpha_1, \alpha_2 \) and \( \alpha_3 \). The interested reader is also invited to read \([36]\) for a similar but milder setting of the inverse Littlewood-Offord for linear forms.

To attack Theorem 2.4, the first step is to study the concentration of a more general linear form \( \sum_{i} a_i x_i + b_i x_i' \), where \( (x_i, x_i') \) are i.i.d copies of a pair random complex variables \( (\xi_1, \xi_2) \) from a given \( (\mu, \rho) \)-family. Intuitively, as \( \mathbb{E} |\xi_1|^2 = \mathbb{E} |\xi_2|^2 = 1 \) and \( |\mathbb{E} \xi_1 \xi_2| = |\rho| < 1 \), the random variables \( \xi_1 \) and \( \xi_2 \) are not totally dependent on each other. (See for instance Claim A.2 of Appendix A for a more precise statement.) This fact may suggest a way to apply Theorem 3.1 with respect to \( x_2, \ldots, x_n \) while holding \( x_2', \ldots, x_n' \) "fixed", and vice versa. One of the main results is to justify this intuition.
Theorem 3.2 (Inverse Littlewood-Offord theorem for mixing linear forms). Let $0 \leq \mu \leq 1, -1 < \rho < 1$ and $0 < \epsilon < 1, B > 0$ be given. Let $\beta > 0$ be an arbitrary real number that may depend on $n$. Suppose that $a_i, b_i \in \mathbb{C}$ such that $\sum_{i=1}^{n} |a_i|^2 + \sum_{i=1}^{n} |b_i|^2 = 1$ and

$$\sup_{a} \mathbb{P}_{\mathbf{x}, \mathbf{x}'} \left( \left| \sum_{i=1}^{n} (a_i x_i + b_i x'_i) - a \right| \leq \beta \right) = \gamma \geq n^{-B},$$

where

- $(x_i, x'_i)$ are i.i.d copies of $(\xi_1, \xi_2)$ from a given $(\mu, \rho)$-family,
- any collection $x_{i_1}, \ldots, x_{i_k}, x'_{i_1}, \ldots, x'_{i_k}$ of random variables are mutually independent as long as the indices $i_1, \ldots, i_k, j_1, \ldots, j_k$ are distinct.

Then there exist positive constants $\alpha, c_0, C_0$ depending on $(\xi_1, \xi_2)$ and two pairs of complex numbers $(c_1, c_2)$ and $(c'_1, c'_2)$ (which may depend on $n$) such that

- $|c_1|, |c_2|, |c'_1|, |c'_2|$ are bounded from below and above by $c_0$ and $C_0$ respectively,
- $|c_1/c_2 - c'_1/c'_2| > \alpha$,
- for any number $n'$ between $n^\epsilon$ and $n$, there exists a proper symmetric GAP $Q = \{ \sum_{i=1}^{r} k_i g_i : k_i \in \mathbb{Z}, |k_i| \leq L_i \} \subset \mathbb{C}$ whose parameters satisfy (i) and (ii) of Theorem 3.1 and for at least $n - n'$ indices $i$, the pairs $c_1 a_i + c_2 b_i, c'_1 a_i + c'_2 b_i$ are $\beta$-close to $Q$.

As Theorem 3.2 can be shown by modifying the proof of Theorem 3.1 from [33], we postpone its proof until Appendix A. We now introduce several useful corollaries of it.

Firstly, by choosing $b_i = 0$, Theorem 3.2 immediately implies the following version of Theorem 3.1.

**Corollary 3.3.** The conclusion of Theorem 3.1 also holds if $x_i$ are i.i.d. copies of a complex random variable $\xi$ satisfying $\mathbb{E}|\xi|^2 = 1$ and $\mathbb{E} \xi = \mathbb{E} \text{Im}(\xi) \text{Re}(\xi) = 0$.

Secondly, if we $\beta$-approximate the components of $c_i, c'_i$ by rational numbers of the form $p/q, |p|, |q| = O(\beta)$, then we obtain the following.

**Corollary 3.4.** Assume as in Theorem 3.2. Then there exist two pairs of complex numbers $(c_1, c_2)$ and $(c'_1, c'_2)$ for which $|c_i|, |c'_i|$ are bounded from below and above by $c_0$ and $C_0$, $|c_1/c_2 - c'_1/c'_2| > \alpha$, and the components of $c_i, c'_i$ are rational numbers of the form $p/q, |p|, |q| = O(\beta)$ such that for any number $n'$ between $n^\epsilon$ and $n$, there exists a proper symmetric GAP $Q = \{ \sum_{i=1}^{r} k_i g_i : k_i \in \mathbb{Z}, |k_i| \leq L_i \} \subset \mathbb{C}$ whose parameters satisfy (i) and (ii) of Theorem 3.1 and for at least $n - n'$ indices $i$ the following holds:

- $a_i$ are $O(\beta)$-close to the GAP $P_1 := \frac{c_2}{c_1 c'_2 - c_1 c_2} \cdot Q + \frac{c'_2}{c_1 c'_2 - c_1 c_2} \cdot Q$;
- $b_i$ are $O(\beta)$-close to the GAP $P_2 := \frac{c_1}{c_1 c'_2 - c_1 c_2} \cdot Q + \frac{c'_1}{c_1 c'_2 - c_1 c_2} \cdot Q$;
• consequently, \(a_i\) and \(b_i\) are \(O(\beta)\)-close to the combined \(GAP\) \(P = c_1 c_2 \cdot Q + \frac{c_1 c_2'}{c_1 c_2 - c_1 c_2'} \cdot Q = \frac{c_1 c_2'}{c_1 c_2 - c_1 c_2'} \cdot Q + \frac{c_1 c_2'}{c_1 c_2 - c_1 c_2'} \cdot Q + \frac{c_1 c_2'}{c_1 c_2 - c_1 c_2'} \cdot Q\).

Notice that the rank of \(P\) is \(O_{\beta, \epsilon}(1)\) and the size of \(P\) is \(n^{O_B(1)}\). Roughly speaking, the fact that the parameters involved in \(P\) are rational numbers will enable us to control the number of such \(GAPs\) easily. We will exploit this pleasant fact in more details in Sections 4 and 6.

We remark that the assumption \(-1 < \rho < 1\) is necessary because Theorem 3.2 is not valid for the boundary case \(|\rho| = 1\). For instance, if \(\xi\) is a symmetric random variable and if \(x'_i = -x_i\) (in which case \(\rho = -1\)), then the assumption \(P_{x, x'}(\|\sum_{i=1}^n a_i x_i + b_i x'_i - a| \leq \beta) \geq n^{-B}\) is equivalent to \(P_x(\|\sum_{i=1}^n (a_i - b_i) x_i - a| \leq \beta) \geq n^{-B}\). From here, only information for the \(a_i - b_i\) can be deduced but not for the individual \(a_i\) and \(b_i\) separately.

Finally, the conclusion of Theorem 3.2 is somewhat optimal. Indeed, assume that there exist \((c_1, c_1'), (c_2, c_2')\) with \(c_1^2 + c_2^2 = c_1'^2 + c_2'^2 = 1\) and \(c_1 c_2' \neq c_1' c_2\) such that \(\xi_1 = c_1 \psi_1 + c_1' \psi_2\) and \(\xi_2 = c_2 \psi_1 + c_2' \psi_2\), where \(\psi_2\) is an independent copy of \(\psi_1\). Then the assumption \(P_{x, x'}(\|\sum_{i=1}^n (a_i x_i + b_i x'_i - a| \leq \beta) \geq n^{-B}\) becomes \(P_{\psi_1}(\|\sum_{i=1}^n (c_1 a_i + c_2 b_i) \psi_1 + (c_1' a_i + c_2' b_i) \psi_2 - a| \leq \beta) \geq n^{-B}\). So, as \(\psi_{ij}\) are independent, structural information for \(c_1 a_i + c_2 b_i\) and \(c_1' a_i + c_2' b_i\) can be deduced using Theorem 3.1 as in the same way we concluded using Theorem 3.2.

4. Anti-concentration of bilinear forms

We will next apply Corollary 3.4 to infer an inverse version for the concentration of the bilinear form \(\sum_{1 \leq i, j \leq n} a_{ij} (x_i + f_i)(x'_j + f'_j)\) appeared in Theorem 2.4.

**Theorem 4.1.** Let \(0 < \epsilon < 1, |\rho| < 1\) and \(B > 0\) be given. Let \(\beta > 0\) be an arbitrary real number that may depend on \(n\). Assume that \(\sum_{i, j} |a_{ij}|^2 = 1\) and

\[
\sup_{a} P_{x, x'} (\|\sum_{1 \leq i, j \leq n} a_{ij} (x_i + f_i)(x'_j + f'_j) - a| \leq \beta) = \gamma \geq n^{-B},
\]

- \((x_i, x'_i)\) are i.i.d copies of \((\xi_1, \xi_2)\) from a given \((\mu, \rho)\)-family
- any collection \(x_{i_1}, \ldots, x_{i_k}, x'_{j_1}, \ldots, x'_{j_l}\) of random variables are mutually independent as long as the indices \(i_1, \ldots, i_k, j_1, \ldots, j_l\) are distinct.

Then, there exist an integer \(k \neq 0, |k| = n^{O_B(1)}\), a set of \(r = O(1)\) rows \(r_{i_1}, \ldots, r_{i_r}\) of the array \(A_n = (a_{ij})_{1 \leq i, j \leq n}\), and set \(I\) of size at least \(n - 2n^\epsilon\) such that for each \(i \in I\), there exists integers \(k_{ii_1}, \ldots, k_{ii_r}\), all bounded by \(n^{O_B(1)}\), such that the following holds.

\[
P_{z} (\|z, kr (A_n) + \sum_{j=1}^r k_{ij} r_{ij} (A_n)\| \leq \beta n^{O_B(1)}) \geq n^{-O_B(1)}, \tag{9}
\]
where \( z = (z_1, \ldots, z_n) \) and \( z_i \) are i.i.d. copies of \( \eta^{(1/2)}(\xi_2 - \xi'_2) \), where \( \xi'_2 \) is an i.i.d. copy of \( \xi_2 \) and \( \eta^{(1/2)} \) is a Bernoulli random variable of parameter 1/2 independent of \( \xi_2 \) and \( \xi'_2 \).

We remark that this result is an analogue of \cite[Theorem 1.8]{31} in which case we studied the concentration of the quadratic forms of type \( \sum_{i,j} a_{ij} x_i x_j \). It seems plausible that after an appropriate linear transform we can trap most of the entries \( a_{ij} \) of \( A_n \) into a GAP of small size and small rank (in the spirit of \cite{30}). However, we do not proceed this matter here. Roughly speaking, in application to justify Theorem \ref{thm:main} we just need the conclusion of Theorem \ref{thm:main} for only one row.

To prove \ref{thm:main}, we will follow the machinery from \cite{31} with some extra twists. As the first step, we free the dependencies between \( x \) and \( x' \).

4.2. Decoupling lemma. Let \( U \) be an arbitrary subset of \( \{1, \ldots, n\} \) such that both of \( U \) and \( \bar{U} \) are of size \( \Theta(n) \). Let \( A_U \) be a matrix of size \( n \) by \( n \) defined as

\[
A_U(ij) = \begin{cases} 
    a_{ij} & \text{if } i \in U \text{ and } j \in \bar{U} \text{ or } i \in \bar{U} \text{ and } j \in U, \\
    0 & \text{otherwise},
\end{cases}
\]

where we denoted by \( A_U(ij) \) the \( ij \) entry of \( A_U \). We prove the following lemma by a series applications of Cauchy-Schwarz inequality.

**Lemma 4.3.** Assume that

\[
\gamma = \sup_{a,b,b'_i} P_{x,x'} \left( \left| \sum_{i,j} a_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a \right| \leq \beta \right) \geq n^{-B},
\]

where \( x, x' \) are defined as in Theorem \ref{thm:main}. Then,

\[
P_{v,w} \left( \left| \sum_{1 \leq i,j \leq n} A_U(ij) v_i w_j \right| = O_B(\beta \sqrt{\log n}) \right) = \Theta(\gamma^4), \tag{10}
\]

where \( v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \), and

- \((v_i, w_i)\) are i.i.d copies of a vector \((\xi_1 - \xi'_1, \xi_2 - \xi'_2)\), where \((\xi'_1, \xi'_2)\) is an independent copy of \((\xi_1, \xi_2)\),
- any collection \(v_{i_1}, \ldots, v_{i_k}, w'_{j_1}, \ldots, w'_{j_l}\) of random variables are mutually independent as long as the indices \(i_1, \ldots, i_k, j_1, \ldots, j_l\) are distinct.

An advantage of considering the sum \( \sum_{1 \leq i,j \leq n} A_U(ij) v_i w_j \) over the original form \( \sum_{i,j} a_{ij} x_i x'_j \) is that we can rewrite the former as \( \sum_{i \in U} (\sum_{j \in \bar{U}} a_{ij} y_{j}) x_i + \sum_{i \in \bar{U}} (\sum_{j \in U} a_{ji} x_{j}) y_{i} \). Thus, if all \( x_j, y_j, j \in \bar{U} \) are held fixed, Theorem \ref{thm:decoupling} applied to \eqref{10} allows us to extract useful information on \( \sum_{j \in \bar{U}} a_{ji} y_{j} \) and \( \sum_{j \in U} a_{ji} x_{j} \). As the proof of Lemma \ref{lem:decoupling} is standard, we postpone it until Appendix \ref{sec:appendix3}.
We next apply Theorem 3.2 to obtain the following key structure for the entries of $A_U$.

**Lemma 4.4.** There exist a set $I_0(U)$ of size $O_{B,\epsilon}(1)$ and a set $I(U)$ of size at least $n - n^\epsilon$, and a nonzero integer $k(U)$ bounded by $n^{O_{B,\epsilon}(1)}$ such that for any $i \in I$, there are integers $k_{i_0}(U), i_0 \in I_0(U)$, all bounded by $n^{O_{B,\epsilon}(1)}$, such that

$$
P_y\left(|(k(U)r_i(A_U) + \sum_{i_0 \in I_0} k_{i_0}(U)r_{i_0}(A_U)) \cdot y| \leq \beta n^{O_{B,\epsilon}(1)}\right) = n^{-O_{B,\epsilon}(1)},$$

where $y = (y_1, \ldots, y_n)$ and $y_i$ are i.i.d copies of $\xi_2 - \xi'_2$.

**Proof.** (of Theorem 4.1 assuming Lemma 4.4) See [31, Section 4]. \(\square\)

For the rest of this section we prove Lemma 4.4 using Lemma 4.3. First of all, as $A_U = \bar{A}_U$, it is enough to verify (11) for any index $i$ from $U$. Also, it suffices to assume $\xi$ to have discrete distribution. The continuous case can be recovered by approximating the continuous distribution by a discrete one while holding $n$ fixed.

We begin by applying Corollary 3.4.

**Lemma 4.5.** Assume as in the conclusion of Lemma 4.3. Then, the following holds with probability at least $3\gamma/4$ with respect to $\bar{v}_U$ and $\bar{w}_U$. There exist a proper symmetric GAP $P_{\bar{w}_U} \subset C$ of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$, and an index set $I_{\bar{w}_U} \subset U$ of size $\lvert U \rvert - n^\epsilon$ such that $\langle r_i(A_U), \bar{w}_U \rangle$ is $\beta$-close to $P_{\bar{w}_U}$ for all $i \in I_{\bar{w}_U}$.

**Proof.** (of Lemma 4.5) Write

$$
\sum_{i \in U, j \in \bar{U}} a_{ij}v_i w_j + \sum_{i \in \bar{U}, j \in U} a_{ij}v_i w_j = \sum_{i \in U} \left(\sum_{j \in \bar{U}} a_{ij} w_j\right) v_i + \sum_{j \in U} \left(\sum_{i \in \bar{U}} a_{ji} v_i\right) w_j = \sum_{i \in U} \langle r_i(A_U), \bar{w}_U \rangle v_i + \sum_{i \in \bar{U}} \langle r_i(A_U^T), \bar{v}_U \rangle w_i.
$$

We say that a pair vector $(\bar{v}_U, \bar{w}_U)$ is *good* if

$$
P_{\bar{v}_U,\bar{w}_U}(\lvert \sum_{i \in U} \langle r_i(A_U), \bar{w}_U \rangle v_i + \sum_{i \in \bar{U}} \langle r_i(A_U^T), \bar{v}_U \rangle w_i - a \rvert \leq \beta) \geq \gamma/4.
$$

We call $(\bar{v}_U, \bar{w}_U)$ *bad* otherwise.

Let $G$ denote the collection of good pairs. We are going to estimate the probability $p$ of a randomly chosen pair $(\bar{v}_U, \bar{w}_U)$ being bad by an averaging method.
\[ P_{v_U, w_U} P_{v_U, w_U} \left( |\sum_{i \in U} \langle r_i(A_U), w_U \rangle v_i + \sum_{i \in U} \langle r_i(A_U^T), v_U \rangle w_i - a| \leq \beta \right) = \gamma \]

Thus, the probability of a randomly chosen \((v_U, w_U)\) belonging to \(G\) is at least

\[ 1 - p \geq \frac{3\gamma}{4}/\left(1 - \frac{\gamma}{4}\right) \geq 3\gamma/4. \]

Consider a good vector \((v_U, w_U) \in G\). By definition, we have

\[ P_{v_U, w_U} \left( |\sum_{i \in U} \langle r_i(A_U), w_U \rangle v_i + \sum_{i \in U} \langle r_i(A_U^T), v_U \rangle w_i - a| \leq \beta \right) \geq \frac{\gamma}{4}. \]

Next, if \(\langle r_i(A_U), w_U \rangle = 0\) for all \(i\), then the conclusion of the lemma holds trivially for \(P_{w_U} := 0\). Otherwise, we apply the last conclusion of Corollary \[3,4\] to the sequence \{\(\langle r_i(A_U), w_U \rangle, \langle r_i(A_U^T), v_U \rangle, i \in U\)\} (after a rescaling). As a consequence, we obtain an index set \(I_{w_U} \subset U\) of size \(|U| - n^e\) and a proper symmetric GAP \(P_{w_U} \subset C\) of rank \(O_{B,e}(1)\) and size \(n^{O_{B,e}(1)}\), together with its elements \(q_i(w_U)\), such that \(|\langle r_i(A_U), w_U \rangle - q_i(w_U)| \leq \beta\) for all \(i \in I_{w_U}\).

4.6. Property of the \(q_i(w_U)\)'s. We now work with the GAP elements \(q_i(w_U)\), where \(w_U \in G\). Because these points occupy the most part of an integer box, we can infer a great deal of structural relation among them. To do this, we first pause to introduce a pleasant property of generalized arithmetic progressions.

Assume that \(P = \{k_1g_1 + \cdots + k_rg_r | -K_i \leq k_i \leq K_i\}\) is a proper symmetric GAP, which contains a set \(U = \{u_1, \ldots, u_n\}\). We consider \(P\) together with the map \(\Phi : P \to \mathbb{R}^r\) which maps \(k_1g_1 + \cdots + k_rg_r\) to \((k_1, \ldots, k_r)\). Because \(P\) is proper, this map is bijective. We know that \(P\) contains \(U\), but we do not know yet that \(U\) is non-degenerate in \(P\) in the sense that the set \(\Phi(U)\) has full rank in \(\mathbb{R}^r\). In the later case, we say \(U\) spans \(P\). The following lemma states that we can always assume this without loss of any additive structure.

**Lemma 4.7.** Assume that \(U\) is a subset of a proper symmetric GAP \(P\) of size \(r\), then there exists a proper symmetric GAP \(Q\) that contains \(U\) such that the followings hold.

- \(\text{rank}(Q) \leq r\) and \(|Q| \leq O_r(1)|P|\).
- \(U\) spans \(Q\), that is, \(\phi(U)\) has full rank in \(\mathbb{R}^{\text{rank}(Q)}\).

We refer the reader to \[31\] Theorem 2.1 for a short proof of this lemma.
Common generating indices. By Lemma 4.7, we may assume that the $q_i(w_{\bar{U}})$ span $P_{w_{\bar{U}}}$. We choose $s$ indices $i_{w_1}, \ldots, i_{w_s}$ from $J_{w_{\bar{U}}}$ such that $q_{i_{w_1}}(w_{\bar{U}})$ span $P_{w_{\bar{U}}}$, where $s$ is the rank of $P_{w_{\bar{U}}}$. Note that $s = O_{B, \epsilon}(1)$ for all $w_{\bar{U}} \in G$.

