H. Nguyen and S. O'Rourke. (2009) "The Elliptic Law,"
International Mathematics Research Notices, Vol. 2009, Article ID rnn999, 45 pages.
doi:10.1093/imrn/rnn999

The Elliptic Law

Hoi H. Nguyen[1] **and Sean O'Rourke**[2]

[1]The Ohio State University
Department of Mathematics
231 West 18th Avenue
Columbus, OH 43210

[2]Yale University
Department of Mathematics
PO Box 208283
New Haven, CT 06520

Correspondence to be sent to: sean.orourke@yale.edu

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 μ_{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

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

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 [52]. In particular, Wigner showed that, under some additional moment

Received 1 January 2009; Revised 11 January 2009
Communicated by A. Editor

and symmetry assumptions on the entries, the ESD of such a matrix converges to the semi-circular law $F_{\rm sc}$ with density given by

$$F'_{\rm sc}(x) := \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & -2 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}.$$

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)**.** Let ζ be a real random variable, and let ξ be a complex random variable with variance one. For each $n \ge 1$, assume 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. copies of ζ and those above the diagonal are i.i.d copies of ξ. 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 [29]. 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 $F_{\rm cir}$ given by

$$F_{\rm cir}(x,y) := \frac{1}{\pi}\,\mathrm{mes}\Big(|z| \le 1 : \mathrm{Re}(z) \le x, \mathrm{Im}(z) \le y\Big).$$

In other words, $F_{\rm cir}$ 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 [11]. The real Gaussian case was studied by Edelman in [8]. For the general (non-Gaussian) case when there is no formula, the problem appears much more difficult. Important results were obtained by Girko [12, 13], Bai [1, 3], and more recently by Götze and Tikhomirov [19], Pan and Zhou [36], and Tao and Vu [45]. 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 [46].

Theorem 1.2 (Circular law for non-Hermitian i.i.d. matrices)**.** Let ξ be a complex-valued random variable with mean zero and variance one. For each $n \ge 1$, assume that the entries of the $n \times n$ matrix X_n are i.i.d. copies of ξ. 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 (ξ_1, ξ_2) be a random vector in $\mathbb{C}^2$ where both ξ_1 and ξ_2 have mean zero and unit variance. Let $\{x_{ij}\}$ be an infinite double array of random variables on $\mathbb{C}$. For each $n \ge 1$ we define the random $n \times n$ matrix $X_n = (x_{ij})_{1 \le i,j \le n}$. We say that the sequence of random matrices $\{X_n\}_{n \ge 1}$ satisfies condition **C0** with atom variables (ξ_1, ξ_2) if the following conditions hold:

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

(ii) (Common distribution) each pair (x_{ij}, x_{ji}), $1 \le i < j$ is an i.i.d. copy of (ξ_1, ξ_2),

(iii) (Flexibility of the main diagonal) the diagonal elements, $\{x_{ii} : i \ge 1\}$, 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 **C0**. We refer the reader to [14, 15, 16, 17, 18] and 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, 8]).

When $\{X_n\}_{n \ge 1}$ is a sequence of random matrices that satisfy condition **C0** with jointly Gaussian atom variables (ξ_1, ξ_2), the joint eigenvalue density can be derived explicitely and the limiting ESD can be computed directly; see [23, 24, 26] and references therein. Recently, Naumov [31] 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

$$\mathcal{E}_\rho := \left\{ z \in \mathbb{C} : \frac{\mathrm{Re}(z)^2}{(1+\rho)^2} + \frac{\mathrm{Im}(z)^2}{(1-\rho)^2} \le 1 \right\}.$$

Theorem 1.4 (Naumov [31])**.** Let $\{X_n\}_{n\geq 1}$ be a sequence of real random matrices that satisfy condition **C0** with real atom variables (ξ_1, ξ_2) where $\mathbf{E}[\xi_1\xi_2] = \rho$, $-1 < \rho < 1$. Also, assume that $\max(\mathbf{E}|\xi_1|^4, \mathbf{E}|\xi_2|^4) < \infty$. Then the ESD of the matrix $\frac{1}{\sqrt{n}}X_n$ converges in probability as $n \to \infty$ to the elliptic law F_ρ with parameter ρ given by

$$F_\rho(x, y) := \frac{1}{\pi(1-\rho^2)} \operatorname{mes}\Big(z \in \mathcal{E}_\rho : \operatorname{Re}(z) \leq x, \operatorname{Im}(z) \leq y\Big).$$

□

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

$$\|M\|_2 := \sqrt{\operatorname{tr}(M^*M)} = \sqrt{\operatorname{tr}(MM^*)}. \tag{1}$$

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 **C0** with real atom variables (ξ_1, ξ_2) where $\mathbf{E}[\xi_1\xi_2] = \rho$, $-1 < \rho < 1$. Assume that $\{F_n\}_{n\geq 1}$ is a sequence of deterministic matrices such that $\operatorname{rank}(F_n) = o(n)$* 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 ρ as $n \to \infty$. □

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

Definition 1.6 ((μ, ρ)-family)**.** Given parameters $0 \leq \mu \leq 1$ and $-1 < \rho < 1$, we say that the complex random variable pair (ξ_1, ξ_2) belongs to the (μ, ρ)-*family* if the following holds.

(i) Both ξ_1 and ξ_2 have mean zero and unit variance;

(ii) $\mathbf{E}[(\operatorname{Re}(\xi_1))^2] = \mathbf{E}[(\operatorname{Re}(\xi_2))^2] = \mu$ and $\mathbf{E}[(\operatorname{Im}(\xi_1))^2] = \mathbf{E}[(\operatorname{Im}(\xi_2))^2] = 1 - \mu$;

(iii) $\mathbf{E}[\operatorname{Re}(\xi_1)\operatorname{Re}(\xi_2)] = \mu\rho$ and $\mathbf{E}[\operatorname{Im}(\xi_1)\operatorname{Im}(\xi_2)] = -(1-\mu)\rho$;

(iv) $\mathbf{E}[\operatorname{Re}(\xi_i)\operatorname{Im}(\xi_j)] = 0$ for any $i, j \in \{1, 2\}$.

□

Remark 1.7. If (ξ_1, ξ_2) belongs to the (μ, ρ)-family, then the covariance matrix of $\xi = (\operatorname{Re}(\xi_1), \operatorname{Im}(\xi_1), \operatorname{Re}(\xi_2), \operatorname{Im}(\xi_2))^{\mathrm{T}}$ is given by

$$\mathbf{E}\xi\xi^{\mathrm{T}} = \begin{pmatrix} \mu & 0 & \mu\rho & 0 \\ 0 & 1-\mu & 0 & -(1-\mu)\rho \\ \mu\rho & 0 & \mu & 0 \\ 0 & -(1-\mu)\rho & 0 & 1-\mu \end{pmatrix}.$$

□

Notice that if (ξ_1, ξ_2) belongs to the (μ, ρ)-family then $\mathbf{E}|\xi_1|^2 = \mathbf{E}|\xi_2|^2 = 1$ and $\mathbf{E}[\xi_1\xi_2] = \rho$. More importantly, we do not require the imaginary and real parts of ξ_1, ξ_2 to be independent.

Theorem 1.8 (Elliptic law for complex random matrices)**.** Let $0 \leq \mu \leq 1$ and $-1 < \rho < 1$ be given. Let $\{X_n\}_{n\geq 1}$ be a sequence of complex matrices such that $\{X_n\}_{n\geq 1}$ satisfies condition **C0** with atom variables (ξ_1, ξ_2) from the (μ, ρ)-family. Assume furthermore that $\{F_n\}_{n\geq 1}$ is a sequence of deterministic matrices such that $\operatorname{rank}(F_n) = o(n)$ 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 ρ as $n \to \infty$. □

In light of the universality phenomenon, we conjecture that Theorem 1.8 continues to hold when $\mathbf{E}|\xi_1|^2 = \mathbf{E}|\xi_2|^2 = 1$ and $\mathbf{E}[\xi_1\xi_2] = \rho$, where ρ is a complex number satisfying $|\rho| < 1$. In this optimal setting, the ESD of $\frac{1}{\sqrt{n}}X_n$ is conjectured to converge to the elliptic law associated with the rotated ellipsoid $\mathcal{E}_\rho$ given by

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

*We use asymptotic notation under the assumption that $n \to \infty$. See Section 1.2 for a complete description of the asymptotic notation used here and throughout the paper.

where $\theta = \mathrm{Arg}(\rho)$. (This formula for the rotated ellipsoid can be derived by multiplying the matrix by $e^{-i\theta/2}$ so that the resulting atom variables have a real-valued correlation.)

One of the key ingredients in the proof of Theorems 1.5 and 1.8 is a lower bound on the least singular value of X_n. If M is a $n \times n$ matrix, 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, the largest and smallest singular values satisfy

$$\sigma_1(M) = \sup_{\|x\|=1} \|Mx\|$$

and

$$\sigma_n(M) = \inf_{\|x\|=1} \|Mx\|,$$

where $\|v\|$ denotes the Euclidean norm of a vector v.

In particular, we will verify the following polynomial bound for the smallest singular value.