Consider the tuples $(i_{w_1}, \ldots, i_{w_s})$ for all $w_{\bar{U}} \in G$. Because there are $\sum_{\epsilon} O_{B, \epsilon}(n^s) = n^{O_{B, \epsilon}(1)}$ possibilities these tuples can take, there exists a tuple, say $(1, \ldots, r)$ (by rearranging the rows of $A_{\bar{U}}$ if needed), such that $(i_{w_1}, \ldots, i_{w_s}) = (1, \ldots, r)$ for all $w_{\bar{U}} \in G'$, a subset $G'$ of $G$ which satisfies

$$P_{w_{\bar{U}}}(w_{\bar{U}} \in G') \geq P_{w_{\bar{U}}}(w_{\bar{U}} \in G)/n^{O_{B, \epsilon}(1)} = \gamma/n^{O_{B, \epsilon}(1)}.$$  \hspace{1cm} (12)

Common coefficient tuple. For each $1 \leq i \leq r$, we express $q_i(w_{\bar{U}})$ in terms of the generators of $P_{w_{\bar{U}}}$ for each $w_{\bar{U}} \in G'$,

$$q_i(w_{\bar{U}}) = c_{i1}(w_{\bar{U}})g_{i1}(w_{\bar{U}}) + \cdots + c_{ir}(w_{\bar{U}})g_{ir}(w_{\bar{U}}),$$

where $c_{i1}(w_{\bar{U}}), \ldots, c_{ir}(w_{\bar{U}})$ are integers bounded by $n^{O_{B, \epsilon}(1)}$, and $g_{i1}(w_{\bar{U}})$ are the generators of $P_{w_{\bar{U}}}$.

We will show that there are many $w_{\bar{U}}$ that correspond to the same coefficients $c_{ij}$.

Consider the collection of the coefficient-tuples \((c_{i1}(w_{\bar{U}}), \ldots, c_{ir}(w_{\bar{U}})); \ldots; (c_{r1}(w_{\bar{U}}), \ldots, c_{rr}(w_{\bar{U}}))\) for all $w_{\bar{U}} \in G'$. Because the number of possibilities these tuples can take is at most

$$(n^{O_{B, \epsilon}(1)})^r = n^{O_{B, \epsilon}(1)}.$$  \hspace{1cm} (13)

There exists a coefficient-tuple, say \((c_{11}, \ldots, c_{1r}), \ldots, (c_{r1}, \ldots, c_{rr})\), such that

\[
\left( (c_{i1}(w_{\bar{U}}), \ldots, c_{ir}(w_{\bar{U}})); \ldots; (c_{r1}(w_{\bar{U}}), \ldots, c_{rr}(w_{\bar{U}})) \right) = \left( (c_{11}, \ldots, c_{1r}), \ldots, (c_{r1}, \ldots, c_{rr}) \right)
\]

for all $w_{\bar{U}} \in G''$, a subset of $G'$ which satisfies

$$P_{w_{\bar{U}}}(w_{\bar{U}} \in G'') \geq P_{w_{\bar{U}}}(w_{\bar{U}} \in G')/n^{O_{B, \epsilon}(1)} \geq \gamma/n^{O_{B, \epsilon}(1)}.$$  \hspace{1cm} (13)

In summary, there exist $r$ tuples \((c_{11}, \ldots, c_{1r}), \ldots, (c_{r1}, \ldots, c_{rr})\), whose components are integers bounded by $n^{O_{B, \epsilon}(1)}$, such that the followings hold for all $w_{\bar{U}} \in G''$.

- $q_i(w_{\bar{U}}) = c_{i1}g_{i1}(w_{\bar{U}}) + \cdots + c_{ir}g_{ir}(w_{\bar{U}})$, for $i = 1, \ldots, r$.
- The vectors \((c_{11}, \ldots, c_{1r}), \ldots, (c_{r1}, \ldots, c_{rr})\) span $\mathbb{Z}^{\text{rank}(P_{w_{\bar{U}}})}$. 

Next, because $|I_{w^U}| \geq |U| - n'$ for each $w^U \in G''$, by an averaging argument, there exists a set $I \subset U$ of size $|U| - 2n'$ such that for each $i \in I$ we have

$$P_{w^U}(i \in I_{w^U}, w^U \in G'') \geq P_{w^U}(w^U \in G'')/2. \quad (14)$$

From now on we fix an arbitrary row $r$ of index from $I$. We will focus on those $w^U \in G''$ where the index of $r$ belongs to $I_{w^U}$.

**Common coefficient tuple for each individual.** Because $q(w^U) \in P_{w^U}$ ($q(w^U)$ is the element of $P_{w^U}$ that is $\beta$-close to $\langle r, w^U \rangle$), we can write

$$q(w^U) = c_1(w^U)g_1(w^U) + \ldots + c_r(w^U)g_r(w^U)$$

where $c_i(w^U)$ are integers bounded by $n^{O_B(r)}$.

For short, for each $i$ we denote by $v_i$ the vector $(c_1, \ldots, c_i)$, we will also denote by $v_{r,w^U}$ the vector $(c_1(w^U), \ldots, c_r(w^U))$.

Because $P_{w^U}$ is spanned by $q_1(w^U), \ldots, q_r(w^U)$, we have $k = \operatorname{det}(v_1, \ldots, v_r) \neq 0$, and that

$$kq(w^U) + \operatorname{det}(v_{r,w^U}, v_2, \ldots, v_r)q_1(w^U) + \ldots + \operatorname{det}(v_{r,w^U}, v_1, \ldots, v_{r-1})q_r(w^U) = 0. \quad (15)$$

It is crucial to note that $k$ is independent of the choice of $r$ and $w^U$.

Next, because each coefficient of $[15]$ is bounded by $n^{O_B(r)}$, there exists a subset $G''_r$ of $G''$ such that all $w^U \in G''_r$ correspond to the same identity, and

$$P_{w^U}(w^U \in G''_r) \geq (P_{w^U}(w^U \in G'')/2)/(n^{O_B(r)})^r = \gamma/n^{O_B(r)} = n^{-O_B(r)}. \quad (16)$$

In other words, there exist integers $k_1, \ldots, k_r$ depending on $r$, all bounded by $n^{O_B(r)}$, such that

$$kq(w^U) + k_1q_1(w^U) + \ldots + k_rq_r(w^U) = 0 \quad (17)$$

for all $w^U \in G''_r$.

4.8. **Passing back to $A_U$.** Because $q_i(w^U)$ are $\beta$-close to $\langle r, w^U \rangle$, it follows from $[17]$ that

$$\left| \langle kr, w^U \rangle + \langle k_1r_1, w^U \rangle + \ldots + \langle k_r r_r, w^U \rangle \right| = \left| \langle kr + k_1r_1 + \ldots + r_r, w^U \rangle \right| \leq n^{O_B(r)} \beta.$$
Furthermore, as \( P_{w_U}(w_U \in G'_r) = n^{-O_{B,\epsilon}(1)} \), we have
\[
P_{w_U}(\|kr + k_r r_1 + \cdots + k_r r, w_U\| \leq n^{O_{B,\epsilon}(1)} \beta) = n^{-O_{B,\epsilon}(1)}.
\] (18)

As (18) holds for any row \( r \) indexing from \( I \), this completes the proof of Lemma 4.4.

5. Random matrix: the inverse step

We now give a proof of Theorem 2.4. We first apply Theorem 4.1 to \( a_{ij} \) to obtain
\[
P_z(|\langle z, k r_i (A n^{-1}) + \sum_j k_{ij} r_j (A n^{-1}) \rangle| \leq n^{-A + O_{B,\epsilon}(1)} \beta) \geq n^{-O_{B,\epsilon}(1)}.
\]

For short, we denote by \( r'_i \) the vector \( k r_i (A n^{-1}) + \sum_j k_{ij} r_j (A n^{-1}) \). Thus, for any \( i \in I \),
\[
P_z(|\langle z, r'_i \rangle| \leq n^{-A + O_{B,\epsilon}(1)}) \geq n^{-O_{B,\epsilon}(1)}.
\] (19)

Set
\[
K = n^{-A/2}.
\]

We consider two cases.

**Case 1. (non-degenerate case).** There exists \( i_0 \in I \) such that \( \|r'_i\|_2 \geq K \). Because \( r'_i = k r_i (A n^{-1}) + \sum_j k_{ij} r_j (A n^{-1}) \), \( r'_i \) is orthogonal to \( n - |I_0| - 1 = n - O_{B,\epsilon}(1) \) column vectors of \( M_{n-1} \).

Set
\[
v := r'_i / \|r'_i\|_2.
\]

Hence, \( \langle v, c_i (M_{n-1}) \rangle = 0 \) for at least \( n - O_{B,\epsilon}(1) \) column vectors of \( M_{n-1} \).

Also, it follows from (19) that
\[
P_z(|\langle z, v \rangle| \leq n^{-A/2 + O_{B,\epsilon}(1)}) \geq n^{-O_{B,\epsilon}(1)}.
\] (20)

Next, Corollary 3.3 applied to (20) implies that \( v \) can be approximated by a vector \( u \) as follows.

- \( |u_i - v_i| \leq n^{-A/2 + O_{B,\epsilon}(1)} \) for all \( i \).
There exists a GAP of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ that contains at least $n - n'$ components $u_i$.

All the components $u_i$, and all the generators of the GAP are rational complex numbers of the form $\frac{p}{q} + \sqrt{-1} \frac{p'}{q'}$, where $|p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\epsilon}(1)}$.

Note that, by the approximation above, we have $\|u\|_2 \approx 1$ and $|\langle u, c_i(M_{n-1}) \rangle| \leq n^{-A/2+O_{B,\epsilon}(1)}$ for at least $n - O_{B,\epsilon}(1)$ column vectors of $M_{n-1}$.

**Case 2.** (degenerate case) $\|r'_i\|_2 \leq K$ for all $i \in I$. Hence, with $I_0 := \{i_1, \ldots, i_r\}$

$$\|kr_i(A_{n-1}) + \sum_{j \in I_0} k_{ij}r_j(A_{n-1})\|_2 = \|r_i'\|_2 \leq K. \quad (21)$$

Next, because $\sum_j \|c_j(A_{n-1})\|^2 = 1$, there exists an index $j_0$ such that $\|c_{j_0}(A_{n-1})\|_2 \geq n^{-1/2}$. Consider this column vector.

It follows from (21) that for any $i \in I$,

$$|kc_{j_0}(i) + \sum_{j \in I_0} k_{ij}c_{j_0}(j)| \leq K.$$

The above inequality means that the components $c_{j_0}(i)$ of $c_{j_0}(A_{n-1})$ belong to a GAP generated by $c_{j_0}(j)/k, j \in I_0$, up to an error $K$. This suggests us the following approximation.

For each $j \notin I$, we approximate $c_{j_0}(j)$ by a number $v_j$ of the form $(1/\lfloor 2K^{-1} \rfloor) \cdot Z$ such that $|v_j - c_{j_0}(j)| \leq K$. We next set

$$v_i := \sum_{j \in I_0} k_{ij}v_j/k$$

for any $i \in I$.

Thus, $v_i$ belongs to a GAP of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ for all $i \in I$.

With $v = (v_1, \ldots, v_{n-1})$, we have

$$\|v - c_{j_0}(A_{n-1})\|_2 \leq Kn^{O_{B,\epsilon}(1)}.$$

Furthermore, by Condition 1 and because $\langle c_{j_0}(A_{n-1}), r_i(M_{n-1}) \rangle = 0$ for $i \neq j_0$, we infer that

$$|\langle v, r_i(M_{n-1}) \rangle| \leq Kn^{O_{B,\epsilon}(1)}.$$
Note that $\|v\|_2 \gg n^{-1/2}$. Set $u := \lfloor 1/\|v\|_2 \rfloor \cdot v$, we then obtain

- $|\langle u, r_i(M_{n-1}) \rangle| \leq n^{-A/2+O_{B,\epsilon}(1)}$ for $n - 2$ rows of $M_{n-1}$.
- There exists a GAP of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ that contains at least $n - 2n^\epsilon$ components $u_i$.
- All the components $u_i$, and all the generators of the GAP are rational complex numbers of the form $\frac{p}{q} + \sqrt{-1} \frac{p'}{q'}$, where $|p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\epsilon}(1)}$.

6. Random matrix: the counting step

We now give a proof of Theorem 2.5. Our argument, which follows the “divide and conquer” strategy, is simple and purely combinatorial. We note that a similar but simpler treatment for symmetric matrices has appeared in [32, Section 5].

For convenience, let us replace $M_{n-1}$ by $M_n$. We will consider the case $|\langle u, r_i(M_n) \rangle| \leq n^{-A/2+O_{B,\epsilon}(1)}$ for $n - O_{B,\epsilon}(1)$ rows of $M_n$ only, the remaining case $|\langle u, c_i(M_n) \rangle| \leq n^{-A/2+O_{B,\epsilon}(1)}$ can be treated identically.

Let $N$ be the number of such structural vectors $u$. Because each GAP is determined by its generators and dimensions, the number of $Q$’s is bounded by

$$\#\{Q, \text{ there exists } u \text{ such that } u \in Q\} = (n^{2A+O_{B,\epsilon}(1)})^{O_{B,\epsilon}(1)}(n^{O_{B,\epsilon}(1)})^{O_{B,\epsilon}(1)} = n^{O_{A,B,\epsilon}(1)}.$$  

Next, for a given $Q$ of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$, there are at most $n^{n-2n^\epsilon} |Q|^{n-2n^\epsilon} = n^{O_{B,\epsilon}(n)}$ ways to choose the $n - 2n^\epsilon$ components $u_i$ that $Q$ contains. Because the remaining components belong to the set $\left\{ \frac{p}{q} + \sqrt{-1} \frac{p'}{q'}, |p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\epsilon}(1)} \right\}$, so there are at most $(n^{2A+O_{B,\epsilon}(1)})^{2n^\epsilon} = n^{O_{A,B,\epsilon}(n^\epsilon)}$ ways to choose them.

Hence, we obtain the key bound

$$N \leq n^{O_{A,B,\epsilon}(1)} n^{O_{B,\epsilon}(n)} n^{O_{A,B,\epsilon}(n^\epsilon)} = n^{O_{B,\epsilon}(n)}.$$  

Set $\beta_0 := n^{-A/2+O_{B,\epsilon}(1)}$, the bound obtained from the conclusion of Theorem 2.4. For a given vector $u$, we define $P_{\beta_0}(u)$ as follows

$$P_{\beta_0}(u) := P \left( |\langle u, r_i(M_n) \rangle| \leq \beta_0 \text{ for } n - O_{B,\epsilon}(1) \text{ rows of } M_{n-1} \right).$$

For the sake of discussion, let us pretend for now that the rows of $X_n$ are independent. By definition, the vector $u$ is orthogonal to almost every row of $M_n$. Thus, if $u$ is fixed, the probability of this event is bounded by
\[ \mathbf{P}_{\beta_0}(\mathbf{u}) \leq (\mathbf{P}_\times(|u_1 x_1 + \cdots + u_n x_n| \leq \beta_0))^{n-O(1)} = \gamma^{n-O(1)}, \]

where \( x_1, \ldots, x_n \) are i.i.d. copies of \( \xi \).

Now, if \( \gamma \) is small, say \( n^{-O(1)} \), then \( \mathbf{P}_{\beta_0}(\mathbf{u}) \) is \( n^{-\Omega(n)} \). Thus the contribution of these \( \mathbf{P}_{\beta_0}(\mathbf{u}) \) in the total sum \( \sum_u \mathbf{P}_{\beta_0}(\mathbf{u}) \) is negligible, taking into account of the bound \( n^{O(n)} \) of \( \mathcal{N} \).

Next, if \( \gamma \) is comparably large, \( \gamma = n^{-O(1)} \), then by Theorem 3.1 most of the components \( u_i \) are close to a new GAP of rank \( O(1) \) and of size \( O(\gamma^{-1}/\sqrt{n}) \). This would then enable us to approximate \( \mathbf{u} \) by a new vector \( \mathbf{u}' \) in such a way that \( |\langle \mathbf{u}', r_i(M_n) \rangle| \) is still of order \( O(\beta_0) \) and the components of \( \mathbf{u}' \) are now from the new GAPs. The number \( \mathcal{N}' \) of these \( \mathbf{u}' \) can be bounded by \( (\gamma^{-1}/n^\ell)^n \), while we recall that \( \mathbf{P}_{\beta_0}(\mathbf{u}') \) is of order \( \gamma^{-n} \). Thus, summing over \( \mathbf{u}' \) we obtain the desired bound

\[ \sum_{\mathbf{u}'} \mathbf{P}_{\beta_0}(\mathbf{u}') \leq \#\{ \text{ new GAPs } \}(\gamma^{-1}/n^\ell)^n \gamma^{-n} = O(n^{-cn+O(1)}). \]

To our model \( M_n = F_n + X_n \) of independent entries, we will mainly follow the heuristic above. Our strategy is to classify \( \mathbf{u} \) into two classes:

(i) \( \mathcal{B}' \) contains of \( \mathbf{u} \) of very small \( \mathbf{P}_{\beta_0}(\mathbf{u}) \), and thus \( \sum_{\mathbf{u} \in \mathcal{B}} \mathbf{P}_{\beta_0}(\mathbf{u}) \) is negligible;

(ii) the other class \( \mathcal{B} \) contains of \( \mathbf{u} \) of relatively large \( \mathbf{P}_{\beta_0}(\mathbf{u}) \). To deal with those \( \mathbf{u} \) of the second type, we will not control \( \sum_{\mathbf{u} \in \mathcal{B}} \mathbf{P}_{\beta_0}(\mathbf{u}) \) directly but passing to a class of new vectors \( \mathbf{u}' \) that are also almost orthogonal to many rows of \( M_n \), while the probability \( \sum_{\mathbf{u}'} \mathbf{P}_{\beta_0}(\mathbf{u}') \) is of order \( O(n^{-cn}) \).

What makes our analysis harder is that the estimate \( \mathbf{P}_{\beta_0}(\mathbf{u}) \leq (\mathbf{P}_\times(|u_1 x_1 + \cdots + u_n x_n| \leq \beta_0))^{n-O(1)} \) is no-longer valid for our random matrix model.

6.1. Technical reductions and upper bounds for \( \mathbf{P}_{\beta_0}(\mathbf{u}) \). By paying a factor of \( n^{O_B,\ell(1)} \) in probability, we may assume that \( |\langle \mathbf{u}, r_i(M_n) \rangle| \leq \beta_0 \) for the first \( n - O_B,\ell(1) \) rows of \( M_n \). Also, by paying another factor of \( n^{n^\ell} \) in probability, we may assume that the first \( n_0 \) components \( u_i \) of \( \mathbf{u} \) belong to a GAP \( Q_i \), and \( u_{n_0} \geq 1/2 \sqrt{n-1} \) (recall that \( \mathbf{u} \propto 1 \)), where

\[ n_0 := n - 2n^\ell. \]

We refer to the remaining \( u_i \)'s as exceptional components. Note that these extra factors do not affect our final bound \( \exp(-\Omega(n)) \).

For given \( \beta > 0 \) and \( i \leq n_0 \), we define

\[ \gamma_{\beta}^{(i)}(\mathbf{u}) := \sup_a \mathbf{P}_{x_{i}, \ldots, x_{n_0}}(|x_i u_i + \cdots + x_{n_0} u_{n_0} - a| \leq \beta), \]
where \( x_1, \ldots, x_{n_0} \) are i.i.d copies of \( \xi \).

A crucial observation is that, by exposing the rows of \( M_{n-1} \) one by one, and due to symmetry (i.e. \( x_{ij} \) is independent from all other entries except \( x_{ji} \)), the probability \( P_{\beta}(u) \) that \(|\langle u, r_i(M_{n-1}) \rangle| \leq \beta \) for all \( i \leq n - O_{B, \epsilon}(1) \) can be bounded by

\[
P_{\beta}(u) \leq \prod_{1 \leq i \leq n - O_{B, \epsilon}(1)} \sup_a P_{x_i, \ldots, x_{n-1}}(|x_i u_i + \ldots + x_{n-1} u_{n-1} - a| \leq \beta) 
\leq \prod_{1 \leq i \leq n_0} \sup_a P_{x_i, \ldots, x_{n_0}}(|x_i u_i + \ldots + x_{n_0} u_{n_0} - a| \leq \beta) 
= \prod_{1 \leq i \leq n_0} \gamma^{(i)}_{\beta}(u). \tag{23}
\]

Also, because \( u_{n_0} \geq 1/2\sqrt{n - 1} \), there exist positive constants \( c_1, c_2 \) such that \( c_2 < 1 \) and for any \( \beta < c_1/2\sqrt{n - 1} \) we have

\[
\gamma^{(k)}_{\beta}(u) \leq \sup_a P_{x_{n_0}}(|x_{n_0} u_{n_0} - a| \leq \beta) 
\leq 1 - c_2. \tag{24}
\]

Thus,

\[
P_{\beta}(u) \leq (1 - c_2)^{n_0} = (1 - c_2)^{(1-o(1))n}.
\]

6.2. **Classification.** Next, let \( C \) be a sufficiently large constant depending on \( B \) and \( \epsilon \) but not \( A \). We classify \( u \) into two classes \( B \) and \( B' \), depending on whether \( P_{\beta_0}(u) \geq n^{-Cn} \) or not.