Theorem 1.9 (Bound on the least singular value for pertubed random matrices)**.** Assume that $M_n = F_n + X_n$, where the entries of the given complex matrix F_n are bounded by n^α in absolute value, and X_n is a random matrix from Theorem 1.8 for given $0 \leq \mu \leq 1$ and $-1 < \rho < 1$. Then for any $B > 0$, there exists $A > 0$ and $n_0 > 0$ (both depending on B, α, μ, ρ, and the distribution of (ξ_1, ξ_2) and x_{11}) such that

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

for all $n > n_0$. □

Our polynomial bound here is motivated by [45, Lemma 4.1] of Tao and Vu, which plays a fundamental rule in their establishment of the circular law (Theorem 1.2). We also refer the reader to the work [38] of Rudelson and Vershynin for an almost complete treatment for the least singular values of random non-Hermitian matrices with independent entries. Similar techniques have also been used by Götze and Tikhomirov [19] to prove a version of Theorem 1.2. Recently, a similar study for random real symmetric matrices has been carried out independently by Vershynin in [51] and by the first author in [34].

1.1 Overview and Outline

Because of its importance, we prove Theorem 1.9 first. Indeed, in Section 2, we outline the proof of Theorem 1.9. We then complete the proof in Sections 3–6. In Section 7, we use Theorem 1.9 to prove our main results, Theorems 1.5 and 1.8. In particular, Section 7 is independent of Sections 2–6 and can be read separately. The appendix contains a number of auxiliary results.

1.2 Notation

For a $m \times n$ matrix M, we let

$$\sigma_1(M) \geq \cdots \geq \sigma_{\min\{m,n\}}(M) \geq 0$$

denote the singular values of M. We use the notations $\mathbf{r}_i(M)$ and $\mathbf{c}_j(M)$ 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 (1)) and let $\|M\| := \sigma_1(M)$ denote the spectral norm of M.

We consider n an asymptotic parameter tending to infinity. We use $Z \ll Y$, $Y \gg Z$, $Y = \Omega(Z)$, or $Z = O(Y)$ to denote the bound $|Z| \leq CY$ for all sufficient large n for some constant C. Notations such as $Z \ll_k Y$, $Z = O_k(Y)$ mean that the hidden constant C depends on another constant k. $Z = o(Y)$ or $Y = \omega(Z)$ means that $Z/Y \to 0$ as $n \to 0$. We write $Z = \Theta(Y)$ or $Z \asymp Y$ for $Y \ll Z \ll Y$.

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

Let A be an event. Sometimes we will write $\mathbf{P}_{y_1,\ldots,y_k}(A)$ to emphasize that the probability under consideration is taken with respect to the specified random variables $y_1, \ldots, y_k$ (while fixing all other random variables).

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

In this section, we begin the proof of Theorem 1.9. Broadly speaking, our proof follows the approach of [34]. Nevertheless, because the matrix X_n under consideration is much more complicated than a Hermitian matrix, it is of great necessity to generalize and string a series of previous results [35, 33, 34] together. 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 our approach; complete details of the proofs will be presented subsequently.

First of all, we will assume n to be sufficiently large. For the sake of simplicity, we will prove our result under the following condition.

Condition 2.1. With probability one, $|x_{ij}| \leq n^{B+1}$ for all i, j. □

In fact, because all x_{ij} have bounded variance, we have $\mathbf{P}(|x_{ij}| \geq n^{B+1}) = O(n^{-2B-2})$. 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 $M_n\mathbf{x} = \mathbf{y}$ for some $\|\mathbf{x}\|_2 = 1$ and $\|\mathbf{y}\|_2 \leq n^{-A}$. There are two cases to consider.

2.1 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, $C(M_n)\mathbf{y} = \det(M_n) \cdot \mathbf{x}$, and thus we have $\|C(M_n)\mathbf{y}\|_2 = |\det(M_n)|$.

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

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

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

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

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 2.1, and by the bounds $f_{ij} \leq n^{\alpha}$ for the entries of F_n, we have

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

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

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

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

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

Hence, for proving Theorem 1.9, it suffices to justify the following result (after an appropriate modification for A).

Theorem 2.2. For any $B > 0$, there exists $A > 0$ such that

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

□

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

$$\frac{1}{c}\det(M_n) = \frac{1}{c}m_{11}c_{11}(M_n) + \sum_{2\leq i,j\leq n} a_{ij}m_{1i}m_{j1}. \tag{6}$$

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 2.2, 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

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

  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 $\{g_0, \dots, g_r\}, \{K_1, \dots, K_r\}$ and $\{K_1', \dots, K_r'\}$.

It is convenient to think of Q as the image of an integer box $B := \{(k_1, \dots, k_r) \in \mathbf{Z}^r | K_i \leq k_i \leq K_i'\}$ under the linear map

$$\Phi : (k_1, \dots, k_r) \mapsto g_0 + k_1g_1 + \cdots + k_rg_r.$$

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 are the right object to study here. In the sequel we state our main steps rigorously with the help of GAPs.

Theorem 2.3 (Step 1)**.** Let $0 < \epsilon < 1$ be a given constant. Assume that M_{n-1} is fixed and

$$\sup_a \mathbf{P}_{x_2,\dots,x_n,x_2',\dots,x_n'}\left(\left|\sum_{2\leq i,j\leq n} a_{ij}(x_i + f_i)(x_j' + f_j') - a\right| \leq n^{-A}\right) \geq n^{-B}$$

for some sufficiently large integer A, where

- $a_{ij} = c_{ij}(M_{n-1})/c$,
- $f_i = f_{1i}, f_i' = f_{i1}$ are the entries of F_n, and thus fixed,
- (x_i, x_i') are i.i.d copies of (ξ_1, ξ_2) of a given (μ, ρ)-family with $0 \leq \mu \leq 1$ and $-1 < \rho < 1$.

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

- (orthogonality) $\|\mathbf{u}\|_2 \asymp 1$ and either $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le n^{-A/2+O_{B,\epsilon}(1)}$ for $n - O_{B,\epsilon}(1)$ rows of M_{n-1} or $|\langle \mathbf{u}, \mathbf{c}_i(M_{n-1})\rangle| \le n^{-A/2+O_{B,\epsilon}(1)}$ 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 $|p|, |q|, |p'|, |q'| \le n^{A/2+O_{B,\epsilon}(1)}$.

□

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

Theorem 2.4 (Step 2)**.** With respect to M_{n-1}, the probability that there exists a vector $\mathbf{u}$ as in Theorem 2.3 is $\exp(-\Omega(n))$. □

2.2 Case 2.

M_n does not have full rank, which is the case to consider if ξ_1, ξ_2 have discrete distribution. We show that for any fixed $B > 0$ this event holds with probability less than n^{-B} for large enough n depending on B.

First, instead of the entries x_{ij} of X_n, consider $x'_{ij} := (1 - \epsilon^2)x_{ij} + \epsilon\xi_{ij}$, where ξ_{ij} are independently uniform on the interval $[-1, 1]$ and ϵ is very small, say $n^{-1000An}$. 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 1.9 obtained from Case 1 for the matrix M'_n, with probability at least $1 - n^{-B}$ one has $\sigma_n(M'_n) \ge n^{-A}$, and thus

$$|\det(M'_n)| \ge n^{-An}. \tag{7}$$

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

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

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

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

The proof of Theorem 2.3 will be given in Section 5 thanks to useful tools from Sections 3 and 4. Theorem 2.4 will be concluded in Section 6.

3 Anti-concentration, a warm-up

Recall that in the inverse step, Theorem 2.3, we assumed that

$$\sup_a \mathbf{P}_{x_2,\dots,x_n,x'_2,\dots,x'_n}\left(\left|\sum_{2\le i,j\le n} a_{ij}(x_i + f_i)(x'_j + f'_j) - a\right| \le n^{-A}\right) \ge n^{-B}. \tag{8}$$

This can be considered as a high concentration of the bilinear form $\sum_{2\le i,j\le n} a_{ij}(x_i + f_i)(x'_j + f'_j)$ 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 [10] and Littlewood-Offord [28] in the 1940s asserts that if a_i are complex numbers of magnitude $|a_i| \ge 1$, 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 [42, 45, 48] 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 [35]. Before stating this result, we pause to introduce some terminology.

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

Theorem 3.1 (Inverse Littlewood-Offord theorem for linear forms, [35])**.** 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

$$\sup_a \mathbf{P}_{\mathbf{x}}\Big(|\sum_{i=1}^n a_i(x_i + f_i) - a| \leq \beta\Big) = \gamma \geq n^{-B},$$

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

(i) (control of rank and size) Q has small rank, $r = O_{B,\epsilon}(1)$, and small cardinality

$$|Q| \leq \max\left(O_{B,\epsilon}\left(\frac{\gamma^{-1}}{\sqrt{n'}}\right), 1\right);$$

(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 $g_i = \beta\frac{p_i}{p}$, with $p_i \in \mathbf{Z}$ and $p_i = O_{B,\epsilon}(\beta^{-1}\sqrt{n'})$;

(iii) (good approximation) at least $n - n'$ elements of a_i are β-close to Q.

□

Here and later, if not specified, the implied constants are allowed to depend on the distribution of the random variables under consideration. Thus, for instance the implied constants in Theorem 3.1 also depend on α_1, α_2 and α_3. The interested reader is also invited to read [38] for a similar but milder setting of the inverse Littlewood-Offord for linear forms.