Because of [22], and \( C \) is large enough,

\[
\sum_{u \in B'} P_{\beta_0}(u) \leq n^{O_{B, \epsilon}(n)} / n^{Cn} \leq n^{-n/2}. \tag{25}
\]

For the rest of this section, we focus on \( u \in B \).

6.3. **Approximation for “compressible” vectors.** Let \( B_1 \) be the collection of \( u \in B \) satisfying the following property: for any \( n' \) components \( u_1, \ldots, u_{n'} \) among the \( u_1, \ldots, u_{n_0} \), we have

\[
\sup_a P_{x_{i_1}, \ldots, x_{i_{n'}}}(|u_1 x_{i_1} + \ldots + u_{n'} x_{i_{n'}} - a| \leq n^{-B-4}) \geq (n')^{-1/2+o(1)}. \tag{26}
\]
Here we set \( n' := n^{1-\epsilon} \).

For concision we set \( \beta = n^{-B-4} \). It follows from Theorem 3.1 that, among any \( u_{i_1}, \ldots, u_{i_{n'}} \), there are, say, at least \( n'/2 + 1 \) components that belong to a ball of radius \( \beta \) (because our GAP now has only one element). A simple covering argument then implies that there is a ball of radius \( 2\beta \) that contains all but \( n' - 1 \) components \( u_i \).

Thus there exists a vector \( u' \in (2\beta) \cdot (\mathbb{Z} + \sqrt{-1}\mathbb{Z}) \) satisfying the following conditions.

- \( |u_i - u'_i| \leq 4\beta \) for all \( i \).
- \( u'_i \) takes the same value \( u \) for at least \( n_0 - n' \) indices \( i \).

Because of the approximation and of Condition 1 whenever \( |\langle u, r_i(M_{n-1}) \rangle| \leq \beta_0 \), we have

\[
|\langle u', r_i(M_{n-1}) \rangle| \leq n(n^{B+1} + n^n)(4\beta) + \beta_0 := \beta'.
\]

It is clear from the bound on \( \beta \) and \( \beta_0 \) that \( \beta' \leq c_1/2\sqrt{n-1} \), and thus by (24),

\[
P_{\beta'}(u') \leq (1 - c_2)^{(1-o(1))n}.
\]

Now we bound the number of \( u' \) obtained from the approximation. First, there are \( O(n^{n-n_0+n'}) = O(n^{2n_1-\epsilon}) \) ways to choose those \( u'_i \) that take the same value \( u \), and there are just \( O(\beta^{-1}) \) ways to choose \( u \). The remaining components belong to the set \( (2\beta)^{-1} \cdot (\mathbb{Z} + i\mathbb{Z}) \), and thus there are at most \( O((\beta^{-1})^{n-n_0+n'}) = O(n^{O(A,B,\epsilon)(n^{1-\epsilon})}) \) ways to choose them.

Hence we obtain the total bound

\[
\sum_{u \in B_1} P_{\beta_0}(u) \leq \sum_{u'} P_{\beta}(u') \leq O(n^{2n_1-\epsilon})O(n^{O(A,B,\epsilon)(n^{1-\epsilon})})(1 - c_2)^{(1-o(1))n}
\]

\[
\leq (1 - c_2)^{(1-o(1))n}.
\]

6.4. Approximation for ”incompressible” vectors. Assume that \( u \in B_2 := B \setminus B_1 \). By exposing the rows of \( M_{n-1} \) accordingly, and by paying an extra factor \( \binom{n_0}{n} = O(n^{n^{1-\epsilon}}) \) in probability, we may assume that the components \( u_{n_0-n'+1}, \ldots, u_{n_0} \) satisfy the property

\[
\sup_{a} P_{x_{n_0-n'+1}, \ldots, x_{n_0}}(|u_{n_0-n'+1}x_{n_0-n'+1} + \cdots + u_{n_0}x_{n_0} - a| \leq n^{-B-4}) \leq (n')^{-1/2+o(1)}
\]

\[
\leq n^{-1/2+\epsilon/2+o(1)}.
\]
Preparation. Next, define a radius sequence \( \beta_k, k \geq 0 \) where \( \beta_0 = n^{-A/2+O_B(1)} \) is the bound obtained from the conclusion of Theorem 2.4 and

\[
\beta_{k+1} := (n^{B+2} + n^{\alpha+1} + 1)2\beta_k.
\]

Recall from (23) that

\[
P_{\beta_k}(u) \leq \prod_{1 \leq i \leq n_0-n'} \gamma^{(i)}_{\beta_k}(u) =: \pi_{\beta_k}(u).
\]

Roughly speaking, the reason we truncated the product here is that whenever \( i \leq n_0 - n' \) and \( \beta_k \) is small enough, the terms \( \gamma^{(i)}_{\beta_k}(u) \) are smaller than \((n')^{-1/2+o(1)}\), owing to (27). This fact will allow us to gain some significant factors when applying Theorem 3.1.

Observe that if \(|\langle u, r_i(M_n) \rangle| \leq \beta_k\) and if \( u' \) is an approximation of \( u \) such that \(|u_i - u'_i| \leq \beta_k\) for all \( i \), then

\[
\pi_{\beta_k}(u) = \prod_{1 \leq i \leq n_0-n'} \sup_a \mathbf{P}_{x_i,\ldots,x_{n_0}} \left( |u_ix_i + \cdots + u_{n_0}x_{n_0} - a| \leq \beta_k \right)
\]

\[
\leq \prod_{1 \leq i \leq n_0-n'} \sup_a \mathbf{P}_{x_i,\ldots,x_{n_0}} \left( |u'_ix_i + \cdots + u'_{n_0}x_{n_0} - a| \leq (n^{B+1} + n^{\alpha})\beta_k + \beta_k \right)
\]

\[
= \prod_{1 \leq i \leq n_0-n'} \sup_a \mathbf{P}_{x_i,\ldots,x_{n_0}} \left( |u'_ix_i + \cdots + u'_{n_0}x_{n_0} - a| \leq (n^{B+2} + n^{\alpha+1} + 1)\beta_k \right)
\]

\[
\leq \prod_{1 \leq i \leq n_0-n'} \sup_a \mathbf{P}_{x_i,\ldots,x_{n_0}} \left( |u_ix_i + \cdots + u_{n_0}x_{n_0} - a| \leq (n^{B+2} + n^{\alpha+1} + 1)^2\beta_k \right)
\]

\[
= \pi_{\beta_{k+1}}(u).
\]

(28)

Naturally, we hope that after the approximation \( \mathbf{P}_{(n^{B+2} + n^{\alpha+1}+1)\beta_k}(u') \) does not increase much compared to the original \( \mathbf{P}_{\beta_k}(u) \). That motivates us to consider a special radius \( \beta_{k_0} \) with respect to \( u \) defined below.

Note that the bounded sequence \( \pi_{\beta_k}(u) \) increases with \( k \), and recall that \( \pi_{\beta_0}(u) \geq n^{-Cn} \) for \( u \in \mathcal{B} \). Thus, by the pigeonhole principle, there exists \( k_0 := k_0(u) \leq C\epsilon^{-1} \) such that

\[
\pi_{\beta_{k_0+1}}(u) \leq n^{tn}\pi_{\beta_{k_0}}(u).
\]

(29)

It is crucial to note that, since \( A \) was chosen to be sufficiently large compared to \( O_B,\epsilon(1) \) and \( C \), we have

\[
\beta_{k_0+1} \leq n^{-B-4}.
\]
Having mentioned the upper bound of $\gamma^{(i)}_{\beta}(u)$, we now turn to its lower bound. Because of Condition II and $u_i \leq 1$ for all $i$, and by pigeonhole principle, the following trivial bound holds for any $\beta \geq \beta_0$ and $i \leq n_0 - n'$,

$$\gamma^{(i)}_{\beta}(u) \geq \beta n^{n-B-2} \geq \beta_0 n^{n-B-2} = n^{-A/2+O_{B, \epsilon}(1)}.$$

**Subclasses of $u$ in terms of the sequence $(\gamma^{(i)}(u))$.** Set

$$I := [n^{-A/2+O_{B, \epsilon}(1)}, n^{-1/2+\epsilon/2+o(1)}] := [I_1, I_2].$$

We next divide it into $K = (A/2 + O_{B, \epsilon}(1))\epsilon^{-1}$ sub-intervals $I_k = [I_{k_{1-2+\epsilon/2+o(1)}}]$. For short, we denote by $l_k$ the left endpoint of each $I_k$. Thus $l_k = n^{-A/2+O_{B, \epsilon}(1)}k\epsilon$.

With all the necessary settings above, we now classify $u$ basing on the distribution of the $\gamma^{(i)}_{\beta_{k_0}}(u), 1 \leq i \leq n_0 - n^{1-\epsilon}$.

For each $0 \leq k_0 \leq C\epsilon^{-1}$ and each tuple $(m_0, \ldots, m_K)$ satisfying $m_0 + \cdots + m_K = n_0 - n'$, we let $B^{(m_0, \ldots, m_K)}_{k_0}$ denote the collection of those $u$ from $B_2$ that satisfy the following conditions.

- $k_0(u) = k_0$.

- There are exactly $m_k$ terms of the sequence $(\gamma^{(i)}_{\beta_{k_0}}(u))$ that belong to the interval $I_k$.

In other words, if $m_0 + \cdots + m_{k-1} + 1 \leq i \leq m_0 + \cdots + m_k$ then $\gamma^{(i)}_{\beta_{k_0}}(u) \in I_k$.

**The approximation.** Now we will use Theorem 3.1 to approximate $u \in B^{(m_0, \ldots, m_K)}_{k_0}$ as follows.

- **First step.** Consider each index $i$ in the range $1 \leq i \leq m_0$. Because $\gamma^{(1)}_{\beta_{k_0}} \in I_0$, we apply Theorem 3.1 to approximate $u_i$ by $u'_i$ such that $|u_i - u'_i| \leq \beta_{k_0}$ and the $u'_i$ belong to a GAP $Q_0$ of rank $O_{B, \epsilon}(1)$ and size $O(\gamma^{-1}_{\beta_{k_0}}/n^\epsilon) = O(\gamma^{-1}_{\beta_{k_0}}/n^{1/2-\epsilon})$ for all but $n^{1-2\epsilon}$ indices $i$. Furthermore, all $u'_i$ have the form $\beta_{k_0} \cdot (\frac{p}{q} + \sqrt{-1}q')$, where $|p|, |q|, |p'|, |q'| = O(n\beta^{-1}_{k_0}) = O(n^{A/2+O_{B, \epsilon}})$.

- **$k$-th step, $1 \leq k \leq K$.** We focus on $i$ from the range $n_0 + \cdots + n_{k-1} + 1 \leq i \leq n_0 + \cdots + n_k$. Because $\gamma^{(m_0+\cdots+n_{k-1}+1)}_{\beta_{k_0}} \in I_k$, we apply Theorem 3.1 to approximate $u_i$ by $u'_i$ such that $|u_i - u'_i| \leq \beta_{k_0}$ and the $u'_i$ belong to a GAP $Q_k$ of rank $O_{B, \epsilon}(1)$ and size $O(\gamma^{-1}_{\beta_{k_0}}/n^{1/2-\epsilon})$ for all but $n^{1-2\epsilon}$ indices $i$. Furthermore, all $u'_i$ have the form $\beta_{k_0} \cdot (p/q + \sqrt{-1}q')$, where $|p|, |q|, |p'|, |q'| = O(n\beta^{-1}_{k_0}) = O(n^{A/2+O_{B, \epsilon}})$.

- For the remaining components $u_i$, we just simply approximate them by the closest point in $\beta_{k_0} \cdot (\mathbb{Z} + \sqrt{-1}\mathbb{Z})$. 


We have thus provided an approximation of \( u \) by \( u' \) satisfying the following properties.

(i) \( |u_i - u'_i| \leq \beta_k \) for all \( i \).

(ii) \( u'_i \in Q_k \) for all but \( n^{1-2\epsilon} \) indices \( i \) in the range \( m_0 + \cdots + m_{k-1} + 1 \leq i \leq m_0 + \cdots + m_k \).

(iii) All the \( u'_i \), including the generators of \( Q_k \), belong to the set \( \{p/q + \sqrt{-1}p'/q', |p|, |q|, |p'|, |q'| \leq n^{A/2 + O_{B,\epsilon}(1)} \} \).

(iv) \( Q_k \) has rank \( O_{B,\epsilon}(1) \) and size \( |Q_k| = O((\gamma_k - 1)^{k/n_1/2 - \epsilon}) \).

**Property of \( u' \).** Let \( B'(m_1, \ldots, m_K) \) be the collection of all \( u' \) obtained from \( u \in B(m_1, \ldots, m_K) \) as above. Observe that, as \( |\langle u, r_i(M_n) \rangle| \leq \beta_k \) for all \( i \leq n - O_{B,\epsilon}(1) \), we have

\[
|\langle u', r_i(M_n) \rangle| \leq (n^{B+2} + n^{\alpha+1} + 1)\beta_k. \tag{30}
\]

Hence, in order to justify Theorem 2.5 in the case \( u \in B_2 \), it suffices to show that the probability that (30) holds for all \( i \leq n - O_{B,\epsilon}(1) \), for some \( u' \in B'(m_1, \ldots, m_K) \), is small.

Consider a \( u' \in B'(m_1, \ldots, m_K) \) and the probability \( P_{(n^{B+2}+n^{\alpha+1}+1)\beta_k} (u') \) that (30) holds for all \( i \leq n - O_{B,\epsilon}(1) \). By the discussion in 28, we have

\[
P_{(n^{B+2}+n^{\alpha+1}+1)\beta_k} (u') \leq \pi_{\beta_k+1}(u) \leq n^{\epsilon n} \pi_{\beta_k}(u),
\]

where in the second inequality we used (29).

We recall from the definition of \( B'(m_1, \ldots, m_K) \) that

\[
\pi_{\beta_k}(u) \leq \prod_{k=1}^{K} f_{k+1}^{m_k} = n^{\epsilon(m_1 + \cdots + m_k)} \prod_{k=1}^{K} f_{k+1}^{m_k} \leq n^{\epsilon n} \prod_{k=1}^{K} f_{k+1}^{m_k}.
\]

Hence,

\[
P_{(n^{B+2}+n^{\alpha+1}+1)\beta_k} (u') \leq n^{2\epsilon n} \prod_{k=1}^{K} f_{k+1}^{m_k}. \tag{31}
\]
The size of $B_{k_0}^{(m_1,\ldots,m_K)}$. In the next step of the argument, we bound the size of $B_{k_0}^{(m_1,\ldots,m_K)}$. Because each $Q_k$ is determined by its $O_{B,\epsilon}(1)$ generators from the set $\beta_{k_0} \cdot \{ \frac{p}{q} + i \frac{p'}{q'}, |p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\epsilon}(1)} \}$, and its dimensions from the integers bounded by $n^{O_{B,\epsilon}(1)}$, there are $n^{O_{A,B,\epsilon}(1)}$ ways to choose each $Q_k$. So the total number of ways to choose $Q_1, \ldots, Q_K$ is bounded by

$$n^{O_{A,B,\epsilon}(1)} K = n^{O_{A,B,\epsilon}(1)}.$$ 

Next, after locating $Q_k$, the number $N_1$ of ways to choose $u'_i$ from each $Q_k$ is

$$N_1 \leq \prod_{k=1}^{K} \left( \frac{m_k}{n^{1-2\epsilon}} \right) |Q_k|^{m_k-n^{1-2\epsilon}}$$

$$\leq 2^{m_1+\cdots+m_K} \prod_{k=1}^{K} |Q_k|^{m_k}$$

$$\leq (O(1))^n \prod_{k=1}^{K} \gamma_k^{-m_k} / n^{(1/2-\epsilon)(m_1+\cdots+m_K)}$$

$$\leq \prod_{k=1}^{K} \gamma_k^{-m_k} / n^{(1/2-\epsilon-o(1))n},$$

where we used the bound $|Q_k| = O(\gamma_k^{-1}/n^{1/2-\epsilon})$ for each $k$.

The remaining components $u'_i$ can take any value from the set $\beta_{k_0} \cdot \{ \frac{p}{q} + i \frac{p'}{q'}, |p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\epsilon}(1)} \}$, so the number $N_2$ of ways to choose them is bounded by

$$N_2 \leq (n^{A+O_{B,\epsilon}(1)})^{2n^* + Kn^{1-2\epsilon}} = n^{O_{A,B,\epsilon}(n^{1-2\epsilon})}.$$ 

Putting the bound for $N_1$ and $N_2$ together, we obtain a bound $N'$ for $|B_{k_0}^{(m_1,\ldots,m_K)}|$,)

$$N' \leq \prod_{k=1}^{K} l_k^{-m_k} / n^{(1/2-\epsilon-o(1))n}. \quad (32)$$

Closing the argument. It follows from (31) and (32) that

$$\sum_{u' \in B_{k_0}^{(m_1,\ldots,m_K)}} P_{(n^{B+2+n^{n+1}+1})\beta_{k_0}}(u') \leq n^{2en} \prod_{k=1}^{K} l_k^{m_k} \prod_{k=1}^{K} l_k^{-m_k} / n^{(1/2-\epsilon-o(1))n}$$

$$\leq n^{-(1/2-3\epsilon-o(1))n}.$$
Summing over the choices of $k_0$ and $(m_1, \ldots, m_K)$ we obtain the bound
\[
\sum_{k_0, m_1, \ldots, m_K} \mathbb{P}_{(n^{B+2+n^{a+1}+1})^{k_0}}(u') \leq n^{-(1/2-3\epsilon-o(1))n},
\]
completing the treatment for incompressible vectors, and hence the proof of Theorem 2.5.

7. Proof of the elliptic law, Theorems 1.5 and 1.7

This section is devoted to the proof of Theorems 1.5 and 1.7. We introduce the following notation. Given a $n \times n$ matrix $A_n$, we let $\mu_{A_n}$ denote the empirical measure built from the eigenvalues of $A_n$ and $\nu_{A_n}$ denote the empirical measure built from the singular values of $A_n$. That is,
\[
\mu_{A_n} := \frac{1}{n} \sum_{i \leq n} \delta_{\lambda_i(A_n)}
\]
and
\[
\nu_{A_n} := \frac{1}{n} \sum_{i \leq n} \delta_{\sigma_i(A_n)},
\]
where $\lambda_1(A_n), \ldots, \lambda_n(A_n)$ are the eigenvalues of $A_n$ and $\sigma_1(A_n) \geq \cdots \geq \sigma_n(A_n)$ are the singular values of $A_n$.

In order to prove Theorems 1.5 and 1.7, we will show that, with probability one,
\[
\mu_{\frac{1}{\sqrt{n}}(X_n + F_n)} \longrightarrow \mu_\rho
\]
as $n \to \infty$, where $\mu_\rho$ is the uniform probability measure on the ellipsoid $E_\rho$. In particular, (33) implies the almost sure convergence of the ESD of $\frac{1}{\sqrt{n}}(X_n + F_n)$ to the elliptic law with parameter $\rho$.

To this end, let $\mathcal{P}(\mathbb{C})$ be the set of probability measures on $\mathbb{C}$ which integrate $\log |\cdot|$ in a neighborhood of infinity. If $\mu \in \mathcal{P}(\mathbb{C})$, we define the logarithmic potential to be the function
\[
U_\mu(z) := \int \log |z - \lambda| d\mu(\lambda).
\]
We will make use of the following uniqueness property [4, Lemma 4.1]: if $\mu, \nu \in \mathcal{P}(\mathbb{C})$ and $U_\mu(z) = U_\nu(z)$ for a.e. $z \in \mathbb{C}$, then $\mu = \nu$.

We say a Borel function $f$ is uniformly integrable for a sequence of probability measures $\{\mu_n\}_{n \geq 1}$ if
\[
\lim_{t \to \infty} \sup_{n \geq 1} \int_{\{|f| > t\}} |f| d\mu_n = 0.
\]
For a complex $n \times n$ random matrix $A_n$, there is a connection between the measure $\mu_{A_n}$ and the family of measures $\{\nu_{A_n-zI}\}_{z \in \mathbb{C}}$. In particular,
\[
U_{\mu_{A_n}}(z) = -\frac{1}{2n} \log \det(A_n - zI)^*(A_n - zI) = -\int_0^\infty \log(s) d\nu_{A_n-zI}(s).
\]
Remark 7.2. Since the singular values (and eigenvalues) of \((\text{non-random}) \text{ probability measures as } n \to \infty C_0 \text{ sequences of random matrices that satisfy condition given by } \mu \text{ fact below. of measures } \{x \text{ for all } \sigma \text{ } A \text{ matrices where } (\text{Hermitization lemma, } [4]) \text{ Lemma 7.1 } 1.5 \text{ and } 1.7 \text{ is the following result from } [4].

We refer the reader to the survey [4] for more details. A key tool in the proof of Theorems \( \mu \text{ Then there exists a probability measure } \mu \in \mathcal{P}(\mathbb{C}) \text{ such that } U_{\mu}(z) = -\int_{0}^{\infty} \log(s) d\nu_{z}(s).

Remark 7.2. Since the singular values (and eigenvalues) of \((A_n - zI)^*(A_n - zI)\) are just \(\sigma_1^2(A_n - zI), \sigma_2^2(A_n - zI), \ldots, \sigma_n^2(A_n - zI),\) it follows that

\[ \nu(A_n - zI)^*(A_n - zI)(-\infty, x) = \nu_{A_n - zI}(-\infty, \sqrt{x}) \]

for all \(x \geq 0.\) As a consequence, Lemma 7.1 can be equivalently formulated with the family of measures \(\nu(A_n - zI)^*(A_n - zI)_{z \in \mathbb{C}}\) rather than \(\{\nu_{A_n - zI}\}_{z \in \mathbb{C}}.\) We will take advantage of this fact below.

In conjunction with Remark 7.2 we define the matrix

\[ H_n := \left( \frac{1}{\sqrt{n}} X_n - zI \right)^* \left( \frac{1}{\sqrt{n}} X_n - zI \right). \]

For our purposes, we will need to show that the limiting measure \(\mu \in \mathcal{P}(\mathbb{C})\) in Lemma 7.1 is given by \(\mu_\rho.\) Fix \(-1 < \rho < 1.\) We say the family of measures \(\{\nu_z\}_{z \in \mathbb{C}}\) determine the elliptic law with parameter \(\rho\) by Lemma 7.1 if

\[ U_{\mu_\rho}(z) = -\int_{0}^{\infty} \log(s) d\nu_{z}(s) \]

for all \(z \in \mathbb{C}.\) The existence of this family of measures was verified and used in [29].

The key tool we use to prove Theorems 1.5 and 1.7 is the following comparison lemma.

Lemma 7.3. Let \(0 \leq \mu \leq 1 \text{ and } -1 < \rho < 1\) be given. Let \(\{X_n\}_{n \geq 1} \text{ and } \{Y_n\}_{n \geq 1}\) be sequences of random matrices that satisfy condition \(\text{C0} \text{ with atom variables } (\xi_1, \xi_2) \text{ and } (\eta_1, \eta_2),\) respectively. Assume \((\xi_1, \xi_2) \text{ and } (\eta_1, \eta_2)\) are from the \((\mu, \rho)-\text{family.}\) Assume for a.a. \(z \in \mathbb{C}\) that a.s.

\[ \nu_{\frac{1}{\sqrt{n}}Y_n - zI} \longrightarrow \nu_{z} \]

as \(n \to \infty\) for a family of deterministic measures \(\{\nu_z\}_{z \in \mathbb{C}}.\) Assume \(\{F_n\}_{n \geq 1}\) is a sequence of deterministic matrices such that \(\text{rank}(F_n) = o(n)\) and \(\sup_n \frac{1}{n^2} \|F_n\|_2 < \infty.\) Then a.s.

\[ \mu_{\frac{1}{\sqrt{n}}(X_n + F_n)} \longrightarrow \mu_{\frac{1}{\sqrt{n}}Y_n} \longrightarrow 0 \]

as \(n \to \infty.\)
Lemma 7.3 is useful when we know the limit of $\mu_{\sqrt{n}Y_n}$. For our purposes, we will take \( \{Y_n\}_{n \geq 1} \) to be a sequence of matrices that satisfy condition C0 with jointly Gaussian entries. In the real case, the limiting ESD of $\frac{1}{\sqrt{n}}Y_n$ was computed in [29].

We divide the proof of Theorems 1.5 and 1.7 into a number of lemmas organized below by sub-section.

1. In order to apply Lemma 7.1, we need to show that $\log$ is uniformly integrable for \( \{\nu_{\sqrt{n}(X_n + F_n) - zI}\}_{n \geq 1} \). We prove this statement in sub-section 7.4. The arguments in this section are based on [4, 29, 42]. We will also require the use of Theorem 2.1 to control the least singular value.

2. In sub-section 7.12 we prove a replacement lemma using a moment matching argument. The lemma will allow us to compute the limit of $\nu_{\sqrt{n}X_n - zI}$ by comparing the Stieltjes transform of this measure to the corresponding Stieltjes transform in the Gaussian case. In order to prove this lemma, we will first need to bound the variance of the resolvent (sub-section 7.8) and apply a truncation argument (sub-section 7.10).

3. In sub-section 7.15, we prove Lemma 7.3. We then apply the results of [29] and Lemma 7.3 to prove Theorem 1.5.

4. In sub-section 7.16, we prove Theorem 1.7.

7.4. Uniform Integrability. In this sub-section, we prove the following Lemma.

**Lemma 7.5.** Let $0 \leq \mu \leq 1$ and $-1 < \rho < 1$ be given. Let \( \{X_n\}_{n \geq 1} \) be a sequence of random matrices that satisfies condition C0 with atom variables $(\xi_1, \xi_2)$ from the $(\mu, \rho)$-family. Assume \( \{F_n\}_{n \geq 1} \) is a sequence of deterministic matrices such that $\text{rank}(F_n) = o(n)$ and $\sup_n \frac{1}{\sqrt{n}} \|F_n\|^2_2 < \infty$. Then for a.a. $z \in \mathbb{C}$ a.s. $\log$ is uniformly integrable for \( \{\nu_{\sqrt{n}(X_n + F_n) - zI}\}_{n \geq 1} \).

The proof of Lemma 7.5 is based on the arguments of [4, 29, 42]. In order to prove Lemma 7.5, we will need the following bound for small singular values.

**Lemma 7.6.** There exists $c_0 > 0$ and $0 < \gamma < 1$ such that the following holds. Let \( \{X_n\}_{n \geq 1} \) be a sequence of random matrices that satisfies condition C0. Then a.s. for $n \gg 1$ and for all $n^{1-\gamma} \leq i \leq n-1$ and all deterministic $n \times n$ matrices $M$,

\[
\sigma_{n-i}(n^{-1/2}X_n + M) \geq c_0 \frac{i}{n}.
\]

**Proof.** Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ denote the singular values of $A = \frac{1}{\sqrt{n}}X_n + M$. It suffices to prove the lemma for $2n^{1-\gamma} \leq i \leq n-1$ for some $0 < \gamma < 1$ to be chosen later. Let $A'$ be the matrix formed from the first $m = [n - i/2]$ rows of $\sqrt{n}A$. Let $\sigma'_1 \geq \cdots \geq \sigma'_m$ denote the singular values of $A'$. From eigenvalue interlacing it follows that

\[
\frac{1}{\sqrt{n}} \sigma'_{n-i} \leq \sigma_{n-i}.
\]
By [42] Lemma A.4,
\[ \sigma_1^{r-2} + \cdots + \sigma_m^{r-2} = \text{dist}_1^{r-2} + \cdots + \text{dist}_m^{r-2} \]
where \( \text{dist}_i = \text{dist}(r_i, H_i) \), \( r_i \) is the \( i \)-th row of \( A' \), and
\[ H_i = \text{Span}\{r_j : j = 1, \ldots, m; j \neq i\}. \]
Since
\[ \sigma_n^{r-2} \leq n\sigma_n^{r-2} \]
it follows that
\[ \frac{i}{2n}\sigma_n^{r-2} \leq \frac{i}{2}\sigma_n^{r-2} \leq \sum_{j=n-i}^{m} \sigma_{n-j}^{r-2} \leq \sum_{j=1}^{m} \text{dist}_j^{r-2}. \]  \( (34) \)

We now wish to estimate \( \text{dist}(r_j, H_j) \). However, \( r_j \) and \( H_j \) are not independent. To work around this problem, we define the matrix \( A'_j \) to be the matrix \( A' \) with the \( j \)-th column removed. Let \( Y_j \) be the \( j \)-th row of \( A'_j \) and let \( H'_j = \text{Span}\{r_k(A'_j) : k = 1, \ldots, m; k \neq j\} \).

Note that \( Y_j \) and \( H'_j \) are independent for each \( j = 1, \ldots, m \).

We also have
\[ \text{dist}(r_j, H_j) = \inf_{v \in H_j} \| r_j - v \| \geq \inf_{v \in H'_j} \| Y_j - v \| = \text{dist}(Y_j, H'_j) \]
where
\[ \dim(H'_j) \leq \dim(H_j) \leq n - 1 - \frac{i}{2} \leq n - 1 - (n - 1)^{1-\gamma}. \]

By Lemma 7.7 below and the union bound, we obtain
\[ \sum_{n=1}^{\infty} P \left( \bigcup_{i=2n^{1-\gamma}}^{n} \bigcup_{j=1}^{m} \{ \text{dist}_j \leq c_0\sqrt{i} \} \right) < \infty. \]
Thus, by the Borel-Cantelli lemma, for all \( 2n^{1-\gamma} \leq i \leq n - 1 \) and all \( 1 \leq j \leq m \)
\[ \text{dist}_j \geq c_0\sqrt{i} \text{ a.s.} \]
The proof of Lemma 7.6 is then complete by the above estimate and \( (34) \). \( \square \)

**Lemma 7.7** (Distance of a random vector to a subspace). Let \( x \) and \( y \) be complex-valued random variables with unit variance. Then there exists \( \gamma > 0 \) and \( \varepsilon > 0 \) such that the following holds. Let \( (\xi_1, \xi_2, \ldots, \xi_n) \) be a random vector in \( \mathbb{C}^n \) with independent entries. Assume further that for each \( 1 \leq i \leq n \), \( \xi_i \) is equal in distribution to either \( x \) or \( y \). Then for all \( n \gg 1 \), any deterministic vector \( v \in \mathbb{C}^n \) and any subspace \( H \) of \( \mathbb{C}^n \) with \( 1 \leq \dim(H) \leq n - n^{1-\gamma} \), we have
\[ P \left( \text{dist}(R, H) \leq \frac{1}{2}\sqrt{n - \dim(H)} \right) \leq \exp(-n^\varepsilon) \]
where \( R = (\xi_1, \xi_2, \ldots, \xi_n) + v \).
Proof. Let $H'$ be the subspace spanned by $H$, $v$, and $E[R]$. Then $\dim(H') \leq \dim(H) + 2$ and $\text{dist}(R, H') \geq \text{dist}(R, H)$ where $R' = R - E[R]$. Thus it suffices to prove the lemma when $v = 0$ and $E[x] = E[y] = 0$.

We now perform a truncation. By Chebyshev’s inequality,
$$P(\{\vert \xi_i \vert > n^\varepsilon\}) \leq n^{-2\varepsilon}.$$  
Furthermore, by Hoeffding’s inequality
$$P\left(\sum_{i=1}^{n} 1_{\{\vert \xi_i \vert \leq n^\varepsilon\}} < n - n^{1-\varepsilon}\right) \leq \exp(-n^{1-2\varepsilon})$$
where we take $\varepsilon \in (0, 1/3)$. Therefore we will prove the lemma by conditioning on the event 
$$\Omega_m = \{\vert \xi_1 \vert \leq n^\varepsilon, \ldots, \vert \xi_m \vert \leq n^\varepsilon\}$$
with $m = \lceil n - n^{1-\varepsilon}\rceil$.

We now deal with the fact that on the event $\Omega_m$, the random vector $(\xi_1, \ldots, \xi_m)$ may have non-zero mean. Let $E_m$ denote the conditional expectation with respect to the event $\Omega_m$ and the $\sigma$-algebra $F_m = \sigma(\xi_{m+1}, \ldots, \xi_n)$. Let $W$ be the subspace spanned by $H$, $u$, and $w$ where
$$u = (0, \ldots, 0, \xi_{m+1}, \ldots, \xi_n), \quad w = (E_m[\xi_1], \ldots, E_m[\xi_m], 0, \ldots, 0).$$
Clearly $W$ is $F_m$-measurable. Moreover, $\dim(W) \leq \dim(H) + 2$. Define
$$Y = (\xi_1 - E_m[\xi_1], \ldots, \xi_m - E_m[\xi_m], 0, \ldots, 0) = R - u - w.$$  
Then $\text{dist}(R, H) \geq \text{dist}(R, W) = \text{dist}(Y, W)$. By construction each entry of $Y$ has mean zero. Since each entry of the original vector $R$ is equal in distribution to either $x$ or $y$, it follows that
$$\sup_{1 \leq i \leq m} \vert \sigma_i^2 - 1 \vert = 1 - o(1)$$
where $\sigma_i^2 = E_m[\xi_i]^2$.

By Talagrand’s concentration inequality [37],
$$P_m(\vert \text{dist}(Y, W) - M_m \vert \geq t) \leq 4 \exp\left(-\frac{t^2}{16n^{2\varepsilon}}\right)$$
(35)
where $M_m$ is the median of $\text{dist}(Y, W)$ under $\Omega_m$. Using (35) one can verify that
$$M_m \geq \sqrt{E_m}\text{dist}^2(Y, W) - Cn^{4\varepsilon}$$
for some positive constant $C$ (see for instance [45, Lemma E.3]). Let $P$ denote the orthogonal projection onto $W^\perp$. Then
$$E_m\text{dist}^2(Y, W) = \sum_{k=1}^{m} E_m[Y_k^2]P_{kk} \geq c \left(\sum_{k=1}^{n} P_{kk} - \sum_{k=m+1}^{n} P_{kk}\right)$$
$$\geq c(n - \dim(H) - (n - m))$$
for any $1/2 < c < 1$ and $n \gg c$. Thus
$$M_m \geq c\sqrt{n - \dim(H)}$$
for $n$ sufficiently large. Finally, we choose $0 < \gamma < \varepsilon/2$ and the proof of the lemma is complete by taking $t = (c - 1/2)\sqrt{n - \dim(H)}$ in (35).  \[\square\]
We now prove Lemma 7.5.

**Proof of Lemma 7.5.** By Markov’s inequality, it suffices to show that there exists $p > 0$ such that for a.a. $z \in \mathbb{C}$ a.s.

$$\limsup_{n \to \infty} \int s^{-p} d\nu \frac{1}{\sqrt{n}} X_n - z I < \infty \quad \text{and} \quad \limsup_{n \to \infty} \int s^{p} d\nu \frac{1}{\sqrt{n}} X_n - z I < \infty.$$  

Fix $z \in \mathbb{C}$. Then

$$\int s^{p} d\nu \frac{1}{\sqrt{n}} X_n + \frac{1}{\sqrt{n}} F_n - z I \leq 1 + \frac{1}{n} \text{tr} \left( \frac{1}{\sqrt{n}} X_n + \frac{1}{\sqrt{n}} F_n - z I \right)^* \left( \frac{1}{\sqrt{n}} X_n + \frac{1}{\sqrt{n}} F_n - z I \right)$$

for $p \leq 2$. We expand out the right-hand side and consider three separate terms. First, by the law of large numbers,

$$\frac{1}{n} \text{tr} \left( \frac{1}{\sqrt{n}} X_n - z I \right)^* \left( \frac{1}{\sqrt{n}} X_n - z I \right) \leq 1 + \frac{1}{n^2} \sum_{i,j=1}^{n} |x_{ij}|^2 - 2 \text{Re} \left( \frac{z}{n^{3/2}} \sum_{k=1}^{n} x_{kk} \right) + |z|^2 \to 2 + |z|^2$$

a.s. as $n \to \infty$. Here, we first divide the sums into three parts in order to apply the law of large numbers. The first when $i < j$, the second when $i > j$, and the third when $i = j$. In this way the summands in each sum are i.i.d. random variables and the law of large numbers applies.

Second,

$$\left| \frac{1}{n} \text{tr} \left( \frac{1}{\sqrt{n}} F_n^* \right) \left( \frac{1}{\sqrt{n}} X_n + \frac{1}{\sqrt{n}} F_n - z I \right) \right| \leq 1 + \frac{1}{n^2} \|F_n\|_2^2 + \frac{1}{n^2} \text{tr}(F_n^* X_n).$$

Since $\sup_n \frac{1}{n^2} \|F_n\|_2^2 < \infty$ by assumption, it suffices to show that $\limsup_{n \to \infty} \frac{1}{n^2} \text{tr}(F_n^* X_n) < \infty$ a.s. We apply the bounds

$$\left| \frac{1}{n^2} \text{tr}(F_n^* X_n) \right| \leq \frac{1}{n^2} \|F_n\|_2 \|X_n\|_2 \leq \frac{1}{n^2} \|F_n\|_2^2 + \frac{1}{n^2} \|X_n\|_2^2.$$

By the law of large numbers (again considering three separate terms), it follows that a.s.

$$\limsup_{n \to \infty} \frac{1}{n^2} \text{tr}(X_n^* X_n) < \infty.$$

Similarly, for the third term, we have that a.s.

$$\limsup_{n \to \infty} \left| \frac{1}{n} \text{tr} \left( \frac{1}{\sqrt{n}} X_n + \frac{1}{\sqrt{n}} F_n - z I \right)^* \left( \frac{1}{\sqrt{n}} F_n \right) \right| < \infty.$$  

We simplify our notation for the remainder of the proof and write $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ for the singular values of $\frac{1}{\sqrt{n}} (X_n + F_n) - z I$. By Theorem 2.1, we have that for some $A > 0$, $\sigma_n > n^{-A}$ a.s.
Thus,
\[
\frac{1}{n} \sum_{i=1}^{n} \sigma_i^{-p} \leq \frac{1}{n} \sum_{i=1}^{n-n^{-1-\gamma}} \sigma_i^{-p} + \frac{1}{n} \sum_{i=n-n^{-1-\gamma}}^{n} \sigma_i^{-p}
\]
\[
\leq \frac{1}{n} \sum_{i=n^{-1-\gamma}}^{n-1} \sigma_{n-i}^{-p} + \frac{1}{n} n^{1-\gamma} n A p
\]
\[
\leq \frac{1}{c_0 n} \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{n - i}{n} \right)^p \right] + n^{A p - \gamma}
\]
a.s. by Lemma 7.6. The remaining sum is just the Riemann sum of the integral \( \int_0^1 u^{-p} du \).
Therefore, we have that
\[
\frac{1}{n} \sum_{i=1}^{n} \sigma_i^{-p} < \infty \text{ a.s.}
\]
for \( p < \min \{1, \gamma / A \}. \)

7.8. Variance Bound. In this sub-section, we prove the following lemma.

Lemma 7.9. There exists a positive constant \( C \) such that the following holds. Let \( \{X_n\}_{n \geq 1} \) be a sequence of random matrices that satisfies condition \( C_0 \) with atom variables \( (\xi_1, \xi_2) \).
Define
\[
R_n := \left( \frac{1}{\sqrt{n}} X_n - z I \right), \quad H_n(\alpha) := (R_n^* R_n - \alpha I)^{-1},
\]
where \( \alpha \in \mathbb{C} \) with \( \text{Im}(\alpha) \neq 0 \). Then
\[
E \left| \frac{1}{n} \text{tr} H_n(\alpha) - E \left[ \frac{1}{n} \text{tr} H_n(\alpha) \right] \right|^4 \leq C \frac{\mathcal{C}_\alpha}{n^2}
\]
(36)
uniformly for \( z \in \mathbb{C} \) where
\[
\mathcal{C}_\alpha = \frac{1}{|\text{Im}(\alpha)|} + \frac{|\alpha|}{|\text{Im}(\alpha)|^2}.
\]
Moreover, for every fixed \( \alpha \),
\[
\frac{1}{n} \text{tr} H_n(\alpha) = E \left[ \frac{1}{n} \text{tr} H_n(\alpha) \right] + O(n^{-1/8}) \text{ a.s.}
\]
(37)
uniformly for \( z \in \mathbb{C} \).