To attack Theorem 2.3, 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 (ξ_1, ξ_2) from a given (μ, ρ)-family. Intuitively, as $\mathbf{E}|\xi_1|^2 = \mathbf{E}|\xi_2|^2 = 1$ and $|\mathbf{E}\xi_1\xi_2| = |\rho| < 1$, the random variables ξ_1 and ξ_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 $x_2, \dots, x_n$ while holding $x_2', \dots, x_n'$ "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 $a_i, b_i \in \mathbf{C}$ such that $\sum_{i=1}^n |a_i|^2 + \sum_{i=1}^n |b_i|^2 = 1$ and

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

where (x_i, x_i') are i.i.d copies of (ξ_1, ξ_2) from a given (μ, ρ)-family. Then there exist positive constants α, c_0, C_0 depending on (ξ_1, ξ_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^ϵ and n, there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r k_i g_i : k_i \in \mathbf{Z}, |k_i| \leq L_i\} \subset \mathbf{C}$ 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 β-close to Q.

□

As Theorem 3.2 can be shown by modifying the proof of Theorem 3.1 from [35], we postpone its proof until Appendix A. We now introduce several useful corollaries.

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 ξ satisfying $\mathbf{E}|\xi|^2 = 1$ and $\mathbf{E}\xi = \mathbf{E}[\mathrm{Im}(\xi)\mathrm{Re}(\xi)] = 0$. □

Secondly, if we β-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^ϵ and n, there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r k_i g_i : k_i \in \mathbf{Z}, |k_i| \leq L_i\} \subset \mathbf{C}$ 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 $P_1 := \frac{c_2}{c_1c_2' - c_1'c_2} \cdot Q + \frac{c_2'}{c_1c_2' - c_1'c_2} \cdot Q$;
- b_i are $O(\beta)$-close to the GAP $P_2 := \frac{c_1}{c_1c_2' - c_1'c_2} \cdot Q + \frac{c_1'}{c_1c_2' - c_1'c_2} \cdot Q$;
- consequently, a_i and b_i are $O(\beta)$-close to the combined GAP $P = \frac{c_2}{c_1c_2' - c_1'c_2} \cdot Q + \frac{c_2'}{c_1c_2' - c_1'c_2} \cdot Q + \frac{c_1}{c_1c_2' - c_1'c_2} \cdot Q + \frac{c_1'}{c_1c_2' - c_1'c_2} \cdot Q$.

□

Notice that the rank of P is $O_{B,\epsilon}(1)$ 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 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 ξ is a symmetric random variable and if $x_i' = -x_i$ (in which case $\rho = -1$), then the assumption $\mathbf{P}_{\mathbf{x},\mathbf{x}'}(|\sum_{i=1}^n a_i x_i + b_i x_i' - a| \leq \beta) \geq n^{-B}$ is equivalent to $\mathbf{P}_{\mathbf{x}}(|\sum_{i=1}^n (a_i - b_i)x_i - a| \leq \beta) \geq n^{-B}$. 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 $c_1^2 + c_1'^2 = c_2^2 + c_2'^2 = 1$ and $c_1c_2' \neq c_1'c_2$ such that $\xi_1 = c_1\psi_1 + c_1'\psi_2$ and $\xi_2 = c_2\psi_1 + c_2'\psi_2$, where ψ_2 is an independent copy of ψ_1. Then the assumption $\mathbf{P}_{\mathbf{x},\mathbf{x}'}(|\sum_{i=1}^n (a_i x_i + b_i x_i') - a| \leq \beta) \geq n^{-B}$ becomes $\mathbf{P}_{\psi_{ij}}(|\sum_{i=1}^n (c_1a_i + c_2b_i)\psi_{1i} + (c_1'a_i + c_2'b_i)\psi_{2i}) - a| \leq \beta) \geq n^{-B}$. So, as ψ_{ij} are independent, structural information for $c_1a_i + c_2b_i$ and $c_1'a_i + c_2'b_i$ can be deduced using Theorem 3.1, in the same way as 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 $\sum_{1 \leq i,j \leq n} a_{ij}(x_i + f_i)(x_j' + f_j')$ appearing in Theorem 2.3.

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

$$\sup_a \mathbf{P}_{\mathbf{x},\mathbf{x}'}\left(\left|\sum_{1 \leq i,j \leq n} a_{ij}(x_i + f_i)(x_j' + f_j') - a\right| \leq \beta\right) = \gamma \geq n^{-B},$$

where (x_i, x_i') are i.i.d copies of (ξ_1, ξ_2) from a given (μ, ρ)-family.

Then, there exist an integer $k \neq 0, |k| = n^{O_{B,\epsilon}(1)}$, a set of $r = O(1)$ rows $\mathbf{r}_{i_1}, \ldots, \mathbf{r}_{i_r}$ 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 exist integers $k_{ii_1}, \ldots, k_{ii_r}$, all bounded by $n^{O_{B,\epsilon}(1)}$, such that the following holds.

$$\mathbf{P}_{\mathbf{z}}\left(\left|\left\langle \mathbf{z}, k\mathbf{r}_i(A_n) + \sum_{j=1}^r k_{ii_j}\mathbf{r}_{i_j}(A_n)\right\rangle\right| \leq \beta n^{O_{B,\epsilon}(1)}\right) \geq n^{-O_{B,\epsilon}(1)}, \tag{9}$$

where $\mathbf{z} = (z_1, \ldots, z_n)$ and z_i are i.i.d. copies of $\eta^{(1/2)}(\xi_2 - \xi_2')$, where ξ_2' is an i.i.d. copy of ξ_2 and $\eta^{(1/2)}$ is a modified-Bernoulli random variable of parameter $1/2$ independent of ξ_2 and ξ_2'. □

We remark that this result is an analogue of [33, Theorem 1.8] in which case we studied the concentration of the quadratic forms of type $\sum_{ij} a_{ij}x_ix_j$. 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 [32]). However, we do not consider this matter here. Roughly speaking, in order to justify Theorem 1.9, we just need the conclusion of Theorem 4.1 for only one row.

To prove 4.1, we will follow the machinery from [33] with some extra twists. As the first step, we free the dependencies between $\mathbf{x}$ and $\mathbf{x}'$.

4.1 Decoupling lemma

Let U be an arbitrary subset of $\{1,\dots,n\}$ such that both of U and $\bar{U}$ are of size $\Theta(n)$. Let A_U be a matrix of size $n\times n$ 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 the Cauchy-Schwarz inequality.

Lemma 4.2. Assume that

$$\gamma=\sup_{a,b_i,b_i'} \mathbf{P}_{\mathbf{x},\mathbf{x}'}\left(\left|\sum_{i,j} a_{ij}x_ix_j'+\sum_i b_ix_i+\sum_i b_i'x_i'-a\right|\le\beta\right)\ge n^{-B},$$

where $\mathbf{x},\mathbf{x}'$ are defined as in Theorem 4.1. Then,

$$\mathbf{P}_{\mathbf{v},\mathbf{w}}\left(\left|\sum_{1\le i,j\le n} A_U(ij)v_iw_j\right|=O_B(\beta\sqrt{\log n})\right)=\Theta(\gamma^4), \tag{10}$$

where $\mathbf{v}=(v_1,\dots,v_n)$, $\mathbf{w}=(w_1,\dots,w_n)$, and (v_i,w_i) are i.i.d copies of a vector $(\xi_1-\xi_1',\xi_2-\xi_2')$, where (ξ_1',ξ_2') is an independent copy of (ξ_1,ξ_2). □

An advantage of considering the sum $\sum_{1\le i,j\le n} A_U(ij)v_iw_j$ over the original form $\sum_{ij} a_{ij}x_ix_j'$ is that we can rewrite the former as $\sum_{i\in U}(\sum_{j\in\bar{U}} a_{ij}w_j)v_i+\sum_{i\in U}(\sum_{j\in\bar{U}} a_{ji}v_j)w_i$. Thus, if all $v_j,w_j,j\in\bar{U}$ are held fixed, Theorem 3.2 applied to (10) allows us to extract useful information on $\sum_{j\in\bar{U}} a_{ij}w_j$ and $\sum_{j\in\bar{U}} a_{ji}v_j$. As the proof of Lemma 4.2 is standard, we postpone it until Appendix B.

We next apply Theorem 3.2 to obtain the following key structure for the entries of A_U.

Lemma 4.3. 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 $k_{ii_0}(U), i_0\in I_0(U)$, all bounded by $n^{O_{B,\epsilon}(1)}$, such that

$$\mathbf{P}_{\mathbf{y}}\left(\left|\left\langle k(U)\mathbf{r}_i(A_U)+\sum_{i_0\in I_0} k_{ii_0}(U)\mathbf{r}_{i_0}(A_U),\mathbf{y}\right\rangle\right|\le\beta n^{O_{B,\epsilon}(1)}\right)=n^{-O_{B,\epsilon}(1)}, \tag{11}$$

where $\mathbf{y}=(y_1,\dots,y_n)$ and y_i are i.i.d copies of $\xi_2-\xi_2'$. □

As the deduction of Theorem 4.1 from Lemma 4.3 is quite straightforward (by gathering the structural information from (11) for each U carefully), we refer the reader to [33, Section 4] for a complete treatment.