Proof. Let \( E_{\leq k} \) denote conditional expectation with respect to the \( \sigma \)-algebra generated by \( r_1(X_n), \ldots, r_k(X_n), c_1(X_n), \ldots, c_k(X_n) \). Define
\[
Y_k := E_{\leq k} \frac{1}{n} \text{tr} H_n(\alpha)
\]
for \( k = 0, 1, \ldots, n \). Clearly \( \{Y_k\}_{k=0}^n \) is a martingale. Define the martingale difference sequence
\[
\alpha_k := Y_k - Y_{k-1}
\]
for \( k = 1, 2, \ldots, n \). Then by construction
\[
\sum_{k=1}^{n} \alpha_k = \frac{1}{n} \text{tr} H_n(\alpha) - E \frac{1}{n} \text{tr} H_n(\alpha).
\]
We will bound the fourth moment of the sum, but first we obtain a bound on the individual summands. Let \( X_{n,k} \) denote the matrix \( X_n \) with the \( k \)-th row and \( k \)-th column replaced by zeros. Let

\[
R_{n,k} := \frac{1}{\sqrt{n}} X_{n,k} - zI, \quad H_{n,k}(\alpha) := (R_{n,k}^* R_{n,k} - \alpha I)^{-1}.
\]

It follows that

\[
E \leq k \frac{1}{n} \text{tr} H_{n,k}(\alpha) = E \leq k - 1 \frac{1}{n} \text{tr} H_{n,k}(\alpha)
\]

and hence

\[
\alpha_k = E \leq k \left[ \frac{1}{n} \text{tr} H_n(\alpha) - \frac{1}{n} \text{tr} H_{n,k}(\alpha) \right] - E \leq k - 1 \left[ \frac{1}{n} \text{tr} H_n(\alpha) - \frac{1}{n} \text{tr} H_{n,k}(\alpha) \right].
\]

By the resolvent identity,

\[
|\text{tr} H_n(\alpha) - \text{tr} H_{n,k}(\alpha)| = |\text{tr} [H_n(R_n^* R_n - R_{n,k}^* R_{n,k}) H_{n,k}]|.
\]

Since \( R_n^* R_n - R_{n,k}^* R_{n,k} \) is at most rank 4, it follows that

\[
|\text{tr} H_n(\alpha) - \text{tr} H_{n,k}(\alpha)| \leq 4 \| H_n(R_n^* R_n - R_{n,k}^* R_{n,k}) H_{n,k} \|.
\]

We then note that

\[
\| H_n(\alpha) R_n R_n \| \leq \sup_{t \geq 0} \frac{t}{|t - \alpha|} \leq 1 + \sup_{t \geq 0} \frac{|\alpha|}{|t - \alpha|} \leq 1 + \frac{|\alpha|}{|\text{Im}(\alpha)|}
\]

since the eigenvalues of \( R_n^* R_n \) are non-negative. Similarly,

\[
\| H_{n,k}(\alpha) R_{n,k} R_{n,k} \| \leq 1 + \frac{|\alpha|}{|\text{Im}(\alpha)|}.
\]

Since we always have the bound \( \| H_n(\alpha) \| \leq |\text{Im}(\alpha)|^{-1} \), it follows that

\[
|\text{tr} H_n(\alpha) - \text{tr} H_{n,k}(\alpha)| \leq 8c_\alpha.
\]

Thus we conclude that

\[
|\alpha_k| \leq \frac{16c_\alpha}{n}.
\]

By the Burkholder inequality (see [3] Lemma 2.12) for a complex martingale version of the Burkholder inequality), there exists an absolute constant \( C > 0 \) such that

\[
E \left| \sum_{k=1}^n \alpha_k \right|^4 \leq C E \left( \sum_{k=1}^n |\alpha_k|^2 \right)^2 \leq C \left( n \frac{16^2 c_\alpha^2}{n^2} \right)^2 \leq 16^4 C \frac{c_\alpha^4}{n^2}.
\]

The proof of (36) is complete.

To prove (37), we use Markov’s inequality and (36) to obtain

\[
P \left( \left| \frac{1}{n} \text{tr} H_n(\alpha) - \frac{1}{n} \text{tr} H_n(\alpha) \right| > \varepsilon \right) \leq C \frac{c_\alpha^4}{n^2 \varepsilon^4}.
\]

The result follows by taking \( \varepsilon = n^{-1/8} \) and applying the Borel-Cantelli Lemma. \( \square \)
7.10. **Truncation.** Given a sequence of random matrices \( \{X_n\}_{n \geq 1} \) that satisfies condition C0, we define the sequences \( \{\hat{X}_n\}_{n \geq 1} \) and \( \{\tilde{X}_n\}_{n \geq 1} \) where for each \( n \geq 1 \), \( \hat{X}_n = (\hat{x}_{ij})_{1 \leq i,j \leq n} \) and \( \tilde{X}_n = (\tilde{x}_{ij})_{1 \leq i,j \leq n} \) with
\[
\hat{x}_{ij} = \begin{cases} 
 x_{ij}1_{\{|x_{ij}| \leq n^\delta\}} - E[x_{ij}1_{\{|x_{ij}| \leq n^\delta\}}, & i \neq j \\
 0, & i = j
\end{cases}
\]
and
\[
\tilde{x}_{ij} = \begin{cases} 
 \frac{\hat{x}_{ij}}{\sqrt{E[|x_{ij}|^2]}}, & i \neq j \\
 0, & i = j
\end{cases}
\]
for some \( \delta > 0 \), which we will choose later. For each \( n \geq 1 \), define the matrices
\[
\hat{H}_n = \left( \frac{1}{\sqrt{n}} \hat{X}_n - zI \right)^* \left( \frac{1}{\sqrt{n}} \hat{X}_n - zI \right)
\]
and
\[
\tilde{H}_n = \left( \frac{1}{\sqrt{n}} \tilde{X}_n - zI \right)^* \left( \frac{1}{\sqrt{n}} \tilde{X}_n - zI \right)
\]
We let \( L(\mu, \nu) \) denote the Levy distance between the probability measures \( \mu \) and \( \nu \). We prove the following truncation lemma.

**Lemma 7.11.** Let \( \{X_n\}_{n \geq 1} \) be a sequence of random matrices that satisfies condition C0. Then uniformly for any \( |z| \leq M \), we have that
\[
L(\nu_{H_n}, \nu_{\tilde{H}_n}) = o(1) \text{ a.s.}
\]
Moreover,
\[
E[\text{Re}(\hat{x}_{ij})^k \text{Im}(\hat{x}_{ij})^l \text{Re}(\tilde{x}_{ji})^m \text{Im}(\tilde{x}_{ji})^p] = E[\text{Re}(x_{ij})^k \text{Im}(x_{ij})^l \text{Re}(x_{ji})^m \text{Im}(x_{ji})^p] + o(1) \quad (38)
\]
uniformly for \( i \neq j \) and all non-negative integers \( k, l, m, p \) such that \( k + l + m + p \leq 2 \).

**Proof.** By \cite{3} Corollary A.42,
\[
L^4(\nu_{H_n}, \nu_{\tilde{H}_n}) \leq 2 \left[ \text{tr}(H_n + \hat{H}_n)\text{tr}((X_n - \hat{X}_n)^*(X_n - \hat{X}_n)) \right]. \quad (39)
\]
By the law of large numbers,
\[
\frac{1}{n} \text{tr}H_n = \frac{1}{n^2} \sum_{i,j=1}^n |x_{ij}|^2 - 2\text{Re}\left( \frac{z}{n^{3/2}} \sum_{k=1}^n x_{kk} \right) + |z|^2
\]
\[
\to 1 + |z|^2
\]
a.s. as \( n \to \infty \). Here, we first divide the sums into three parts in order to apply the law of large numbers. The first when \( i < j \), the second when \( i > j \), and the third when \( i = j \). In this way the summands in each sum are i.i.d. and the law of large numbers applies.

Similarly,
\[
\frac{1}{n} \text{tr}\hat{H}_n \to 1 + |z|^2
\]
a.s. as \( n \to \infty \).
For the remaining terms, we note that
\[
\frac{1}{n^2} \text{tr} \left( (X_n - \tilde{X}_n)^*(X_n - \tilde{X}_n) \right) \leq \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \left[ |x_{ij}|^2 1_{\{|x_{ij}| > n^\delta \}} + E|x_{ij}|^2 1_{\{|x_{ij}| > n^\delta \}} \right] \\
+ \frac{2}{n^2} \sum_{1 \leq j < i \leq n} \left[ |x_{ij}|^2 1_{\{|x_{ij}| > n^\delta \}} + E|x_{ij}|^2 1_{\{|x_{ij}| > n^\delta \}} \right] \\
+ \frac{1}{n^2} \sum_{i=1}^{n} |x_{ii}|^2.
\]
By the law of large numbers, each sum on the right-hand side converges to zero a.s. as \( n \to \infty \). Combining these estimates into (39), yields
\[
L(\nu_{H_n}, \nu_{\tilde{H}_n}) = o(1) \quad (40)
\] a.s. as \( n \to \infty \).

Again using [3, Corollary A.42],
\[
L^A(\nu_{H_n}, \nu_{\tilde{H}_n}) \leq \frac{2}{n^3} \text{tr}(\tilde{H}_n + \tilde{H}_n) \text{tr} \left( (\tilde{X}_n - \tilde{X}_n)^*(\tilde{X}_n - \tilde{X}_n) \right).
\]
It then follows that
\[
L(\nu_{\tilde{H}_n}, \nu_{\tilde{H}_n}) = o(1) \quad (41)
\] a.s. since
\[
1 - \sqrt{E|\tilde{x}_{ij}|^2} = o(1)
\]
uniformly for all \( i \neq j \) by the identical distribution assumption of condition C0. The result then follows from estimates (40) and (41).

(38) can be obtained from the dominated convergence theorem; the identical distribution portion of condition C0 gives uniform control for all \( i \neq j \).

7.12. Replacement. In this sub-section, we prove a comparison lemma based on moment matching. We begin with a definition.

**Definition 7.13 (Moment matching).** Let \((\xi_1, \xi_2)\) and \((\eta_1, \eta_2)\) be two random vectors in \( \mathbb{C}^2 \). We say that \((\xi_1, \xi_2)\) and \((\eta_1, \eta_2)\) match to order \( k \) if
\[
E[\text{Re}(\xi_1)^i \text{Im}(\xi_1)^j \text{Re}(\xi_2)^l \text{Im}(\xi_2)^m] = E[\text{Re}(\eta_1)^i \text{Im}(\eta_1)^j \text{Re}(\eta_2)^l \text{Im}(\eta_2)^m]
\]
for all non-negative integers \( i, j, l, m \) with \( i + j + l + m \leq k \).

The goal of this sub-section is to prove the following lemma.

**Lemma 7.14.** Let \( \{X_n\}_{n \geq 1} \) and \( \{Y_n\}_{n \geq 1} \) be sequences of random matrices that satisfy condition C0 with with atom variables \((\xi_1, \xi_2)\) and \((\eta_1, \eta_2)\), respectively. Assume the moments of \((\xi_1, \xi_2)\) and \((\eta_1, \eta_2)\) match to order 2. Then for a.a. \( z \in \mathbb{C} \) a.s.
\[
\nu^{1/n} X_n - zI - \nu^{1/n} Y_n - zI \to 0
\]
as \( n \to \infty \).
We proceed using the Stieltjes transform. For a $n \times n$ matrix $A$, we define the matrices

$$R_n(A) = \frac{1}{\sqrt{n}} A - z I, \quad G_n(A) = (R_n(A)^* R_n(A) - \alpha I)^{-1}$$

where $z, \alpha \in \mathbb{C}$ with $\text{Im}(\alpha) > 0$. Using the resolvent identity, we can compute

$$\frac{\partial (G_n(A))_{ij}}{\partial \text{Re}(A_{st})} = -\frac{1}{\sqrt{n}} [(G_n(A) R_n(A)^*)_{is} (G_n(A))_{tj} + (G_n(A))_{it} (R_n(A) G_n(A))_{sj}]$$

and

$$\frac{\partial (G_n(A))_{ij}}{\partial \text{Im}(A_{st})} = -\frac{\sqrt{-1}}{\sqrt{n}} [(G_n(A) R_n(A)^*)_{is} (G_n(A))_{tj} - (G_n(A))_{it} (R_n(A) G_n(A))_{sj}].$$

Fix the indices $a \neq b$. Let $V_1 = e_a e_b^*$ and $V_2 = e_b e_a^*$ where $e_1, \ldots, e_n$ is the standard basis in $\mathbb{C}^n$. Let $x_1, x_2, x_3, x_4$ be real variables. We define the function

$$f(x_1, x_2, x_3, x_4) = \frac{1}{2} \text{tr} G_n(A + x_1 V_1 + \sqrt{-1} x_2 V_1 + x_3 V_2 + \sqrt{-1} x_4 V_2).$$

Using the derivatives above, we write out the power series

$$f(x_1, x_2, x_3, x_4) = f(0, 0, 0, 0) + \sum_{k=1}^{4} \frac{\partial f}{\partial x_k}(0, 0, 0, 0) x_k + \sum_{i,j=1}^{4} \frac{\partial^2 f}{\partial x_i \partial x_j}(0, 0, 0, 0) x_i x_j + \varepsilon$$

where $|\varepsilon| \leq C M (|x_1|^3 + |x_2|^3 + |x_3|^3 + |x_4|^4)$ with $M$ defined by

$$M = \sup_{1 \leq i,j,k \leq 4} \sup_{x_1, x_2, x_3, x_4} \left| \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} (x_1, x_2, x_3, x_4) \right|.$$

We now obtain a bound for $M$ and the partial derivatives of $f$. Note that the bounds we derive below hold uniformly for any matrix $A$. We can write $R_n(A) = U \sqrt{R_n(A)^* R_n(A)}$ where $U$ is a partial isometry. So

$$\|R_n(A) G_n(A)\| \leq \|U \sqrt{R_n(A)^* R_n(A)} (R_n^* R_n(A) - \alpha I)^{-1}\|
\leq \|\sqrt{R_n(A)^* R_n(A)} (R_n^* R_n(A) - \alpha I)^{-1}\|
\leq \sup_{t \geq 0} \left| \frac{\sqrt{t}}{t - \alpha} \right|
\leq \frac{1}{|\text{Im}(\alpha)|} \sup_{t \geq 0} \left| \frac{t}{t - \alpha} \right|
\leq 1 + \frac{|\alpha| + 1}{|\text{Im}(\alpha)|}$$

and similarly

$$\|G_n(A) R_n(A)^*\| \leq 1 + \frac{|\alpha| + 1}{|\text{Im}(\alpha)|}.$$

Thus, by (42), (43), and the bounds above, it follows that

$$\frac{\partial f}{\partial x_k} = O_{\alpha} \left( \frac{1}{n} \right), \quad \frac{\partial^2 f}{\partial x_k \partial x_i} = O_{\alpha} \left( \frac{1}{n} \right), \quad \frac{\partial^3 f}{\partial x_k \partial x_i \partial x_j} = O_{\alpha} \left( \frac{1}{n} \right)$$

(45)

uniformly for $1 \leq i, j, k \leq 4$, any $x_1, x_2, x_3, x_4 \in \mathbb{R}$, and any $A$. 
We are now ready to prove Lemma 7.14. By Lemma 7.11 and Remark 7.2, it suffices to show that a.s.
\[ \nu_{R_n}(\tilde{X}_n)^* R_n(\tilde{X}_n) \rightarrow 0, \]  
\[ \nu_{R_n}(\tilde{Y}_n)^* R_n(\tilde{Y}_n) \rightarrow 0. \]  
(46)
as \( n \to \infty \). So, without loss of generality, we assume \( \xi_1, \xi_2, \eta_1, \eta_2 \) have mean zero, unit variance, and are bounded almost surely in magnitude by \( n^\delta \) for some \( 0 < \delta < \frac{1}{2} \).

By [3, Theorem B.9], we can equivalently state (46) as
\[ \frac{1}{n} \text{tr} G_n(\tilde{X}_n) = \frac{1}{n} \text{tr} G_n(\tilde{Y}_n) \rightarrow 0 \]a.s. for each fixed \( \alpha \) with \( \text{Im}(\alpha) > 0 \). However by Lemma 7.9, this reduces to showing that
\[ E \frac{1}{n} \text{tr} G_n(\tilde{X}_n) - E \frac{1}{n} \text{tr} G_n(\tilde{Y}_n) \rightarrow 0. \]  
(47)
We will verify (47) by showing that for each fixed \( \alpha \) with \( \text{Im}(\alpha) > 0 \),
\[ E f \left( \frac{\text{Re}(\xi_1)}{\sqrt{n}}, \frac{\text{Im}(\xi_1)}{\sqrt{n}}, \frac{\text{Re}(\xi_2)}{\sqrt{n}}, \frac{\text{Im}(\xi_2)}{\sqrt{n}} \right) - E f \left( \frac{\text{Re}(\eta_1)}{\sqrt{n}}, \frac{\text{Im}(\eta_1)}{\sqrt{n}}, \frac{\text{Re}(\eta_2)}{\sqrt{n}}, \frac{\text{Im}(\eta_2)}{\sqrt{n}} \right) = o_a(n^{-2}) \]where we take \( A \) to be any matrix independent of \( (\xi_1, \xi_2, \eta_1, \eta_2) \). Indeed, by allowing \( a \) and \( b \) to range over all \( O(n^2) \) indices, and by the triangle inequality, we obtain
\[ E \frac{1}{n} \text{tr} G_n(\tilde{X}_n) = E \frac{1}{n} \text{tr} G_n(\tilde{Y}_n) + o_a(1) \]as desired.

It suffices to verify (48) for the off-diagonal entries \( a \neq b \). Indeed, all diagonal entries are assumed to be zero by our previous application of Lemma 7.11.

Using (44), (45), and the independence assumption from condition C0, we obtain
\[ E[f(x_1, x_2, x_3, x_4)] = E[f(0, 0, 0, 0)] + \sum_{i,j=1}^4 E \frac{\partial^2 f}{\partial x_i \partial x_j}(0, 0, 0, 0)E[x_i x_j] + E[\varepsilon] \]
where
\[ x_1 = \frac{\text{Re}(\xi_1)}{\sqrt{n}}, \quad x_2 = \frac{\text{Im}(\xi_1)}{\sqrt{n}}, \quad x_3 = \frac{\text{Re}(\xi_2)}{\sqrt{n}}, \quad x_4 = \frac{\text{Im}(\xi_2)}{\sqrt{n}}, \]
and
\[ E |\varepsilon| = O_a \left( \frac{n^\delta}{n^{2.5}} \right) \]
for some \( 0 < \delta < 1/2 \) from Lemma 7.11.

We repeat the same procedure for \( (\eta_1, \eta_2) \) and obtain
\[ E[f(y_1, y_2, y_3, y_4)] = E[f(0, 0, 0, 0)] + \sum_{i,j=1}^4 E \frac{\partial^2 f}{\partial y_i \partial y_j}(0, 0, 0, 0)E[y_i y_j] + O_a \left( \frac{n^\delta}{n^{2.5}} \right) \]
where
\[ y_1 = \frac{\text{Re}(\eta_1)}{\sqrt{n}}, \quad y_2 = \frac{\text{Im}(\eta_1)}{\sqrt{n}}, \quad y_3 = \frac{\text{Re}(\eta_2)}{\sqrt{n}}, \quad y_4 = \frac{\text{Im}(\eta_2)}{\sqrt{n}}. \]
By (38), we have that 
\[ E[x_i x_j] = E[y_i y_j] + o(n^{-1}) \] uniformly for \(1 \leq i, j \leq 4\). Combining this with (45) yields
\[ E[f(y_1, y_2, y_3, y_4)] = E[f(x_1, x_2, x_3, x_4)] + o(n^{-2}) \] and the proof of Lemma 7.14 is complete.

7.15. **Proof of Lemma 7.3 and Theorem 1.5.** This sub-section is devoted to Lemma 7.3 and Theorem 1.5. The proof of Lemma 7.3 relies on Lemmas 7.14, 7.5, and 7.1.