For the rest of this section we prove Lemma 4.3 using Lemma 4.2. First of all, as $A_U=A_{\bar{U}}$, it is enough to verify (11) for any index i from U. Also, it suffices to assume ξ 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 4.4. Assume as in the conclusion of Lemma 4.2 where (without loss of generality) $\beta\sqrt{\log n}$ and $\Theta(\gamma^4)$ are replaced by β and γ respectively. Then, the following holds with probability at least $3\gamma/4$ with respect to $\mathbf{v}_{\bar{U}}$ and $\mathbf{w}_{\bar{U}}$. There exist a proper symmetric GAP $P_{\mathbf{w}_{\bar{U}}}\subset\mathbf{C}$ of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$, and an index set $I_{\mathbf{w}_{\bar{U}}}\subset U$ of size $|U|-n^{\epsilon}$ such that $\langle\mathbf{r}_i(A_U),\mathbf{w}_{\bar{U}}\rangle$ is β-close to $P_{\mathbf{w}_{\bar{U}}}$ for all $i\in I_{\mathbf{w}_{\bar{U}}}$. □

Proof. (of Lemma 4.4) Write

$$\begin{aligned}\sum_{i\in U,j\in\bar{U}} a_{ij}v_iw_j+\sum_{i\in\bar{U},j\in U} a_{ij}v_iw_j&=\sum_{i\in U}(\sum_{j\in\bar{U}} a_{ij}w_j)v_i+(\sum_{j\in\bar{U}} a_{ji}v_j)w_i\\&=\sum_{i\in U}\langle\mathbf{r}_i(A_U),\mathbf{w}_{\bar{U}}\rangle v_i+\sum_{i\in U}\langle\mathbf{r}_i({A_U}^T),\mathbf{v}_{\bar{U}}\rangle w_i.\end{aligned}$$

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

$$\mathbf{P}_{\mathbf{v}_U,\mathbf{w}_U}\left(\left|\sum_{i\in U}\langle\mathbf{r}_i(A_U),\mathbf{w}_{\bar{U}}\rangle v_i+\sum_{i\in U}\langle\mathbf{r}_i({A_U}^T),\mathbf{v}_{\bar{U}}\rangle w_i-a\right|\le\beta\right)\ge\gamma/4.$$

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

Let G denote the collection of good pairs. We are going to estimate the probability p of a randomly chosen pair $(\mathbf{v}_{\bar{U}}, \mathbf{w}_{\bar{U}})$ being bad by an averaging method.

$$\mathbf{P}_{\mathbf{v}_{\bar{U}}, \mathbf{w}_{\bar{U}}, \mathbf{v}_U, \mathbf{w}_U} \left(\left| \sum_{i \in U} \langle \mathbf{r}_i(A_U), \mathbf{w}_{\bar{U}} \rangle v_i + \sum_{i \in U} \langle \mathbf{r}_i({A_U}^T), \mathbf{v}_{\bar{U}} \rangle w_i - a \right| \leq \beta \right) = \gamma$$

$$p\gamma/4 + 1 - p \geq \gamma$$

$$(1-\gamma)/(1-\gamma/4) \geq p.$$

Thus, the probability of a randomly chosen $(\mathbf{v}_{\bar{U}}, \mathbf{w}_{\bar{U}})$ belonging to G is at least

$$1 - p \geq (3\gamma/4)/(1-\gamma/4) \geq 3\gamma/4.$$

Consider a good vector $(\mathbf{v}_{\bar{U}}, \mathbf{w}_{\bar{U}}) \in G$. By definition, we have

$$\mathbf{P}_{\mathbf{v}_U, \mathbf{w}_U} \left(\left| \sum_{i \in U} \langle \mathbf{r}_i(A_U), \mathbf{w}_{\bar{U}} \rangle v_i + \sum_{i \in U} \langle \mathbf{r}_i({A_U}^T), \mathbf{v}_{\bar{U}} \rangle w_i - a \right| \leq \beta \right) \geq \gamma/4.$$

Next, if $\langle \mathbf{r}_i(A_U), \mathbf{w}_{\bar{U}} \rangle = \mathbf{0}$ for all i, then the conclusion of the lemma holds trivially for $P_{\mathbf{w}_{\bar{U}}} := \mathbf{0}$. Otherwise, we apply the last conclusion of Corollary 3.4 to the sequence $\{\langle \mathbf{r}_i(A_U), \mathbf{w}_{\bar{U}} \rangle, \langle \mathbf{r}_i({A_U}^T), \mathbf{v}_{\bar{U}} \rangle, i \in U\}$ (after a rescaling). As a consequence, we obtain an index set $I_{\mathbf{w}_{\bar{U}}} \subset U$ of size $|U| - n^{\epsilon}$ and a proper symmetric GAP $P_{\mathbf{w}_{\bar{U}}} \subset \mathbf{C}$ of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$, together with its elements $q_i(\mathbf{w}_{\bar{U}})$, such that $|\langle \mathbf{r}_i(A_U), \mathbf{w}_{\bar{U}} \rangle - q_i(\mathbf{w}_{\bar{U}})| \leq \beta$ for all $i \in I_{\mathbf{w}_{\bar{U}}}$. ■

4.2 Property of the $q_i(\mathbf{w}_{\bar{U}})$'s.

We now work with the GAP elements $q_i(\mathbf{w}_{\bar{U}})$, where $\mathbf{w}_{\bar{U}} \in G$. Because these points occupy a large 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 $P = \{k_1 g_1 + \cdots + k_r g_r | - K_i \leq k_i \leq K_i\}$ is a proper symmetric GAP, which contains a set $U = \{u_1, \ldots . u_n\}$. We consider P together with the map $\Phi : P \to \mathbf{R}^r$ which maps $k_1 g_1 + \cdots + k_r g_r$ 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 $\mathbf{R}^r$. 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 4.5. Assume that U is a subset of a proper symmetric GAP P of rank r, then there exists a proper symmetric GAP Q that contains U such that the followings hold.

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

□

We refer the reader to [33, Theorem 2.1] for a short proof of this lemma.

Common generating indices. By Lemma 4.5, we may assume that the $q_i(\mathbf{w}_{\bar{U}})$ span $P_{\mathbf{w}_{\bar{U}}}$. We choose s indices $i_{w_1}, \ldots, i_{w_s}$ from $I_{\mathbf{w}_{\bar{U}}}$ such that $q_{i_{y_j}}(\mathbf{w}_{\bar{U}})$ span $P_{\mathbf{w}_{\bar{U}}}$, where s is the rank of $P_{\mathbf{w}_{\bar{U}}}$. Note that $s = O_{B,\epsilon}(1)$ for all $\mathbf{w}_{\bar{U}} \in G$.

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

$$\mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G') \geq \mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G)/n^{O_{B,\epsilon}(1)} = \gamma/n^{O_{B,\epsilon}(1)}. \tag{12}$$

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

$$q_i(\mathbf{w}_{\bar{U}}) = c_{i1}(\mathbf{w}_{\bar{U}}) g_1(\mathbf{w}_{\bar{U}}) + \cdots + c_{ir}(\mathbf{w}_{\bar{U}}) g_r(\mathbf{w}_{\bar{U}}),$$

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

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

Consider the collection of the coefficient-tuples $\Big(\big(c_{11}(\mathbf{w}_{\bar{U}}),\dots,c_{1r}(\mathbf{w}_{\bar{U}})\big);\dots;\big(c_{r1}(\mathbf{w}_{\bar{U}}),\dots c_{rr}(\mathbf{w}_{\bar{U}})\big)\Big)$ for all $\mathbf{w}_{\bar{U}} \in G'$. The number of possibilities these tuples can take is at most

$$(n^{O_{B,\epsilon}(1)})^{r^2} = n^{O_{B,\epsilon}(1)}.$$

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

$$\Big(\big(c_{11}(\mathbf{w}_{\bar{U}}),\dots,c_{1r}(\mathbf{w}_{\bar{U}})\big);\dots;\big(c_{r1}(\mathbf{w}_{\bar{U}}),\dots c_{rr}(\mathbf{w}_{\bar{U}})\big)\Big) = \Big((c_{11},\dots,c_{1r}),\dots,(c_{r1},\dots c_{rr})\Big)$$

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

$$\mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G'') \geq \mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G')/n^{O_{B,\epsilon}(1)} \geq \gamma/n^{O_{B,\epsilon}(1)}. \tag{13}$$

In summary, there exist r tuples $(c_{11},\dots,c_{1r}),\dots,(c_{r1},\dots c_{rr})$ (where we recall that $r = O(1)$ is the rank of the GAP) whose components are integers bounded by $n^{O_{B,\epsilon}(1)}$, such that the following holds for all $\mathbf{w}_{\bar{U}} \in G''$.

- $q_i(\mathbf{w}_{\bar{U}}) = c_{i1}g_1(\mathbf{w}_{\bar{U}}) + \dots + c_{ir}g_r(\mathbf{w}_{\bar{U}})$, for $i = 1,\dots,r$.
- The vectors $(c_{11},\dots,c_{1r}),\dots,(c_{r1},\dots c_{rr})$ span $\mathbf{R}^{\mathrm{rank}(P_{\mathbf{w}_{\bar{U}}})}$.

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

$$\mathbf{P}_{\mathbf{w}_{\bar{U}}}(i \in I_{\mathbf{w}_{\bar{U}}}, \mathbf{w}_{\bar{U}} \in G'') \geq \mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G'')/2. \tag{14}$$

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

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

$$q(\mathbf{w}_{\bar{U}}) = c_1(\mathbf{w}_{\bar{U}})g_1(\mathbf{w}_{\bar{U}}) + \dots + c_r(\mathbf{w}_{\bar{U}})g_r(\mathbf{w}_{\bar{U}})$$

where $c_i(\mathbf{w}_{\bar{U}})$ are integers bounded by $n^{O_{B,\epsilon}(1)}$.

For short, for each i we denote by $\mathbf{v}_i$ the vector $(c_{i1},\dots,c_{ir})$, we will also denote by $\mathbf{v}_{\mathbf{r},\mathbf{w}_{\bar{U}}}$ the vector $(c_1(\mathbf{w}_{\bar{U}}),\dots c_r(\mathbf{w}_{\bar{U}}))$.

Because $P_{\mathbf{w}_{\bar{U}}}$ is spanned by $q_1(\mathbf{w}_{\bar{U}}),\dots,q_r(\mathbf{w}_{\bar{U}})$, we have $k = \det(\mathbf{v}_1,\dots\mathbf{v}_r) \neq 0$, and by basic linear algebra

$$kq(\mathbf{w}_{\bar{U}}) + \det(\mathbf{v}_{\mathbf{r},\mathbf{w}_{\bar{U}}},\mathbf{v}_2,\dots,\mathbf{v}_r)q_1(\mathbf{w}_{\bar{U}}) + \dots + \det(\mathbf{v}_{\mathbf{r},\mathbf{w}_{\bar{U}}},\mathbf{v}_1,\dots,\mathbf{v}_{r-1})q_r(\mathbf{w}_{\bar{U}}) = 0. \tag{15}$$

It is crucial to note that k is independent of the choice of $\mathbf{r}$ and $\mathbf{w}_{\bar{U}}$.

Next, because each coefficient of (15) is bounded by $n^{O_{B,\epsilon}(1)}$, there exists a subset $G''_{\mathbf{r}}$ of G'' such that all $\mathbf{w}_{\bar{U}} \in G''_{\mathbf{r}}$ correspond to the same identity, and by (14)

$$\mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G''_{\mathbf{r}}) \geq (\mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G'')/2)/(n^{O_{B,\epsilon}(1)})^r = \gamma/n^{O_{B,\epsilon}(1)} = n^{-O_{B,\epsilon}(1)}. \tag{16}$$

In other words, there exist integers $k_1,\dots,k_r$ depending on $\mathbf{r}$, all bounded by $n^{O_{B,\epsilon}(1)}$, such that

$$kq(\mathbf{w}_{\bar{U}}) + k_1q_1(\mathbf{w}_{\bar{U}}) + \dots + k_rq_r(\mathbf{w}_{\bar{U}}) = 0 \tag{17}$$

for all $\mathbf{w}_{\bar{U}} \in G''_{\mathbf{r}}$.

4.3 Passing back to A_U.

Because $q_i(\mathbf{w}_{\bar{U}})$ are β-close to $\langle \mathbf{r}_i, \mathbf{w}_{\bar{U}} \rangle$, it follows from (17) that

$$\Big|\langle k\mathbf{r}, \mathbf{w}_{\bar{U}}\rangle + \langle k_1\mathbf{r}_1, \mathbf{w}_{\bar{U}}\rangle + \dots + \langle k_r\mathbf{r}_r, \mathbf{w}_{\bar{U}}\rangle\Big| = \Big|\langle k\mathbf{r} + k_1\mathbf{r}_1 + \dots + \mathbf{r}_r, \mathbf{w}_{\bar{U}}\rangle\Big| \leq n^{O_{B,\epsilon}(1)}\beta.$$

Furthermore, as $\mathbf{P}_{\mathbf{w}_{\bar{U}}}(\mathbf{w}_{\bar{U}} \in G''_{\mathbf{r}}) = n^{-O_{B,\epsilon}(1)}$, we have

$$\mathbf{P}_{\mathbf{w}_{\bar{U}}}\Big(|\langle k\mathbf{r} + k_1\mathbf{r}_1 + \dots + k_r\mathbf{r}_r, \mathbf{w}_{\bar{U}}\rangle| \leq n^{O_{B,\epsilon}(1)}\beta\Big) = n^{-O_{B,\epsilon}(1)}. \tag{18}$$

As (18) holds for any row $\mathbf{r}$ indexing from I, this completes the proof of Lemma 4.3.

5 Random matrix: the inverse step

We now give a proof of Theorem 2.3. We first apply Theorem 4.1 to a_{ij} to obtain

$$\mathbf{P}_{\mathbf{z}}\left(\left|\left\langle \mathbf{z}, k\mathbf{r}_i(A_{n-1}) + \sum_j k_{ii_j}\mathbf{r}_{i_j}(A_{n-1})\right\rangle\right| \le n^{-A+O_{B,\epsilon}(1)}\right) \ge n^{-O_{B,\epsilon}(1)},$$

where $A_{n-1} = (a_{ij})_{2\le i,j\le n}$.

For short, we denote by $\mathbf{r}'_i$ the vector $k\mathbf{r}_i(A_{n-1}) + \sum_j k_{ii_j}\mathbf{r}_{i_j}(A_{n-1})$. Thus, for any $i \in I$,

$$\mathbf{P}_{\mathbf{z}}\left(|\langle \mathbf{z}, \mathbf{r}'_i\rangle| \le n^{-A+O_{B,\epsilon}(1)}\right) \ge n^{-O_{B,\epsilon}(1)}. \tag{19}$$

Set

$$K = n^{-A/2}.$$

We consider two cases.

Case 1.(*non-degenerate case*). There exists $i_0 \in I$ such that $\|\mathbf{r}'_{i_0}\|_2 \ge K$. Because $\mathbf{r}'_{i_0} = k\mathbf{r}_{i_0}(A_{n-1}) + \sum_{j\in I_0} k_{i_0 j}\mathbf{r}_j(A_{n-1}y)$, $\mathbf{r}'_{i_0}$ is orthogonal to $n - |I| - 1 = n - O_{B,\epsilon}(1)$ column vectors of M_{n-1}.

Set

$$\mathbf{v} := \mathbf{r}'_{i_0}/\|\mathbf{r}'_{i_0}\|_2.$$

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

Also, it follows from (19) that

$$\mathbf{P}_{\mathbf{z}}\left(|\langle \mathbf{z}, \mathbf{v}\rangle| \le n^{-A/2+O_{B,\epsilon}(1)}\right) \ge n^{-O_{B,\epsilon}(1)}. \tag{20}$$

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

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

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

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

$$\left\|k\mathbf{r}_i(A_{n-1}) + \sum_{j\in I_0} k_{ij}\mathbf{r}_j(A_{n-1})\right\|_2 = \|\mathbf{r}'_i\|_2 \le K. \tag{21}$$

Without loss of generality we can assume that I and I_0 are disjoint. Next, because $\sum_j \|\mathbf{c}_j(A_{n-1})\|_2^2 = 1$, there exists an index j_0 such that $\|\mathbf{c}_{j_0}(A_{n-1})\|_2 \ge n^{-1/2}$. Consider this column vector.

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

$$\left|k\mathbf{c}_{j_0}(i) + \sum_{j\in I_0} k_{ij}\mathbf{c}_{j_0}(j)\right| \le K.$$

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

For each $j \notin I$, we approximate $\mathbf{c}_{j_0}(j)$ by a number v_j of the form $(1/\lfloor 2K^{-1}\rfloor)\cdot\mathbf{Z}^2$ such that $|v_j - \mathbf{c}_{j_0}(j)| \le K$. 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 $\mathbf{v} = (v_1, \dots, v_{n-1})$, we have

$$\|\mathbf{v} - \mathbf{c}_{j_0}(A_{n-1})\|_2 \le Kn^{O_{B,\epsilon}(1)}.$$

Furthermore, by Condition 2.1, and because $\langle \mathbf{c}_{j_0}(A_{n-1}), \mathbf{r}_i(M_{n-1})\rangle = 0$ for $i \ne j_0$, we infer that

$$|\langle \mathbf{v}, \mathbf{r}_i(M_{n-1})\rangle| \le Kn^{O_{B,\epsilon}(1)}.$$

Note that $\|\mathbf{v}\|_2 \gg n^{-1/2}$. Set $\mathbf{u} := \lfloor 1/\|\mathbf{v}\|_2\rfloor \cdot \mathbf{v}$, we then obtain

- $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le n^{-A/2+O_{B,\epsilon}(1)}$ 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 $|p|,|q|,|p'|,|q'| \le n^{A/2+O_{B,\epsilon}(1)}$.

6 Random matrix: the counting step

We now give a proof of Theorem 2.4. Without loss of generality, we assume ϵ to be sufficiently small. 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 [34, Section 5].

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

Let $\mathcal{N}$ be the number of such structural vectors $\mathbf{u}$. Because each GAP is determined by its generators and dimensions, the number of Q's is bounded by

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

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 $\{\frac{p}{q}+i\frac{p'}{q'}, |p|,|q|,|p'|,|q'| \le n^{A/2+O_{B,\epsilon}(1)}\}$, 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

$$\mathcal{N} \le n^{O_{A,B,\epsilon}(1)}n^{O_{B,\epsilon}(n)}n^{O_{A,B,\epsilon}(n^\epsilon)} = n^{O_{B,\epsilon}(n)}. \tag{22}$$

Set $\beta_0 := n^{-A/2+O_{B,\epsilon}(1)}$, the bound obtained from the conclusion of Theorem 2.3. For a given vector $\mathbf{u}$, we define $\mathbf{P}_{\beta_0}(\mathbf{u})$ as follows

$$\mathbf{P}_{\beta_0}(\mathbf{u}) := \mathbf{P}\Big(|\langle \mathbf{u}, \mathbf{r}_i(M_n)\rangle| \le \beta_0 \text{ for } n-O_{B,\epsilon}(1) \text{ rows of } M_{n-1}\Big).$$

For the sake of discussion, let us pretend for now that the rows of X_n are independent. By definition, the vector $\mathbf{u}$ is orthogonal to almost every row of M_n. Thus, if $\mathbf{u}$ is fixed, the probability of this event is bounded by

$$\mathbf{P}_{\beta_0}(\mathbf{u}) \le (\mathbf{P}_{\mathbf{x}}(|u_1x_1+\cdots+u_nx_n| \le \beta_0))^{n-O(1)} := \gamma^{n-O(1)},$$

where $x_1,\dots,x_n$ are i.i.d. copies of ξ.