**Proof of Lemma 7.3.** Assume \( \mu, \rho, \{X_n\}_{n \geq 1}, \{Y_n\}_{n \geq 1}, \{F_n\}_{n \geq 1} \) satisfy the assumptions in the statement of Lemma 7.3. By \([3, \text{Theorem A.44}]\), it follows that for a.a. \( z \in \mathbb{C} \) a.s.
\[ \nu_1^{1/2n}(X_n + F_n) - zI \longrightarrow 0 \] as \( n \to \infty \). Since both \((\xi_1, \xi_2)\) and \((\eta_1, \eta_2)\) are from the \((\mu, \rho)\)-family, then \((\xi_1, \xi_2)\) and \((\eta_1, \eta_2)\) match to order 2. Thus by Lemma 7.14 for a.a. \( z \in \mathbb{C} \) a.s.
\[ \nu_1^{1/2n}X_n - zI \longrightarrow \nu_z \] as \( n \to \infty \). Therefore, for a.a. \( z \in \mathbb{C} \) a.s.
\[ \nu_1^{1/2n}(X_n + F_n) - zI \longrightarrow \nu_z \] as \( n \to \infty \).

Furthermore, by Lemma 7.5 for a.a. \( z \in \mathbb{C} \) a.s. log is uniformly integrable for \( \{\nu_1^{1/2n}Y_n - zI\}_{n \geq 1} \) and hence by Lemma 7.1, we conclude that \( \mu_1^{1/2n}Y_n \longrightarrow \mu_\rho \) a.s. as \( n \to \infty \).

We can now prove Theorem 1.5.

**Proof of Theorem 1.5.** Let \( \{X_n\}_{n \geq 1} \) be a sequence of real random matrices that satisfies condition C0 with atom variables \((\xi_1, \xi_2)\) and \( \rho = E[\xi_1 \xi_2] \) for some \(-1 < \rho < 1\). Let \( \{Y_n\}_{n \geq 1} \) be the sequence of random matrices that satisfies condition C0 with atom variables \((\eta_1, \eta_2)\) where \( \eta_1 \) and \( \eta_2 \) are jointly Gaussian and \( E[\eta_1 \eta_2] = \rho \). In \([29, \text{Theorem 5.2}]\), it is shown that
\[ E\nu_1^{1/2n}Y_n - zI \longrightarrow \nu_z \] as \( n \to \infty \) where the family \( \{\nu_z\}_{z \in \mathbb{C}} \) determines the elliptic law with parameter \( \rho \) by Lemma 7.1. In fact, using the variance bound in Lemma 7.9 and \([3, \text{Theorem B.9}]\) it can be shown that a.s.
\[ \nu_1^{1/2n}Y_n - zI \longrightarrow \nu_z \] as \( n \to \infty \). By Lemma 7.5 for a.a. \( z \in \mathbb{C} \) a.s. log is uniformly integrable for \( \{\nu_1^{1/2n}Y_n - zI\}_{n \geq 1} \) and hence by Lemma 7.1 we conclude that \( \mu_1^{1/2n}Y_n \longrightarrow \mu_\rho \) a.s. as \( n \to \infty \).

Since both \((\xi_1, \xi_2)\) and \((\eta_1, \eta_2)\) are from the \((1, \rho)\)-family, the proof of the theorem is complete by an application of Lemma 7.3. \qed
7.16. **Proof of Theorem 1.7.** In the proof of Theorem 1.5 above, we relied on the previous results in [29], where the entries are assumed to be real. In order to prove Theorem 1.7, we first need to study the complex Gaussian case.

**Lemma 7.17.** Let \( 0 \leq \mu < 1 \) and \(-1 < \rho < 1\) be given. Assume \( \{X_n\}_{n \geq 1} \) is a sequence of complex matrices that satisfy condition C0 with atom variables \((\xi_1, \xi_2)\) from the \((\mu, \rho)\)-family, where \(\text{Re}(\xi_1), \text{Im}(\xi_1), \text{Re}(\xi_2), \text{Im}(\xi_2)\) are jointly Gaussian. Then for a.a. \( z \in \mathbb{C} \) a.s.

\[
\sqrt{n} X_n - zI \overset{\nu}{\longrightarrow} \nu z
\]

as \( n \to \infty \) where \( \{\nu_z\}_{z \in \mathbb{C}} \) determines the elliptic law with parameter \( \rho \) by Lemma 7.1.

Let us assume Lemma 7.17 for now and complete the proof of Theorem 1.7.

**Proof of Theorem 1.7.** Let \( \{X_n\}_{n \geq 1} \) be a sequence of complex random matrices that satisfy condition C0 with atom variables \((\xi_1, \xi_2)\) from the \((\mu, \rho)\)-family. Let \( \{Y_n\}_{n \geq 1} \) be the sequence of complex random matrices that satisfy condition C0 with atom variables \((\eta_1, \eta_2)\) from the \((\mu, \rho)\)-family, where \(\text{Re}(\eta_1), \text{Im}(\eta_1), \text{Re}(\eta_2), \text{Im}(\eta_2)\) are jointly Gaussian. By Lemma 7.17 for a.a. \( z \in \mathbb{C} \) a.s.

\[
\sqrt{n} Y_n - zI \overset{\nu}{\longrightarrow} \nu z
\]

as \( n \to \infty \) where the family \( \{\nu_z\}_{z \in \mathbb{C}} \) determines the elliptic law with parameter \( \rho \) by Lemma 7.1. Moreover, log is uniformly integrable for \( \{\sqrt{n} Y_n - zI\}_{n \geq 1} \) by Lemma 7.5. Therefore, by Lemma 7.1 it follows that a.s.

\[
\sqrt{n} Y_n \overset{\mu}{\longrightarrow} \mu \rho
\]

as \( n \to \infty \). The proof of Theorem 1.7 is now complete by Lemma 7.3. \( \square \)

All that remains is to prove Lemma 7.17. Let \( \{X_n\}_{n \geq 1} \) be the sequence of random matrices defined in Lemma 7.17 with jointly Gaussian off-diagonal entries. We follow [29] and introduce the following notation.

For \( n \times n \) matrices \( A \) and \( B \), we define the \( 2n \times 2n \) block matrices

\[
V := \begin{bmatrix} \frac{1}{\sqrt{n}} A & 0 \\ 0 & \frac{1}{\sqrt{n}} B^* \end{bmatrix}, \quad J(z) := \begin{bmatrix} 0 & zI \\ zI & 0 \end{bmatrix}
\]

and set

\[
V(z) := VJ - J(z)
\]

where \( J := J(1) \). We let \( R \) denote the resolvent of \( V(z) \). That is,

\[
R := [V(z) - \alpha I]^{-1}
\]

for \( \alpha \in \mathbb{C} \).
Using the resolvent identity, we can compute
\[
\frac{\partial R_{ab}}{\partial \text{Re}(A_{cd})} = -\frac{1}{\sqrt{n}}R_{ac}R_{d+n,b},
\]
\[
\frac{\partial R_{ab}}{\partial \text{Im}(A_{cd})} = -\sqrt{-1}\frac{1}{\sqrt{n}}R_{ac}R_{d+n,b},
\]
\[
\frac{\partial R_{ab}}{\partial \text{Re}(B_{cd})} = -\frac{1}{\sqrt{n}}R_{a,d+n,R_{c+b}},
\]
\[
\frac{\partial R_{ab}}{\partial \text{Im}(B_{cd})} = \sqrt{-1}\frac{1}{\sqrt{n}}R_{a,d+n,R_{c+b}},
\]
for \(1 \leq c,d \leq n\) and \(1 \leq a,b \leq 2n\). For the remainder of the paper, we will take \(A = B = X_n\).

We will make use of the multivariate Gaussian decoupling formula \[35\]. That is, if \(Y = \{\xi_i\}_{i=1}^p\) is a real random Gaussian vector such that
\[
E[\xi_j] = 0, \quad E[\xi_j \xi_k] = C_{jk}
\]
for \(j,k = 1,2,\ldots,p\) and if \(\Phi : \mathbb{R}^p \to \mathbb{C}\) has bounded partial derivatives, then
\[
E[\xi_j \Phi] = \sum_{k=1}^p C_{jk} E[(\nabla \Phi)_k].
\]

Using the partial derivatives above and the Gaussian decoupling formula, we obtain
\[
E[R_{ab} x_{cd}] = -\frac{1}{\sqrt{n}} E[R_{a,d+n,R_{c+b}}] - \frac{\rho}{\sqrt{n}} E[R_{a+d,R_{c+n,b}}] + \frac{1-2\mu}{\sqrt{n}} E[R_{a,c}R_{d+n,b}] + \frac{1-2\mu}{\sqrt{n}} E[R_{a,c+n,R_{d+b}}]
\]
and
\[
E[R_{ab} \bar{x}_{cd}] = -\frac{1}{\sqrt{n}} E[R_{ac,R_{d+n,b}}] - \frac{\rho}{\sqrt{n}} E[R_{a,c+n,R_{d+b}}] + \frac{1-2\mu}{\sqrt{n}} E[R_{a,d}R_{c+n,b}] + \frac{1-2\mu}{\sqrt{n}} E[R_{a,d+n,R_{c+b}}]
\]
for \(1 \leq c,d \leq n\), \(c \neq d\), and \(1 \leq a,b \leq 2n\).

Following \[29\], we define the functions
\[
s_n := s_n(\alpha, z) = \frac{1}{2n} E[\text{tr} R] = \frac{1}{n} \sum_{i=1}^n E[R_{ii}] = \frac{1}{n} \sum_{i=1}^n E[R_{i+n,i+n}]
\]
and
\[
t_n := t_n(\alpha, z) = \frac{1}{n} \sum_{i=1}^n E[R_{i+n,i}], \quad u_n := u_n(\alpha, z) = \frac{1}{n} \sum_{i=1}^n E[R_{i,i+n}].
\]

We now fix \(z, \alpha \in \mathbb{C}\) with \(\text{Im}(\alpha) > 0\). In the definitions above, we deal with the expectation of the summands instead of the random elements. In order to justify this, we need control of the variance, which we obtain in the following lemma.
Lemma 7.18.

\[
\text{Var}\left(\frac{1}{2n} \text{tr} R\right) = O_{\alpha,z}\left(\frac{1}{n}\right), \quad (51)
\]

\[
\text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} R_{i+n,i}\right) = O_{\alpha,z}\left(\frac{1}{n}\right), \quad (52)
\]

\[
\text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} R_{i,i+n}\right) = O_{\alpha,z}\left(\frac{1}{n}\right). \quad (53)
\]

Proof. We begin by noting that

\[
\frac{1}{n} \sum_{i=1}^{n} R_{i+n,i} = \frac{1}{n} \text{tr}(P_2 R P_1)
\]

and

\[
\frac{1}{n} \sum_{i=1}^{n} R_{i,i+n} = \frac{1}{n} \text{tr}(P_1 R P_1^*)
\]

where \(P_1\) and \(P_2\) are partial isometries. Thus, it suffices to prove

\[
\text{Var}\left(\frac{1}{n} \text{tr}(PRQ)\right) = O_{\alpha,z}\left(\frac{1}{n}\right)
\]

for arbitrary partial isometries \(P\) and \(Q\).

Let \(E_{\leq k}\) denote conditional expectation with respect to the \(\sigma\)-algebra generated by the random vectors

\[
r_1(X_n), \ldots, r_k(X_n), c_1(X_n), \ldots, c_k(X_n).
\]

Define

\[
Y_k := E_{\leq k} \frac{1}{n} \text{tr}(PRQ)
\]

for \(k = 0, 1, \ldots, 2n\). Clearly \(\{Y_k\}_{k=0}^{2n}\) is a martingale. Define the martingale difference sequence

\[
\alpha_k := Y_k - Y_{k-1}
\]

for \(k = 1, 2, \ldots, 2n\). Then by construction

\[
\sum_{k=1}^{2n} \alpha_k = \frac{1}{n} \text{tr}(PRQ) - E\frac{1}{n} \text{tr}(PRQ).
\]

Thus we need to show that

\[
E\left|\sum_{k=1}^{2n} \alpha_k\right|^2 = O_{\alpha,z}\left(\frac{1}{n}\right).
\]

Again we introduce the notation \(X_{n,k}\) to denote the matrix \(X_n\) with the \(k\)-th row and \(k\)-th column replaced by zeros. Let

\[
V_k := \begin{bmatrix}
\frac{1}{\sqrt{n}} X_{n,k} & 0 \\
0 & \frac{1}{\sqrt{n}} X_{n,k}^*
\end{bmatrix}
\]

and define

\[
R_k := [V_k J - J(z) - \alpha I]^{-1}.
\]
Since
\[ E_{\leq k} \frac{1}{n} \text{tr}(PR_k Q) = E_{\leq k-1} \frac{1}{n} \text{tr}(PR_k Q) \]
it follows that
\[ \alpha_k = E_{\leq k} \left[ \frac{1}{n} \text{tr}(PR Q) - \frac{1}{n} \text{tr}(PR_k Q) \right] - E_{\leq k-1} \left[ \frac{1}{n} (PR Q) - \frac{1}{n} \text{tr}PR_k Q \right]. \]

Because \( V_k - V \) is at most rank 4, we have that
\[ |\text{tr}(PR Q) - \text{tr}(PR_k Q)| \leq 4 \| R((V_k J - J(z)) - (V J - J(z))) J R_k \|. \]
We now claim that
\[ \| R((V_k J - J(z)) - (V J - J(z))) J R_k \| = O_{\alpha,z}(1) \] (54)
uniformly in \( k \). Indeed,
\[ \| R(V J - J(z)) R_k \| \leq \frac{1}{|\text{Im}(\alpha)|} \| R(V J - J(z)) \| \leq \frac{1}{|\text{Im}(\alpha)|} \sup_{t \in \mathbb{R}} \frac{|t|}{|t - \alpha|} = O_{\alpha,z}(1) \]
since \( V J - J(z) \) is Hermitian. Similarly,
\[ \| R(V_k J - J(z)) R_k \| = O_{\alpha,z}(1) \]
and (54) follows. Thus
\[ \alpha_k = O_{\alpha,z} \left( \frac{1}{n} \right) \]
uniformly in \( k \).

By the Burkholder inequality (see [3, Lemma 2.12] for a complex martingale version of the Burkholder inequality), there exists an absolute constant \( C > 0 \) such that
\[ E \left| \sum_{k=1}^{2n} \alpha_k \right|^2 \leq C E \sum_{k=1}^{2n} |\alpha_k|^2 = O_{\alpha,z} \left( \frac{1}{n} \right). \]

\[ \square \]

We are now ready to prove Lemma 7.17.

\textbf{Proof of Lemma 7.17} Fix \( z, \alpha \in \mathbb{C} \) with \( \text{Im}(\alpha) > 0 \). By the resolvent identity,
\[ 1 + \alpha s_n = \frac{1}{2n} E \text{tr}(RV J) - \frac{z}{2} u_n - \frac{z}{2} t_n. \]
We decompose,
\[ \frac{1}{2n} E \text{tr}(RV J) = \frac{1}{2} A_1 + \frac{1}{2} A_2 \]
where
\[ A_1 = \frac{1}{n} E \sum_{i=1}^{n} (RV J)_{ii} \quad \text{and} \quad A_2 = \frac{1}{n} E \sum_{i=1}^{n} (RV J)_{i+n,i+n}. \]
By (50) and Lemma 7.18, we have

\[ A_1 = \frac{1}{n^{3/2}} \mathbb{E} \sum_{i,j=1}^{n} R_{i,j+n} \bar{x}_{ij} \]

\[ = -\frac{1}{n^2} \mathbb{E} \sum_{i,j=1}^{n} [R_{ii}R_{j+n,j+n} + \rho R_{ii} R_{j,j+n} - (1 - 2\mu)\rho R_{i,j} R_{i+n,j+n} \]

\[ - (1 - 2\mu) R_{i,j+n} R_{i,j+n}] + o_{\alpha,z}(1) \]

\[ = -s_n^2 - \rho \bar{u}_n^2 + o_{\alpha,z}(1). \]

For the second line, we used that the diagonal entries \(i = j\) give total contribution \(O_{\alpha,z}(n^{-1/2})\) which we write as the \(o_{\alpha,z}(1)\) term. For the third line, we used that if

\[ R = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} \]

where \(R_1, R_2, R_3, R_4\) are \(n \times n\) matrices, then

\[ \left| \frac{1}{n^2} \sum_{i,j=1}^{n} R_{ii} R_{j+n,j+n} \right| = \left| \frac{1}{n^2} \text{tr}(R_1^T R_4) \right| \leq \frac{1}{n} \|R\|^2 \leq \frac{1}{n|\text{Im}(\alpha)|^2} \]

and

\[ \left| \frac{1}{n^2} \sum_{i,j=1}^{n} R_{ij+n} R_{i,j+n} \right| = \left| \frac{1}{n^2} \text{tr}(R_2^T R_2) \right| \leq \frac{1}{n|\text{Im}(\alpha)|^2}. \]

Similarly (using (49) and Lemma 7.18),

\[ A_2 = -s_n^2 - \rho t_n^2 + o_{\alpha,z}(1). \]

Thus

\[ 1 + \alpha s_n = -s_n^2 - \rho \frac{u_n^2}{2} - \rho \frac{\bar{u}_n^2}{2} - \bar{z} u_n - \frac{z}{2} t_n + o_{\alpha,z}(1). \] \( (55) \)

We now obtain an equation for \(t_n\). Again by the resolvent identity

\[ \alpha t_n = \frac{1}{n^{3/2}} \mathbb{E} \sum_{i,j=1}^{n} R_{i+n,j+n} \bar{x}_{ij} - \bar{z} \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} R_{i+n,i+n} = A_3 - \bar{z} s_n, \]

where

\[ A_3 = \frac{1}{n^{3/2}} \mathbb{E} \sum_{i,j=1}^{n} R_{i+n,j+n} \bar{x}_{ij}. \]

We repeat almost exactly the same procedure as above (using (50) and Lemma 7.18) and obtain

\[ A_3 = -\frac{1}{n^2} \mathbb{E} \sum_{i,j=1}^{n} [R_{i+n,i} R_{j+n,j+n} + \rho R_{i+n,i+n} R_{j,j+n} \]

\[ - (1 - 2\mu)\rho R_{i+n,j} R_{i+n,j+n} - (1 - 2\mu) R_{i+n,j+n} R_{i,j+n}] + o_{\alpha,z}(1) \]

\[ = -t_n s_n - \rho s_n u_n + o_{\alpha,z}(1). \]
Again we used the fact the the diagonal entries give contribution $O_{\alpha,z}(n^{-1/2})$ and control the remaining error terms by writing each as a trace of products of $R_1, R_2, R_3, R_4$. Thus we conclude

$$\alpha t_n = -t_n s_n - \rho s_n u_n - \bar{z}s_n + o_{\alpha,z}(1).$$

Similarly, we obtain an equation for $u_n$:

$$\alpha u_n = -u_n s_n - \rho s_n t_n - z s_n + o_{\alpha,z}(1).$$

Combining the above equations for $t_n$ and $u_n$ with (55), we arrive at the following system of three equations:

$$1 + \alpha s_n = -s_n^2 - \frac{\rho}{2} u_n^2 - \frac{\rho}{2} t_n^2 - \frac{\bar{z}}{2} u_n - \frac{z}{2} t_n + o_{\alpha,z}(1)$$

$$\alpha t_n = -t_n s_n - \rho s_n u_n - \bar{z}s_n + o_{\alpha,z}(1)$$

$$\alpha u_n = -u_n s_n - \rho s_n t_n - z s_n + o_{\alpha,z}(1).$$

We note that the system of equations above does not depend on $\mu$. In addition to the case above, we also consider the case when $\mu = 1$. This corresponds to the real Gaussian case studied in [29]. Repeating the same calculations as above, we obtain the following system of equations in the real Gaussian case:

$$1 + \alpha \hat{s}_n = -\hat{s}_n^2 - \frac{\rho}{2} \hat{u}_n^2 - \frac{\rho}{2} \hat{t}_n^2 - \frac{\bar{z}}{2} \hat{u}_n - \frac{z}{2} \hat{t}_n + o_{\alpha,z}(1)$$

$$\alpha \hat{t}_n = -\hat{t}_n \hat{s}_n - \rho \hat{s}_n \hat{u}_n - \bar{z}\hat{s}_n + o_{\alpha,z}(1)$$

$$\alpha \hat{u}_n = -\hat{u}_n \hat{s}_n - \rho \hat{s}_n \hat{t}_n - z \hat{s}_n + o_{\alpha,z}(1).$$

One can also check that this system matches (5.4), (5.5), and (5.6) from [29]. In [29], it is shown that for every $\alpha, z \in \mathbb{C}$ with $\text{Im}(\alpha) > 0$,

$$\lim_{n \to \infty} \hat{s}_n = s_0$$

where $s_0 = s_0(\alpha, z)$ is given by

$$s_0 = \frac{1}{2} \int_R \frac{d\nu_z(x)}{x-\alpha} - \frac{1}{2} \int_R \frac{d\nu_z(x)}{x + \alpha}$$

and the family $\{\nu_z\}_{z \in \mathbb{C}}$ determines the elliptic law with parameter $\rho$ by Lemma 7.1.