Now, if γ is small, say $n^{-\Omega(1)}$, 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 $\sum_{\mathbf{u}}\mathbf{P}_{\beta_0}(\mathbf{u})$ is negligible, taking into account of the bound $n^{O(n)}$ of $\mathcal{N}$.

Next, if γ 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 $|\langle \mathbf{u}', \mathbf{r}_i(M_n)\rangle|$ 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 $(\gamma^{-1}/n^\epsilon)^n$, while we recall that $\mathbf{P}_{\beta_0}(\mathbf{u}')$ is of order γ^{-n}. Thus, summing over $\mathbf{u}'$ we obtain the desired bound

$$\sum_{\mathbf{u}'}\mathbf{P}_{\beta_0}(\mathbf{u}') \le \#\{\text{ new GAPs }\}(\gamma^{-1}/n^\epsilon)^n\gamma^{-n} = O(n^{-\epsilon n+O(1)}).$$

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

(i) $\mathcal{B}'$ contains those $\mathbf{u}$ for which $\mathbf{P}_{\beta_0}(\mathbf{u})$ is very small, 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 pass 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 $O(n^{-\epsilon n})$.

What makes our analysis harder is that the estimate $\mathbf{P}_{\beta_0}(\mathbf{u}) \le (\mathbf{P}_{\mathbf{x}}(|u_1x_1+\cdots+u_nx_n| \le \beta_0))^{n-O(1)}$ 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 $n^{O_{B,\epsilon}(1)}$ in probability, we may assume that $|\langle \mathbf{u}, \mathbf{r}_i(M_n)\rangle| \le \beta_0$ for the first $n - O_{B,\epsilon}(1)$ rows of M_n. Also, by paying another factor of n^{n^ϵ} in probability, we may assume that the first n_0 components u_i of $\mathbf{u}$ [†] belong to a GAP Q, and $u_{n_0} \ge 1/2\sqrt{n-1}$ (recall that $\mathbf{u} \asymp 1$), where

$$n_0 := n - 2n^\epsilon.$$

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 $i \le n_0$, we define

$$\gamma_\beta^{(i)}(\mathbf{u}) := \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}(|x_i u_i + \dots + x_{n_0} u_{n_0} - a| \le \beta),$$

where $x_i, \dots, x_{n_0}$ are i.i.d copies of ξ.

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 $\mathbf{P}_\beta(\mathbf{u})$ that $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le \beta$ for all $i \le n - O_{B,\epsilon}(1)$ can be bounded by

$$\begin{aligned}
\mathbf{P}_\beta(\mathbf{u}) &\le \prod_{1\le i\le n-O_{B,\epsilon}(1)} \sup_a \mathbf{P}_{x_i,\dots,x_{n-1}}(|x_i u_i + \dots + x_{n-1} u_{n-1} - a| \le \beta) \\
&\le \prod_{1\le i\le n_0} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}(|x_i u_i + \dots + x_{n_0} u_{n_0} - a| \le \beta) \\
&= \prod_{1\le i\le n_0} \gamma_\beta^{(i)}(\mathbf{u}). \qquad (23)
\end{aligned}$$

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

$$\begin{aligned}
\gamma_\beta^{(k)}(\mathbf{u}) &\le \sup_a \mathbf{P}_{x_{n_0}}(|x_{n_0} u_{n_0} - a| \le \beta) \\
&\le 1 - c_2. \qquad (24)
\end{aligned}$$

Thus,

$$\mathbf{P}_\beta(\mathbf{u}) \le (1-c_2)^{n_0} = (1-c_2)^{(1-o(1))n}.$$

6.2 Classification

Next, let C be a sufficiently large constant depending on B and ϵ but not A. We classify $\mathbf{u}$ into two classes $\mathcal{B}$ and $\mathcal{B}'$, depending on whether $\mathbf{P}_{\beta_0}(\mathbf{u}) \ge n^{-Cn}$ or not.

Because of (22) and with C sufficiently large,

$$\sum_{\mathbf{u}\in\mathcal{B}'} \mathbf{P}_{\beta_0}(\mathbf{u}) \le n^{O_{B,\epsilon}(n)}/n^{Cn} \le n^{-n/2}. \qquad (25)$$

For the rest of this section, we focus on $\mathbf{u} \in \mathcal{B}$.

6.3 Approximation for vectors of "low complexity"

Let $\mathcal{B}_1$ be the collection of $\mathbf{u} \in \mathcal{B}$ satisfying the following property: for any n' components $u_{i_1}, \dots, u_{i_{n'}}$ among the $u_1, \dots, u_{n_0}$, we have

$$\sup_a \mathbf{P}_{x_{i_1},\dots,x_{i_{n'}}}\left(\left|u_{i_1} x_{i_1} + \dots + u_{i_{n'}} x_{i_{n'}} - a\right| \le n^{-B-4}\right) \ge (n')^{-1/2+o(1)}. \qquad (26)$$

Here we set

$$n' := n^{1-\epsilon}.$$

[†]Roughly speaking, in later analysis we will be fixing the columns of M_n that correspond to the unstructured components of $\mathbf{u}$. Thus the assumption that the first n_0 components of $\mathbf{u}$ come from a GAP is slightly more difficult than other cases as it involves more dependencies and less structural components. However, there is no major difference among the treatments.

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

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

- $|u_i - u_i'| \le 4\beta$ for all i.
- u_i' takes the same value u for at least $n_0 - n'$ indices i.

In other words, $\mathbf{u}$ can be approximated by a vector of "low complexity". Because of the approximation and Condition 2.1, whenever $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le \beta_0$, we have

$$|\langle \mathbf{u}', \mathbf{r}_i(M_{n-1})\rangle| \le n(n^{B+1}+n^\alpha)(4\beta) + \beta_0 := \beta'.$$

It is clear from the bound on β and β_0 that $\beta' \le c_1/2\sqrt{n-1}$, and thus by (24),

$$\mathbf{P}_{\beta'}(\mathbf{u}') \le (1-c_2)^{(1-o(1))n}.$$

Now we bound the number of $\mathbf{u}'$ 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 $(2\beta)^{-1}\cdot(\mathbf{Z}+\sqrt{-1}\mathbf{Z})$, 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

$$\begin{aligned}\mathbf{P}(\exists \mathbf{u}\in\mathcal{B}_1 \text{ such that for all } i \le n - O_{B,\epsilon}(1), \langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle \le \beta_0) &\le \sum_{\mathbf{u}'} \mathbf{P}_{\beta'}(\mathbf{u}') \\ &\le O(n^{2n^{1-\epsilon}})O(n^{O_{A,B,\epsilon}(n^{1-\epsilon})})(1-c_2)^{(1-o(1))n} \\ &\le (1-c_2)^{(1-o(1))n}.\end{aligned}$$

6.4 Approximation for vectors of "high complexity"

Assume that $\mathbf{u} \in \mathcal{B}_2 := \mathcal{B}\backslash\mathcal{B}_1$. 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 $u_{n_0-n'+1}, \dots, u_{n_0}$ satisfy the property

$$\begin{aligned}\sup_a \mathbf{P}_{x_{n_0-n'+1},\dots,x_{n_0}}\left(|u_{n_0-n'+1}x_{n_0-n'+1} + \dots + u_{n_0}x_{n_0} - a| \le n^{-B-4}\right) &\le (n')^{-1/2+o(1)} \\ &\le n^{-1/2+\epsilon/2+o(1)}. \end{aligned} \tag{27}$$

Preparation. Next, define a radius sequence $\beta_k, k \ge 0$ where $\beta_0 = n^{-A/2+O_{B,\epsilon}(1)}$ is the bound obtained from the conclusion of Theorem 2.3, and

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

Recall from (23) that

$$\mathbf{P}_{\beta_k}(\mathbf{u}) \le \prod_{1\le i\le n_0-n'} \gamma_{\beta_k}^{(i)}(\mathbf{u}) =: \pi_{\beta_k}(\mathbf{u}).$$

Roughly speaking, the reason we truncated the product here is that whenever $i \le n_0 - n'$ and β_k is small enough, the terms $\gamma_{\beta_k}^{(i)}(\mathbf{u})$ 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 $|\langle \mathbf{u}, \mathbf{r}_i(M_n)\rangle| \le \beta_k$ and if $\mathbf{u}'$ is an approximation of $\mathbf{u}$ such that $|u_i - u_i'| \le \beta_k$ for all i, then

$$\begin{aligned}\pi_{\beta_k}(\mathbf{u}) &= \prod_{1\le i\le n_0-n'} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}\Big(|u_i x_i + \dots + u_{n_0}x_{n_0} - a| \le \beta_k\Big) \\ &\le \prod_{1\le i\le n_0-n'} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}\Big(|u_i' x_i + \dots + u_{n_0}'x_{n_0} - a| \le (n(n^{B+1})+n^\alpha)\beta_k + \beta_k\Big) \\ &= \prod_{1\le i\le n_0-n'} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}\Big(|u_i' x_i + \dots + u_{n_0}'x_{n_0} - a| \le (n^{B+2}+n^{\alpha+1}+1)\beta_k\Big) \\ &\le \prod_{1\le i\le n_0-n'} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}\Big(|u_i x_i + \dots + u_{n_0}x_{n_0} - a| \le (n^{B+2}+n^{\alpha+1}+1)^2\beta_k\Big) \\ &= \pi_{\beta_{k+1}}(\mathbf{u}). \end{aligned} \tag{28}$$

Naturally, we hope that after the approximation $\mathbf{P}_{(n^{B+2}+n^{\alpha+1}1)\beta_k}(\mathbf{u}')$ does not increase much compared to the original $\mathbf{P}_{\beta_k}(\mathbf{u})$, where we recall that n^α is the upper bound for the entries of F_n. That motivates us to consider a special radius β_{k_0} with respect to $\mathbf{u}$ defined below.

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

$$\pi_{\beta_{k_0+1}}(\mathbf{u}) \leq n^{\epsilon n}\pi_{\beta_{k_0}}(\mathbf{u}). \tag{29}$$

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 $\gamma_{\beta_i}^{(i)}(\mathbf{u})$, we now turn to its lower bound. Because of Condition 2.1, and $u_i \leq 1$ for all i, and by the pigeonhole principle, the following trivial bound holds for any $\beta \geq \beta_0$ and $i \leq n_0 - n'$,

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

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

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

We next divide it into $K = (A/2 + O_{B,\epsilon}(1))\epsilon^{-1}$ sub-intervals $I_k = [l_I n^{k\epsilon}, l_I n^{(k+1)\epsilon}]$. 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 $\mathbf{u}$ based on the distribution of the $\gamma_{\beta_{k_0}}^{(i)}(\mathbf{u}), 1 \leq i \leq n_0 - n^{1-\epsilon}$.

For each $0 \leq k_0 \leq C\epsilon^{-1}$ and each tuple $(m_0, \dots, m_K)$ satisfying $m_0 + \cdots + m_K = n_0 - n'$, we let $\mathcal{B}_{k_0}^{(m_0,\dots,m_K)}$ denote the collection of those $\mathbf{u}$ from $\mathcal{B}_2$ that satisfy the following conditions.

- $k_0(\mathbf{u}) = k_0$.
- There are exactly m_k terms of the sequence $(\gamma_{\beta_{k_0}}^{(i)}(\mathbf{u}))$ 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 $\gamma_{\beta_{k_0}}^{(i)}(\mathbf{u}) \in I_k$.

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

- *First step.* Consider each index i in the range $1 \leq i \leq m_0$. Because $\gamma_{\beta_{k_0}}^{(1)} \in I_0$, 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 $O(l_0^{-1}/\sqrt{n'}) = O(l_0^{-1}/n^{1/2-\epsilon})$ for all but $n^{1-2\epsilon}$ indices i. Furthermore, all u_i' have the form $\beta_{k_0} \cdot (\frac{p}{q} + \sqrt{-1}\frac{p'}{q'})$, where $|p|, |q|, |p'|, |q'| = O(n\beta_{k_0}^{-1}) = O(n^{A/2+O_{B,\epsilon}(1)})$.
- *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 $\gamma_{\beta_{k_0}}^{(n_0+\cdots+n_{k-1}+1)} \in I_k$, 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 $O(l_k^{-1}/n^{1/2-\epsilon})$ for all but $n^{1-2\epsilon}$ indices i. Furthermore, all u_i' have the form $\beta_{k_0} \cdot (p/q + \sqrt{-1}p'/q')$, where $|p|, |q|, |p'|, |q'| = O(n\beta_{k_0}^{-1}) = O(n^{A/2+O_{B,\epsilon}(1)})$.
- For the remaining components u_i, we just simply approximate them by the closest point in $\beta_{i_0} \cdot (\mathbf{Z} + \sqrt{-1}\mathbf{Z})$.

We have thus provided an approximation of $\mathbf{u}$ by $\mathbf{u}'$ satisfying the following properties.

(i) $|u_i - u_i'| \leq \beta_{k_0}$ 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 $\beta_{k_0} \cdot \{p/q + \sqrt{-1}p'/q', |p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\epsilon}(1)}\}$.

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

Property of u′. Let $\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$ be the collection of all $\mathbf{u}'$ obtained from $\mathbf{u} \in \mathcal{B}_{k_0}^{(m_1,\dots,m_K)}$ as above. Observe that, as $|\langle \mathbf{u}, \mathbf{r}_i(M_n)\rangle| \leq \beta_{k_0}$ for all $i \leq n - O_{B,\epsilon}(1)$, we have

$$|\langle \mathbf{u}', \mathbf{r}_i(M_n)\rangle| \leq (n^{B+2} + n^{\alpha+1} + 1)\beta_{k_0}. \tag{30}$$

Hence, in order to justify Theorem 2.4 in the case $\mathbf{u} \in \mathcal{B}_2$, it suffices to show that the probability that (30) holds for all $i \leq n - O_{B,\epsilon}(1)$, for some $\mathbf{u}' \in \mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$, is small.

Consider a $\mathbf{u}' \in \mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$ and the probability $\mathbf{P}_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}(\mathbf{u}')$ that (30) holds for all $i \leq n - O_{B,\epsilon}(1)$. By the discussion about (28), we have

$$\mathbf{P}_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}(\mathbf{u}') \leq \pi_{\beta_{k_0+1}}(\mathbf{u}) \leq n^{\epsilon n}\pi_{\beta_{k_0}}(\mathbf{u}),$$

where in the second inequality we used (29).

We recall from the definition of $\mathcal{B}^{(m_1,\dots,m_K)}_{k_0}$ that

$$\begin{aligned}\pi_{\beta_{k_0}}(\mathbf{u}) &\leq \prod_{k=1}^{K} l_{k+1}^{m_k} = n^{\epsilon(m_1+\dots+m_k)}\prod_{k=1}^{K} l_k^{m_k} \\ &\leq n^{\epsilon n}\prod_{k=1}^{K} l_k^{m_k}.\end{aligned}$$

Hence,

$$\mathbf{P}_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}(\mathbf{u}') \leq n^{2\epsilon n}\prod_{k=1}^{K} l_k^{m_k}. \tag{31}$$

The size of $\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$. In the next step of the argument, we bound the size of $\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$. 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 $Q_1,\dots,Q_K$ is bounded by

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

Next, after locating Q_k, the number $\mathcal{N}_1$ of ways to choose u_i' from each Q_k is

$$\begin{aligned}\mathcal{N}_1 &\leq \prod_{k=1}^{K}\binom{m_k}{n^{1-2\epsilon}}|Q_k|^{m_k-n^{1-2\epsilon}} \\ &\leq 2^{m_1+\dots+m_K}\prod_{k=1}^{K}|Q_k|^{m_k} \\ &\leq (O(1))^n\prod_{k=1}^{K} l_k^{-m_k}/n^{(1/2-\epsilon)(m_1+\dots+m_k)} \\ &\leq \prod_{k=1}^{K} l_k^{-m_k}/n^{(1/2-\epsilon-o(1))n},\end{aligned}$$

where we used the bound $|Q_k| = O(l_k^{-1}/n^{1/2-\epsilon})$ 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 $\mathcal{N}_2$ of ways to choose them is bounded by

$$\mathcal{N}_2 \leq (n^{A+O_{B,\epsilon}(1)})^{2n^\epsilon+Kn^{1-2\epsilon}} = n^{O_{A,B,\epsilon}(n^{1-2\epsilon})}.$$

Putting the bound for $\mathcal{N}_1$ and N_2 together, we obtain a bound $\mathcal{N}'$ for $|\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}|$,

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

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

$$\begin{aligned}\sum_{\mathbf{u}'\in\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}} \mathbf{P}_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}(\mathbf{u}') &\leq n^{2\epsilon n}\prod_{k=1}^{K} l_k^{m_k}\prod_{k=1}^{K} l_k^{-m_k}/n^{(1/2-\epsilon-o(1))n} \\ &\leq n^{-(1/2-3\epsilon-o(1))n}.\end{aligned}$$

Summing over the choices of k_0 and $(m_1, \dots, m_K)$ we obtain the bound

$$\sum_{k_0, m_1, \dots, m_K} \sum_{\mathbf{u}' \in \mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}} \mathbf{P}_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}(\mathbf{u}') \leq n^{-(1/2-3\epsilon-o(1))n},$$

completing the treatment for incompressible vectors, and hence the proof of Theorem 2.4.

7 Proof of the elliptic law, Theorems 1.5 and 1.8

This section is devoted to the proof of Theorems 1.5 and 1.8. We introduce the following notation. Given a $n \times n$ matrix A_n, we let μ_{A_n} denote the empirical measure built from the eigenvalues of A_n and ν_{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 $\lambda_1(A_n), \dots, \lambda_n(A_n) \in \mathbb{C}$ 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 1.8, we will show that, with probability one,

$$\mu_{\frac{1}{\sqrt{n}}(X_n + F_n)} \longrightarrow \mu_\rho \tag{33}$$

as $n \to \infty$, where μ_ρ is the uniform probability measure on the ellipsoid $\mathcal{E}_\rho$. 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 ρ.