By Lemma 7.18 and [3, Lemma B.9], it suffices to show that for every $\alpha, z \in \mathbb{C}$ with $\text{Im}(\alpha) > 0$,

$$\lim_{n \to \infty} s_n = s_0$$

in order to complete the proof of Lemma 7.17.

Since $\|R\| \leq \frac{1}{\text{Im}(\alpha)}$, it follows that $|s_n|, |t_n|, |u_n| \leq \frac{1}{\text{Im}(\alpha)}$. Similarly, $|\hat{s}_n|, |\hat{t}_n|, |\hat{u}_n| \leq \frac{1}{\text{Im}(\alpha)}$. So by Vitali’s convergence theorem, it suffices to show that for any fixed $z \in \mathbb{C}$, (58) holds for any $\alpha \in \mathbb{C}$ with $\text{Im}(\alpha)$ sufficiently large.
Taking $\text{Im} (\alpha) > 1$, we subtract the last two equations of (56) and (57) to obtain

$$|t_n - \hat{t}_n| \leq \frac{1}{\text{Im}(\alpha)^2 - 1} |u_n - \hat{u}_n| + \frac{1 + \rho + \text{Im}(\alpha)|z|}{\text{Im}(\alpha)^2 - 1} |s_n - \hat{s}_n| + o_{\alpha,z}(1)$$

and

$$|u_n - \hat{u}_n| \leq \frac{1}{\text{Im}(\alpha)^2 - 1} |t_n - \hat{t}_n| + \frac{1 + \rho + \text{Im}(\alpha)|z|}{\text{Im}(\alpha)^2 - 1} |s_n - \hat{s}_n| + o_{\alpha,z}(1).$$

Taking $\text{Im}(\alpha)$ sufficiently large (in terms of $\rho$ and $|z|$) we can write (say)

$$|t_n - \hat{t}_n| \leq \frac{2}{99} |s_n - \hat{s}_n| + o_{\alpha,z}(1)$$

and

$$|u_n - \hat{u}_n| \leq \frac{2}{99} |s_n - \hat{s}_n| + o_{\alpha,z}(1).$$

We now subtract the first equations of (56) and (57) and apply the bounds above to obtain

$$|s_n - \hat{s}_n| \leq \frac{8}{99} |s_n - \hat{s}_n| + o_{\alpha,z}(1)$$

for $\text{Im}(\alpha) > \max\{99, |z|\}$.

This implies that for any $\alpha, z \in \mathbb{C}$ fixed with $\text{Im}(\alpha)$ sufficiently large,

$$s_n = \hat{s}_n + o(1)$$

and the proof of Lemma 7.17 is complete. \qed

APPENDIX A. PROOF OF THEOREM 3.2

Our first step is to obtain the following.

**Claim A.1** (upper bound for small ball probability). We have

$$\sup_a \mathbb{P}(\left| \sum_i (a_i x_i + b_i x'_i) - a \right| \leq r) \leq \exp(\pi r^2) \int_{\mathbb{C}} \exp\left(-\sum_{i=1}^n E_{\xi_i,\xi'_i,\xi_2} \|\text{Re}(2(\xi_1 - \xi'_1)a_i t + 2(\xi_2 - \xi'_2)b_i t)\|_{\mathbb{R}/\mathbb{Z}}^2 - \pi |t|^2\right) dt,$$

where $\|z\|_{\mathbb{R}/\mathbb{Z}}$ is the distance from a real number $z$ to its nearest integer.

**Proof.** (of Claim A.1) First of all, we have
\[
\Pr(\sum_{i=1}^{n} (a_i x_i + b_i x'_i) \in B(a, r)) = \Pr(\sum_{i=1}^{n} a_i x_i + b_i x'_i - a^2 \leq r^2)
\]
\[
= \Pr(\exp(-\pi)\sum_{i=1}^{n} a_i x_i + b_i x'_i - a^2 \geq \exp(-\pi r^2))
\]
\[
\leq \exp(\pi r^2) \Pr(\sum_{i=1}^{n} a_i x_i + b_i x'_i - a^2)\].
\]

Note that for any \(z \in \mathbb{R}^2\), \(\exp(-\pi |z|^2) = \int_{\mathbb{C}} e(z t) \exp(-\pi |t|^2) dt\). Thus,
\[
\Pr(\sum_{i=1}^{n} a_i x_i + b_i x'_i \in B(a, r)) \leq \exp(\pi r^2) \int_{\mathbb{C}} \Pr(\sum_{i=1}^{n} a_i x_i + b_i x'_i t e(-at) \exp(-\pi |t|^2) dt.
\]

Next, because of independence we have \(|\mathbb{E}(\sum_{i=1}^{n} a_i x_i + b_i x'_i t)| = \prod_{i=1}^{n} |\mathbb{E}(a_i x_i + b_i x'_i)|\), and so
\[
|\mathbb{E}(a_i x_i + b_i x'_i t)| \leq |\mathbb{E}(a_i x_i + b_i x'_i t)|^2 / 2 + 1 / 2
\]
\[
= \mathbb{E}(\sum_{i=1}^{n} a_i x_i + (a_i x_i + b_i x'_i t) / 2 + 1 / 2
\]
\[
= \mathbb{E}(\sum_{i=1}^{n} a_i x_i + (a_i x_i + b_i x'_i t) \cos(2\pi \text{Re}(\xi_1 - \xi'_1) a_i t + (\xi_2 - \xi'_2) b_i t)) / 2 + 1 / 2
\]
\[
\leq \exp\left(-\mathbb{E}(\sum_{i=1}^{n} a_i x_i + (a_i x_i + b_i x'_i t) \|\text{Re}(2(\xi_1 - \xi'_1) a_i t + (\xi_2 - \xi'_2) b_i t)\|^2_{\mathbb{R}/\mathbb{Z}})\right),
\]

where the random vector \((\xi_1', \xi_2')\) is an identical independent copy of \((\xi_1, \xi_2)\), and in the last inequality we estimated crudely \(|\cos(\pi z)| \leq 1 - \sin^2(\pi z) / 2 \leq 1 - 2\|z\|^2_{\mathbb{R}/\mathbb{Z}} < \exp(-\|z\|^2_{\mathbb{R}/\mathbb{Z}})\).

Observe that, as \((\xi_1, \xi_2)\) belongs to a given \((\mu, \rho)\)-family, so does the pair \((\omega_1, \omega_2) := ((\xi_1 - \xi'_1) / 2, (\xi_2 - \xi'_2) / 2)\). Intuitively, for \(E|\psi_1|^2 = E|\psi_2|^2 = 1\) and \(|\rho| = |E|\psi_1 \psi_2|| < 1\), these two random variables are essentially not multiple of each other. We summarize this useful fact as a claim below.

**Claim A.2.** Assume that \((\omega_1, \omega_2)\) belongs to a given \((\mu, \rho)\)-family. Then there exist positive numbers \(\alpha, \delta, c_0, C_0\) and two Lebesgue-measurable sets \(R_1\) and \(R_2\) in the set \(\{x, y \in \mathbb{C}^2, c_0 < |x|, |y| < C_0\}\) such that \(\Pr((\omega_1, \omega_2) \in R_1), \Pr((\omega_1, \omega_2) \in R_2) \geq \delta\) and \(|a/b - c/d| > \alpha\) for any \((a, b) \in R_1\) and \((c, d) \in R_2\).

**Proof. (of Claim A.2)** Let \(\epsilon_0\) be a sufficiently small positive constant to be chosen. There exist positive numbers \(c_0, C_0\) depending on \(\omega_1, \omega_2\) and on \(\epsilon_0\) such that the truncated random variables \(\psi_1 := \omega_1 1_{|\omega_1| < C_0}, \psi_2 := \omega_2 1_{|\omega_2| < C_0}\) satisfy the following
In summary, we have obtained two closed sub-regions \( R_i, R_j \) of \( R \) such that the corresponding squares \( Q_{ia}, Q_{ja} \) are greater than \( \delta \). By definition, as \( Q_{ia} \) and \( Q_{ja} \) are not adjacent, we have \( |a/b - c/d| \geq 2C_0/kc_0 \) as long as \( (a, b) \in R_{ia} \) and \( (c, d) \in R_{ja} \), completing the proof. \( \square \)

We now apply Claim \( A.1 \) and \( A.2 \) to prove Theorem 3.2. Our method here follows [33] with non-trivial modifications.

**Proof.** (of Theorem 3.2) For short, set \( a_i' := \beta^{-1}a_i, b_i' := \beta^{-1}b_i \). Also, we will denote by \( z' \) the random variables \( 2(\xi_1 - \xi_2) \) and \( 2(\xi_1' - \xi_2') \) respectively, where \( (\xi_1, \xi_2) \) is an identical independent of \( (\xi_1', \xi_2') \). By definition, we have

\[
(1) \quad 1 - \epsilon_0 \leq E[\psi_1^2, E[\psi_2]^2] \leq 1 + \epsilon_0.
\]

\[
(2) \quad |\rho| - \epsilon_0 \leq |E[\psi_1 \psi_2]| \leq |\rho| + \epsilon_0.
\]

Observe that it suffices to justify the claim for the truncated pair \((\psi_1, \psi_2)\). Set \( k \) to be a sufficiently large integer. We divide the square \( Q := \{ z \in C, |\text{Im}(z)|, |\text{Re}(z)| \leq C_0/c_0 \} \) into \( k^2 \) closed smaller squares \( Q_1, \ldots, Q_{k^2} \) of size \( 2C_0/kc_0 \) each, and then divide the region \( R := \{(x, y) \in C^2, c_0 < |x|, |y| < C_0 \} \) into \( k^2 \) closed regions \( R_i, i = 1, \ldots, k^2 \) depending on whether \( x/y \) belongs to \( Q_i \) or not. Note that if \((x, y) \in R \) then the complex number \( x/y \) has absolute value bounded from above and below by \( C_0/c_0 \) and \( c_0/C_0 \) respectively, and so \( x/y \in Q \).

We next claim that for sufficiently small \( \delta > 0 \) (chosen dependently on \( c_0, C_0, k \)), there are squares \( Q_{ia}, Q_{ja} \) that are not adjacent (i.e. sharing a common edge) and that \( P(\psi_1/\psi_2 \in Q_{ia}) \geq \delta \) and \( P(\psi_1/\psi_2 \in Q_{ja}) \geq \delta \). Indeed, assuming otherwise, then \( P(\psi_1/\psi_2 \in Q_i) < \delta \) holds for all but at most \( 9 \) adjacent squares. The larger square \( Q' \) formed by these adjacent ones has size at most \( 6C_0/kc_0 \) which satisfies

\[
P(\psi_1/\psi_2 \in Q') \geq 1 - (k^2 - 9)\delta.
\]

We now concentrate on the event \( \psi_1/\psi_2 \in Q' \). Because of the definition, there exists a number \( c \) such that if \( x/y \in Q' \) then the difference \( |x/y - c| \) can be bounded crudely by \( 6C_0/kc_0 \). Without loss of generality, we assume that \( |c| \geq 1 \). (Otherwise we consider the ratio \( \psi_2/\psi_1 \) instead). Clearly,

\[
|E[\psi_1 \psi_2]| \geq |E[\psi_1 \psi_21_{\psi_1/\psi_2 \in Q'}]| - |E[\psi_1 \psi_2(1 - 1_{\psi_1/\psi_2 \in Q'})]|.
\]

The expectation of the second term can be bounded crudely from above by \( C_0^2(k^2 - 9)\delta \), while the expectation of the first term can be bounded from below by \((|c| - 6C_0/kc_0)|E[\psi_1|^2 - C_0^2(k^2 - 9)\delta \), which is at least \((1 - 6C_0/kc_0)(1 - \epsilon_0) - C_0^2(k^2 - 9)\delta \) because \( |c| \geq 1 \) and \( E[\psi_1|^2 \geq 1 - \epsilon_0 \) from item (1) above. Finally, by choosing \( k \) to be large enough (depending on \( \epsilon_0, C_0, k \)) and \( \delta \) to be small enough (depending on \( \epsilon_0, C_0, k \)), we obtain a lower bound \( 1 - 2\epsilon_0 \) for \( E[\psi_1 \psi_2] \). This is impossible as from item (2) we have \( |E[\psi_1 \psi_2]| \leq |\rho| + \epsilon_0 < 1 - 2\epsilon_0 \).

In summary, we have obtained two closed sub-regions \( R_{ia}, R_{ja} \) of \( R \) such that the corresponding squares \( Q_{ia}, Q_{ja} \) are not adjacent and that both \( P(\psi_1/\psi_2 \in Q_{ia}) \) and \( P(\psi_1/\psi_2 \in Q_{ja}) \) are greater than \( \delta \). By definition, as \( Q_{ia} \) and \( Q_{ja} \) are not adjacent, we have \( |a/b - c/d| \geq 2C_0/kc_0 \) as long as \( (a, b) \in R_{ia} \) and \( (c, d) \in R_{ja} \), completing the proof. \( \square \)
\[ \gamma = \sup_a \mathbf{P}(|\sum_i (a_i x_i + b_i x_i') - a| \leq \beta) = \sup_a \mathbf{P}(|\sum_i (a'_i x_i + b'_i x_i') - a| \leq 1) = n^{-O(1)}. \]

Set \( M := 2A \log n \) where \( A \) is large enough. From Claim A.1 and the fact that \( \gamma \geq n^{-O(1)} \) we easily obtain

\[
\frac{\gamma}{2} \leq \int_{|t| \leq M} \exp\left(-\sum_{i=1}^n \mathbf{E}_{\xi_i, \xi_i' \xi_i''} \| \text{Re} \left( 2(\xi_1 - \xi_1') a'_i t + 2(\xi_2 - \xi_2') b'_i t \right) \|_{\mathbb{R}/\mathbb{Z}}^2 - \pi |t|^2 \right) dt \\
= \int_{|t| \leq M} \exp\left(-\sum_{i=1}^n \mathbf{E}_{z, z'} \| \text{Re} \left( za'_i t + z'b'_i t \right) \|_{\mathbb{R}/\mathbb{Z}}^2 - \pi |t|^2 \right) dt. \tag{59}
\]

**Large level sets.** For each integer \( 0 \leq m \leq M \) we define the level set

\[ S_m := \left\{ t \in \mathbb{C} : \sum_{i=1}^n \mathbf{E}_{z, z'} \| \text{Re} \left( za'_i t + z'b'_i t \right) \|_{\mathbb{R}/\mathbb{Z}}^2 + |t|^2 \leq m \right\}. \]

Then it follows from (59) that \( \sum_{m \leq M} \mu(S_m) \exp\left(-\frac{m}{2} + 1\right) \geq \gamma \), where \( \mu(.) \) denotes the Lebesgue measure of a measurable set. Hence there exists \( m \leq M \) such that \( \mu(S_m) \geq \gamma \exp\left(\frac{m}{2} - 2\right) \).

Next, since \( S_m \subset B(0, \sqrt{m}) \), by the pigeon-hole principle there exist an absolute constant \( c \) and a ball \( B(x_0, \frac{1}{2}) \subset B(0, \sqrt{m}) \) such that

\[ \mu(B(x_0, \frac{1}{2}) \cap S_m) \geq c \mu(S_m)m^{-1} \geq c \gamma \exp\left(\frac{m}{4} - 2\right)m^{-1}. \]

Consider \( t_1, t_2 \in B(x_0, 1/2) \cap S_m \). By Cauchy-Schwarz inequality (note that \( \mathbf{E}_{z, z'} \| \text{Re}(za'_i t + z'b'_i t) \|_{\mathbb{R}/\mathbb{Z}}^2 \) is a norm in \( t \)) we have

\[
\sum_{i=1}^n \mathbf{E}_{z, z'} \| \text{Re} \left( za'_i (t_1 - t_2) + z'b'_i (t_1 - t_2) \right) \|_{\mathbb{R}/\mathbb{Z}}^2 \\
\leq 2 \left( \sum_{i=1}^n \mathbf{E}_{z, z'} \| \text{Re} \left( za'_i t_1 + z'b'_i t_1 \right) \|_{\mathbb{R}/\mathbb{Z}}^2 + \sum_{i=1}^n \mathbf{E}_{z, z'} \| \text{Re} \left( za'_i t_2 + z'b'_i t_2 \right) \|_{\mathbb{R}/\mathbb{Z}}^2 \right) \leq 4m.
\]

Since \( t_1 - t_2 \in B(0, 1) \) and \( \mu(B(x_0, \frac{1}{2}) \cap S_m - B(x_0, \frac{1}{2}) \cap S_m) \geq \mu(B(x_0, \frac{1}{2}) \cap S_m) \), if we put

\[ T := \{ t \in B(0, 1), \sum_{i=1}^n \mathbf{E}_{z, z'} \| \text{Re} \left( za'_i t + z'b'_i t \right) \|_{\mathbb{R}/\mathbb{Z}}^2 \leq 4m \}, \]
then

\[ \mu(T) \geq c\gamma \exp\left(\frac{m}{4} - 2m^{-1}\right). \]

**Discretization.** Choose \( N \) to be a sufficiently large prime (depending on the set \( T \)). Define the discrete box

\[ B_0 := \left\{ \frac{k_1}{N} + \sqrt{-1}k_2/N : k_1, k_2 \in \mathbb{Z}, -N \leq k_1, k_2 \leq N \right\}. \]

We consider all the shifted boxes \( z + B_0 \), where \((\text{Re} z, \text{Im} z) \in [0, 1/N]^2\). By the pigeon-hole principle, there exists \( z_0 \) such that the size of the discrete set \((z_0 + B_0) \cap T\) is at least the expectation, \(|(z_0 + B_0) \cap T| \geq N^2 \mu(T)\) (to see this, we first consider the case when \( T \) is a box itself).

Let us fix some \( t_0 \in (z_0 + B_0) \cap T \). Then for any \( t \in (t_0 + B_0) \cap T \) we have

\[
\sum_{i=1}^{n} \mathbb{E}_{z,z'} \| \text{Re} \left( za'_i(t - t_0) + z'b'_i(t - t_0) \right) \|_{R/Z}^2 \\
\leq 2 \sum_{i=1}^{n} \mathbb{E}_{z,z'} \| \text{Re} \left( za'_i + z'b'_i \right) \|_{R/Z}^2 \\
+ 2 \sum_{i=1}^{n} \mathbb{E}_{z,z'} \| \text{Re} \left( za'_i t_0 + z'b'_i t_0 \right) \|_{R/Z}^2 \leq 16m.
\]

Notice that \( t_0 - t \in B_1 := B_0 - B_0 = \{ \frac{k_1}{N} + \sqrt{-1}k_2/N : k_1, k_2 \in \mathbb{Z}, -2N \leq k_1, k_2 \leq 2N \}\). Thus there exists a subset \( S \) of size at least \( cN^2 \gamma \exp\left(\frac{m}{4} - 2m^{-1}\right) \) of \( B_1 \) such that the following holds for any \( s \in S \)

\[
\sum_{i=1}^{n} \mathbb{E}_{z,z'} \| \text{Re} \left( za'_i + z'b'_i s \right) \|_{R/Z}^2 \leq 16m.
\]

**Double counting and separation.** By definition of \( S \), we have

\[
\mathbb{E}_{z,z'} \sum_{s \in S} \sum_{i=1}^{n} \| \text{Re} \left( za'_i s + z'b'_i s \right) \|_{R/Z}^2 \leq 16m|S|.
\]

Notice that, for \( z = 2(\xi_1 - \xi'_1) \) and \( z' = 2(\xi_2 - \xi'_2) \), \((z/4, z'/4)\) belongs to the \((\mu, \rho)\)-family. By Claim A.2 there exist \((c_1, c_2) \in \mathcal{R}_1\) and \((c'_1, c'_2) \in \mathcal{R}_2\) such that
\[ \sum_{s \in S} \sum_{i=1}^{n} \| \text{Re} \left( (4c_1 a'_i + 4c_2 b'_i)s \right) \|_{\mathbb{R}/\mathbb{Z}}^2 \leq 16\delta^{-1} m|S| \]

and

\[ \sum_{s \in S} \sum_{i=1}^{n} \| \text{Re} \left( (4c'_1 a'_i + 4c'_2 b'_i)s \right) \|_{\mathbb{R}/\mathbb{Z}}^2 \leq 16\delta^{-1} m|S|. \]

From now on, for brevity, we denote by \( v_i \) the complex number \( 4c_1 a'_i + 4c_2 b'_i \) for \( 1 \leq i \leq n \), and by \( v_{n+i} \) the complex number \( 4c'_1 a'_i + 4c'_2 b'_i \) for \( 1 \leq i \leq n \). We then have

\[ \sum_{s \in S} \sum_{i=1}^{2n} \| \text{Re}(v_i s) \|_{\mathbb{R}/\mathbb{Z}}^2 \leq 32\delta^{-1} m|S|. \]

**Switching to \( \mathbb{R}^2 \).** Next, for convenience, we view each \( v_i \) as the vector \((\text{Re} v_i, \text{Im} v_i)\) and each \( s \in S \) as the vector \((\text{Re} s, -\text{Im} s)\) of \( \mathbb{R}^2 \). So we can write \( \text{Re}(v_i s) \) as \( \langle v_i, s \rangle \), and thus obtain the new estimate in \( \mathbb{R}^2 \),

\[ \sum_{s \in S} \sum_{i=1}^{2n} \| \langle v_i, s \rangle \|_{\mathbb{R}/\mathbb{Z}}^2 \leq 32\delta^{-1} m|S|. \]

Let \( n' \) be any number between \( n' \) and \( 2n \). We say that an index \( 1 \leq i \leq 2n \) is bad if

\[ \sum_{s \in S} \| \langle v_i, s \rangle \|_{\mathbb{R}/\mathbb{Z}}^2 \geq \frac{32\delta^{-1} m|S|}{n'}. \]

Then the number of bad indices is at most \( n' \). Let \( V \) be the set of remaining \( v_i \)'s. Thus \( V \) contains at least \( 2n - n' \) elements. In the rest of the proof, we are going to show that the set \( V \) is close to a GAP.

**Dual sets.** Consider an arbitrary good index \( i \), we have

\[ \sum_{s \in S} \| \langle s, v_i \rangle \|_{\mathbb{R}/\mathbb{Z}}^2 \leq 32\delta^{-1} m|S|/n'. \]

Set \( k := \sqrt{ \frac{n'}{2048\delta^2 m^{-1}}} \) and let \( V_k := k(V \cup \{0\}) \). By Cauchy-Schwarz inequality, for any \( v \in V_k \) we have
The elliptic law

\[ \sum_{s \in S} 2\pi^2 \|\langle s, v \rangle \|_{\mathbb{R}/\mathbb{Z}}^2 \leq \frac{|S|}{2}, \]

which implies

\[ \sum_{s \in S} \cos(2\pi \langle s, v \rangle) \geq \frac{|S|}{2}. \]

Observe that for any \( x \in C(0, \frac{1}{512}) \) (the ball of radius \( 1/512 \) in the \( \|\cdot\|_\infty \) norm) and any \( s \in S \subset C(0, 2) \) we always have \( \cos(2\pi \langle s, x \rangle) \geq 1/2 \) and \( \sin(2\pi \langle s, x \rangle) \leq 1/12 \). Thus for any \( x \in C(0, \frac{1}{512}) \),

\[ \sum_{s \in S} \cos(2\pi \langle s, (v + x) \rangle) \geq \frac{|S|}{4} - \frac{|S|}{12} = \frac{|S|}{6}. \]

On the other hand,

\[ \int_{x \in [0,N]^2} \left( \sum_{s \in S} \cos(2\pi \langle s, x \rangle) \right)^2 dx \leq \sum_{s_1, s_2 \in S} \int_{x \in [0,N]^2} \exp(2\pi i \langle s_1 - s_2, x \rangle) dx \ll |S| N^2. \]

Hence we deduce the following

\[ \mu_{x \in [0,N]^2} \left( \left( \sum_{s \in S} \cos(2\pi \langle s, x \rangle) \right)^2 \right) \geq \left( \frac{|S|}{6} \right)^2 \ll \frac{|S| N^2}{(|S|/6)^2} \ll \frac{N^2}{|S|}. \]

Now use the fact that \( S \) has large size, \( |S| \gg N^2 \gamma \exp(\frac{m}{4} - 2)m^{-1} \), and \( N \) was chosen to be large enough so that \( V_k + C(0, \frac{1}{512}) \subset [0,N]^2 \), we have

\[ \mu(V_k + C(0, \frac{1}{512})) \ll \gamma^{-1} \exp(-\frac{m}{4} + 2)m. \]

Thus, we have obtained the following

\[ \mu \left( k(V \cup \{0\}) + C(0, \frac{1}{256}) \right) \ll \gamma^{-1} \exp(-\frac{m}{4} + 2)m. \quad (60) \]

The long range inverse theorem. Our next analysis relies on the following result from [33].
Theorem A.3. [33, Theorem 3.2] Let $\alpha > 0$ be constant. Assume that $X$ is a subset of a torsion-free group such that $0 \in X$ and $|kX| \leq k^\alpha |X|$ for some integer $k \geq 2$ that may depend on $|X|$. Then, there is proper symmetric GAP $P$ of rank $r = O(\alpha)$ and cardinality $O_\alpha(k^{-r}|kX|)$ such that $X \subset P$.

Let $D := 2048 \times 16 \times \delta^{-1} = \Theta(\delta^{-1})$. We approximate each vector $v$ of $V$ by a closest vector $u \in (\mathbb{Z}/D)^2$, $\|v - u\|_2 \leq \frac{\sqrt{d}}{D}$, with $u \in \mathbb{Z}^2$.

Let $U$ be the collection of all such $u$. Since $\sum_{v \in V} \|v\|_2^2 = O(\beta^{-2})$, we have

$$\sum_{u \in U} \|a\|_2^2 = O_\delta(\beta^2 \delta^{-2}). \quad (61)$$

It follows from (60) that

$$|k(U + C_0(0, 1))| = O_\gamma(Dk)^2(z_0)^{-2} \exp(-\frac{m}{4} + 2)m)$$

$$= O_\gamma^{-1}k^2 \exp(-\frac{m}{4} + 2)m ,$$

where $C_0(0, r)$ is the discrete cube $\{(x_1, x_2) \in \mathbb{Z}^2 : |x_i| \leq r\}$.

Now we apply Theorem A.3 to the set $U + C_0(0, 1)$ (notice that $0 \in U$). That lemma implies there exists a proper GAP $P = \{\sum_{i=1}^r x_i g_i : |x_i| \leq N_i \} \subset \mathbb{Z}^2$ containing $U + C_0(0, 1)$ which has small rank $r = O(1)$, and small size

$$|P| = O_\gamma^{-1}k^2 \exp(-\frac{m}{4} + 2)m k^{-r}$$

$$= O(\gamma^{-1}n^{(-r+2)/2}).$$

Moreover, we have learned from Lemma [33, Lemma 4.4] that $kP$ can be contained in a set $ck(U + C_0(0, 1))$ for some $c = O(1)$. Using (61), we conclude that all the generators $g_i$ of $P$ are bounded,

$$\|g_i\|_2 = O(k\beta^{-1}).$$

Next, since $C_0(0, 1) \subset Q$, the rank $r$ of $P$ is at least 2. We consider the following two cases.

Case 1: $r \geq 3$. Recall that $|P| = O(\gamma^{-1}n^{(3-r)/2}) = O(\gamma^{-1}/\sqrt{n^r})$. Let
\[ Q := \frac{\beta}{Dk} \cdot P. \]

It is clear that \( Q \) satisfies all of the conditions of Theorem 3.2. (Note that, in this case, we obtain a stronger approximation; almost all elements of \( V \) are \( O(\frac{\beta \log n}{\sqrt{m}}) \)-close to \( Q \).)

**Case 2**: \( r = 2 \). Because the unit vectors \( e_1 = (1, 0), e_2 = (0, 1) \) belong to \( P = \{ \sum_{i=1}^{2} x_i g_i : |x_i| \leq N_i \} \subset \mathbb{Z}^2 \); the set of generators \( g_1, g_2 \) forms a base with the unit determinant of \( \mathbb{R}^2 \).

In \( P \), consider the set of lattice points with all coordinates divisible by \( k \). We observe that (for instance, by [46, Theorem 3.36]) this set can be contained in a GAP \( P' \) of rank 2 and cardinality at most \( \max (O(\frac{1}{k^r}|P|, 1) = \max \left( O(\frac{\gamma - 1}{n^{r/2}}), 1 \right) \). (Here, we use the bound \( |P| = O(\gamma^{-1} \exp(-\frac{m}{4}) m) \).) Next, define

\[ Q := \frac{\beta}{Dk} \cdot P'. \]

It is easy to verify that \( Q \) satisfies all of the conditions of Theorem 3.2. (Note that, in this case, we obtain a stronger bound on the size of \( Q \).) □

**Appendix B. Proof of Lemma 4.3**

Set \( a'_{ij} := a_{ij}/\beta \). By definition,

\[ \gamma = \sup_{a,b,b_i} \mathbf{P}_{x,x'} \left( \left| \sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a \right| \leq 1 \right) \geq n^B. \]

By Markov’s inequality we have

\[
\mathbf{P}_{x,x'} \left( \left| \sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a \right| \leq 1 \right) \\
= \mathbf{P} \left( \exp \left( -\frac{\pi}{2} \left| \sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a \right|^2 \right) \geq \exp \left( -\frac{\pi}{2} \right) \right) \\
\leq \exp \left( \frac{\pi}{2} \mathbf{E}_{x,x'} \exp \left( -\frac{\pi}{2} \left| \sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a \right|^2 \right) \right) \\
\leq \exp \left( \frac{\pi}{2} \int_{C} |\mathbf{E}_{x,x'} e \left( \left( \sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i \right) \cdot t \right)| \exp \left( -\frac{\pi}{2} |t|^2 \right) dt \right) \\
\leq \exp \left( \frac{\pi}{2} (\sqrt{2}\pi)^2 \int_{C} |\mathbf{E}_{x,x'} e \left( \left( \sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i \right) \cdot t \right)| \exp \left( -\frac{\pi}{2} |t|^2 \right) / (\sqrt{2}\pi)^2 dt \right)
\]
where in the fourth equation we used the identity \( \exp(-\frac{\pi}{2} |x|^2) = \int_C e(xt) \exp(-\frac{\pi}{2} |t|^2) dt \).

Consider \( x = (x_1, \ldots, x_n) \) as \((x_U, x_U')\) and \( x' = (x'_1, \ldots, x'_n) \) as \((x'_U, x'_U')\), where \( x_U, x'_U \) and \( x_U', x'_U' \) are the vectors corresponding to \( i \in U \) and \( i \notin U \) respectively. After a series of applications of the identity \( \int_C \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt = 1 \) and Cauchy-Schwarz inequality, we obtain

\[
\left[ \int_C \left| \mathbf{E}_{x, x'} e \left( \sum_{i,j} a_{ij}^{(x)} x_i x_j' + \sum_i b_i x_i + \sum_i b'_i x_i' \cdot t \right) \right| \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \right]^4 \\
\leq \left[ \int_C \left| \mathbf{E}_{x, x'} e \left( \sum_{i,j} a_{ij}^{(x)} x_i x_j' + \sum_i b_i x_i + \sum_i b'_i x_i' \cdot t \right) \right|^2 \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \right]^2 \\
= \left[ \int_C \left| \mathbf{E}_{x_U, x'_U} e \left( \sum_{i,j} a_{ij}^{(x)} x_i x_j' + \sum_i b_i x_i + \sum_i b'_i x_i' \cdot t \right) \right|^2 \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \right]^2 \\
\leq \int_C \left| \mathbf{E}_{x_U, x'_U} e \left( \sum_{i,j} a_{ij}^{(x)} x_i x_j' + \sum_i b_i x_i + \sum_i b'_i x_i' \cdot t \right) \right|^2 \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \\
= \int_C \left| \mathbf{E}_{x_U, x'_U} e \left( \sum_{i,j} a_{ij}^{(x)} x_i x_j' + \sum_i b_i x_i + \sum_i b'_i x_i' \cdot t \right) \right|^2 \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \\
\leq \int_C \left| \mathbf{E}_{x_U, x'_U} e \left( \sum_{i,j} a_{ij}^{(x)} x_i x_j' + \sum_i b_i x_i + \sum_i b'_i x_i' \cdot t \right) \right|^2 \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \\
= \int_C \left( \sum_{i,j} a_{ij}^{(x)} v_i w_j + \sum_{i,j} a_{ij}^{(x')} v_i w_j \right) \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \\
= (1/\sqrt{2\pi})^2 \mathbf{E}_{v, w} \exp(-\frac{\pi}{2} |v|^2) \sum_{i,j} a_{ij}^{(x)} v_i w_j + \sum_{i,j} a_{ij}^{(x')} v_i w_j, \quad (62)
\]

where \((y_U, y'_U)\) and \((y'_U, y'_U')\) are independent identical copies of \((x_U, x'_U)\) and \((x'_U, x'_U')\) respectively, and \(v := x - y, \ w := x' - y'.\)

Thus
\[\gamma^4 = \left( P_{x,x'}\left( |\sum_{i,j} a'_{ij} x_i x_j + \sum_i b_i x_i + \sum_i b'_i x_i - a| \leq 1 \right) \right)^4 \]
\[\leq \exp(4\pi)(2\pi)^4 \left( \int_C \left| E_{x,x'} e^{\left( \sum_{i,j} a'_{ij} x_i x_j + \sum_i b_i x_i + \sum_i b'_i x_i \right) \cdot t} \right| \exp(-\frac{\pi}{2} |t|^2)/(\sqrt{2\pi})^2 dt \right)^4 \]
\[\leq \exp(4\pi)(2\pi)^3 E_{v,w} \exp\left( -\frac{\pi}{2} \sum_{i,j} a'_{ij} v_i w_j + \sum_{i,j} a'_{ij} v_i w_j \right). \]

Because \( \gamma \geq n^{-B} \), the inequality above implies that

\[ P_{v,w}\left( |\sum_{i,j} a'_{ij} v_i w_j + \sum_{i,j} a'_{ij} v_i w_j| = O_B(\sqrt{\log n}) \right) \geq \frac{1}{2} \gamma^4 / ((2\pi)^3 \exp(4\pi)). \]

Scaling back to \( a_{ij} \), we thus obtain

\[ P_{v,w}\left( |\sum_{i,j} a_{ij} v_i w_j + \sum_{i,j} a_{ij} v_i w_j| = O_B(\beta \sqrt{\log n}) \right) \geq \frac{1}{2} \gamma^4 / ((2\pi)^3 \exp(4\pi)), \]

completing the proof.

**References**

[1] Z. D. Bai, *Circular law*, Ann. Probab. 25 (1997), 494-529.

[2] Z. D. Bai, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica 9, 611-677 (1999).

[3] Z. D. Bai and J. Silverstein, *Spectral analysis of large dimensional random matrices*, Mathematics Monograph Series 2, Science Press, Beijing 2006.

[4] C. Bordenave and D. Chafai, *Around the circular law*, Probability Surveys 9 (2012) 1-89.

[5] K. Costello, T. Tao and V. Vu, *Random symmetric matrices are almost surely non-singular*, Duke Math. J. 135 (2006), 395-413.

[6] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. 9 (1988), no. 4, 543-560.

[7] A. Edelman, *The Probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular Law*, J. Multivariate Anal. 60, 203-232, (1997).

[8] L. Erdős, *Universality of Wigner random matrices: a survey of recent results*, arxiv.org/abs/1004.0881.

[9] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. 51 (1945), 898-902.

[10] J. Ginibre, *Statistical ensembles of complex, quaternion and real matrices*, J. Math. Phys. 6 (1965), 440-449.

[11] V. L. Girko, *Circular law*, Theory Probab. Appl. (1984), 694-706.

[12] V. L. Girko, *The strong circular law, twenty years later*, II. Random Oper. Stochastic Equations 12 (2004), no. 3, 255-312.
[13] V. L. Girko, *Elliptic law: ten years later I*, Random Operators and Stochastic Equations, V. 3, N. 3, (1995), 257-302.
[14] V. L. Girko, *The Elliptic law: ten years later II*, Random Operators and Stochastic Equations, V. 3, N.4, (1995), 377-398.
[15] V. L. Girko, *Strong elliptic law*, Random Operators and Stochastic Equations, V.5, N.3, (1997), 269-306.
[16] V. L. Girko, *The strong elliptic law, twenty years later*, I. Random Oper. Stochastic Equations 12 (2006), no. 1, 59-102.
[17] V. L. Girko, *The strong elliptical galactic law. Sand clock density, twenty years later*, II. Random Oper. Stochastic Equations 12 (2006), no. 2, 157-208.
[18] F. Götze and A. N. Tikhomirov, *The circular law for random matrices*, Ann. Prob., 38 (2010), no. 4, 1444-1491.
[19] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab.5 (2000) 119-136.
[20] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. 8 (1977), no. 3-4, 197-211.
[21] K. Johansson, *From Gumbel to Tracy-Widom*, Probab. Theory Related Fields, 138 (2007), no. 1-2, p. 75–112.
[22] B. Khoruzhenko, H.-J. Sommers, Chapter 18 of *The Oxford Handbook of Random Matrix Theory*, edited by Gernot Akemann, Jinho Baik, and Philippe Di Francesco, Oxford University Press (2011).
[23] D. Kleitman, *On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors*, Advances in Math. 5 (1970), 155-157.
[24] M. Ledoux, *Complex Hermite polynomials: from the semi-circular law to the circular law*, Commun. Stoch. Anal. 2 (2008), no. 1, p. 27–32.
[25] J. W. Lindeberg, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, Math. Z. 15 (1922), 211-225.
[26] J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation*. III. Rec. Math. Mat. Sbornik N. S. 12, (1943), 277-286.
[27] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Acad. Press. (1967).
[28] M. L. Mehta, *Random Matrices*, third edition. Elsevier/Academic Press, Amsterdam, (2004).
[29] A. Naumov, *Elliptic law for real random matrices*, arxiv.org/abs/1201.1639.
[30] H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Mathematics Journal Vol. 161, 4 (2012), 545-586.
[31] H. Nguyen, *A continuous variant of the inverse Littlewood-Offord problem for quadratic forms*, to appear in Contribution to Discrete Mathematics.
[32] H. Nguyen, *On the least singular value of random symmetric matrices*, Electron. J. Probab.,17 (2012), no. 53 1-19.
[33] H. Nguyen and V. Vu, *Optimal Littlewood-Offord theorems*, Advances in Math., Vol. 226 6 (2011), 5298-5319.
[34] G. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, Journal of Multivariate Analysis, 101 (2010), 645-656.
[35] L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, Mathematical Surveys and Monographs, American Mathematical Society, 2011.
[36] M. Rudelson and R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600-633.
[37] M. Talagrand *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes tudes Sci. Publ. Math. No. 81 (1995), 73-205.
[38] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. (N.S.) 46 (2009), no. 3, 377-396.
[39] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2) 169 (2009), no 2, 595-632.

[40] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, Journal of the A. M. S 20 (2007), 603-673.

[41] T. Tao and V. Vu, *Random matrices: the circular law*, Communication in Contemporary Mathematics 10 (2008), 261-307.

[42] T. Tao and V. Vu, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. 38 (2010), no. 5p. 2023-2065, with an appendix by M. Krishnapur.

[43] T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. 206 (2011), no. 1, 127-204.

[44] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, Mathematics of Computation, 79 (2010), 2333-2352.

[45] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values*, Geom. Funct. Anal. 20 (2010), no. 1, 260-297.

[46] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Univ. Press, 2006.

[47] R. Vershynin, *Invertibility of symmetric random matrices*, to appear in Random Structure and Algorithms.

[48] E. P. Wigner, *On the distributions of the roots of certain symmetric matrices*, Ann. Math. 67, 325-327.

**Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA**

*E-mail address*: hoing@math.upenn.edu

**Department of Mathematics, Rutgers, Piscataway, NJ 08854, USA**

*E-mail address*: sdo21@math.rutgers.edu