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

$$U_\mu(z) := -\int_{\mathbb{C}} \log|z - \lambda| d\mu(\lambda).$$

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 μ_{A_n} and the family of measures $\{\nu_{A_n - zI}\}_{z \in \mathbb{C}}$. In particular,

$$U_{\mu_{A_n}}(z) = -\frac{1}{2n} \log\det(A_n - zI)^*(A_n - zI) = -\int_0^\infty \log(s) d\nu_{A_n - zI}(s).$$

The work of Goldsheid and Khoruzhenko [20] is one of the first rigorous uses of the logarithmic potential to study random matrices. We also refer the reader to the survey [4] for more details. A key tool in the proof of Theorems 1.5 and 1.8 is the following result from [4].

Lemma 7.1 (Hermitization lemma, [4])**.** Let $\{A_n\}_{n \geq 1}$ be a sequence of complex random matrices where A_n is of size $n \times n$ for every $n \geq 1$. Suppose that there exists a family of (non-random) probability measures $\{\nu_z\}_{z \in \mathbb{C}}$ such that for a.a. $z \in \mathbb{C}$, a.s.

(i) $\nu_{A_n - zI} \to \nu_z$ as $n \to \infty$

(ii) log is uniformly integrable for $\{\nu_{A_n - zI}\}_{n \geq 1}$.

Then there exists a probability measure $\mu \in \mathcal{P}(\mathbb{C})$ such that

(i) a.s. $\mu_{A_n} \to \mu$ as $n \to \infty$

(ii) for a.a. $z \in \mathbb{C}$,

$$U_\mu(z) = -\int_0^\infty \log(s) d\nu_z(s).$$

□

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 μ_ρ. Fix $-1 < \rho < 1$. We say the family of measures $\{\nu_z\}_{z \in \mathbb{C}}$ determine the elliptic law with parameter ρ 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 [31].

The key tool we use to prove Theorems 1.5 and 1.8 is the following comparison lemma.

Lemma 7.3. Let $0 \leq \mu \leq 1$ and $-1 < \rho < 1$ be given. Let $\{X_n\}_{n \geq 1}$ and $\{Y_n\}_{n \geq 1}$ be sequences of random matrices that satisfy condition **C0** with atom variables (ξ_1, ξ_2) and (η_1, η_2), respectively. Assume (ξ_1, ξ_2) and (η_1, η_2) are from the (μ, ρ)-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 $\sup_n \frac{1}{n^2} \|F_n\|_2^2 < \infty$. Then a.s.

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

as $n \to \infty$. □

Lemma 7.3 is useful when we know the limit of $\mu_{\frac{1}{\sqrt{n}} Y_n}$. 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 [31].

We divide the proof of Theorems 1.5 and 1.8 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 $\{\nu_{\frac{1}{\sqrt{n}}(X_n + F_n) - zI}\}_{n \geq 1}$. We prove this statement in sub-section 7.1. The arguments in this section are based on [4, 31, 46]. We will also require the use of Theorem 1.9 to control the least singular value.

2. In sub-section 7.4 we prove a replacement lemma using a moment matching argument. The lemma will allow us to compute the limit of $\nu_{\frac{1}{\sqrt{n}} X_n - zI}$ 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 7.2) and apply a truncation argument (sub-section 7.3).

3. In sub-section 7.5, we prove Lemma 7.3. We then apply the results of [31] and Lemma 7.3 to prove Theorem 1.5.

4. In sub-section 7.6, we prove Theorem 1.8.

7.1 Uniform Integrability

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

Lemma 7.4. 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 (ξ_1, ξ_2) from the (μ, ρ)-family. Assume $\{F_n\}_{n\geq 1}$ is a sequence of deterministic matrices such that $\text{rank}(F_n) = o(n)$ and $\sup_n \frac{1}{n^2}\|F_n\|_2^2 < \infty$. Then for a.a. $z \in \mathbb{C}$ a.s. log is uniformly integrable for $\{\nu_{\frac{1}{\sqrt{n}}(X_n+F_n)-zI}\}_{n\geq 1}$. □

The proof of Lemma 7.4 is based on the arguments of [4, 31, 46]. In order to prove Lemma 7.4, we will need the following bound for small singular values.

Lemma 7.5. 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 $m = \lceil n - i/2 \rceil$ 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 [46, Lemma A.4],

$$\sigma_1'^{-2} + \cdots + \sigma_m'^{-2} = \text{dist}_1^{-2} + \cdots + \text{dist}_m^{-2}$$

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

$$\sigma_{n-i}^{-2} \leq n\sigma_{n-i}'^{-2}$$

it follows that

$$\frac{i}{2n}\sigma_{n-i}^{-2} \leq \frac{i}{2}\sigma_{n-i}'^{-2} \leq \sum_{j=n-i}^{m} \sigma_{n-j}'^{-2} \leq \sum_{j=1}^{m} \text{dist}_j^{-2}. \tag{34}$$

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

$$H'_j = \text{Span}\{\mathbf{r}_k(A'_j) : k = 1, \ldots, m; k \neq j\}.$$

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 7.6 below and the union bound, we obtain

$$\sum_{n=1}^{\infty} \mathbf{P}\left(\bigcup_{i=2n^{1-\gamma}}^{n} \bigcup_{j=1}^{m} \left\{\text{dist}_j \leq c_o\sqrt{i}\right\}\right) < \infty.$$

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 7.5 is then complete by the above estimate and (34). ■

Lemma 7.6 (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 $(\xi_1, \xi_2, \dots, \xi_n)$ be a random vector in $\mathbb{C}^n$ with independent entries. Assume further that for each $1 \leq i \leq n$, ξ_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

$$\mathbf{P}\left(\operatorname{dist}(R, H) \leq \frac{1}{2}\sqrt{n - \dim(H)}\right) \leq \exp(-n^{\varepsilon})$$

where $R = (\xi_1, \xi_2, \dots, \xi_n) + v$. □

Proof. Let H' be the subspace spanned by H, v, and $\mathbf{E}[R]$. Then $\dim(H') \leq \dim(H) + 2$ and $\operatorname{dist}(R, H) \geq \operatorname{dist}(R, H') = \operatorname{dist}(R', H')$ where $R' = R - \mathbf{E}[R]$. Thus it suffices to prove the lemma when $v = 0$ and $\mathbf{E}[x] = \mathbf{E}[y] = 0$.

We now perform a truncation. By Chebyshev's inequality,

$$\mathbf{P}(|\xi_i| > n^{\varepsilon}) \leq n^{-2\epsilon}.$$

Furthermore, by Hoeffding's inequality

$$\mathbf{P}\left(\sum_{i=1}^{n} \mathbf{1}_{\{|\xi_i| \leq n^{\epsilon}\}} < n - n^{1-\varepsilon}\right) \leq \exp(-n^{1-2\varepsilon})$$

where we take $\varepsilon \in (0, 1/3)$. Therefore we will prove the lemma by conditioning on the event

$$\Omega_m = \{|\xi_1| \leq n^{\varepsilon}, \dots, |\xi_m| \leq n^{\varepsilon}\}$$

with $m = \lceil n - n^{1-\varepsilon} \rceil$.

We now deal with the fact that on the event Ω_m, the random vector $(\xi_1, \dots, \xi_m)$ may have non-zero mean. Let $\mathbf{E}_m$ denote the conditional expectation with respect to the event Ω_m and the σ-algebra $\mathcal{F}_m = \sigma(\xi_{m+1}, \dots, \xi_n)$. Let W be the subspace spanned by H, u, and w where

$$u = (0, \dots, 0, \xi_{m+1}, \dots, \xi_n), \qquad w = (\mathbf{E}_m[\xi_1], \dots, \mathbf{E}_m[\xi_m], 0, \dots, 0).$$

Clearly W is $\mathcal{F}_m$-measurable. Moreover, $\dim(W) \leq \dim(H) + 2$. Define

$$Y = (\xi_1 - \mathbf{E}_m[\xi_1], \dots, \xi_m - \mathbf{E}_m[\xi_m], 0, \dots, 0) = R - u - w.$$

Then $\operatorname{dist}(R, H) \geq \operatorname{dist}(R, W) = \operatorname{dist}(Y, W)$. 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

$$\sup_{1 \leq i \leq m} |\sigma_i^2 - 1| = o(1)$$

where $\sigma_i^2 = \mathbf{E}_m|Y_i|^2$.

By Talagrand's concentration inequality [39],

$$\mathbf{P}_m(|\operatorname{dist}(Y, W) - M_m| \geq t) \leq 4\exp\left(-\frac{t^2}{16n^{2\varepsilon}}\right) \tag{35}$$

where M_m is the median of $\operatorname{dist}(Y, W)$ under Ω_m. Using (35) one can verify that

$$M_m \geq \sqrt{\mathbf{E}_m \operatorname{dist}^2(Y, W)} - Cn^{4\varepsilon}$$

for some positive constant C (see for instance [49, Lemma E.3]). Let P denote the orthogonal projection onto $W^{\perp}$. Then

$$\begin{aligned}\mathbf{E}_m \operatorname{dist}^2(Y, W) = \sum_{k=1}^{m} \mathbf{E}_m[Y_k^2] P_{kk} &\geq c\left(\sum_{k=1}^{n} P_{kk} - \sum_{k=m+1}^{n} P_{kk}\right) \\ &\geq c(n - \dim(H) - (n - m))\end{aligned}$$

for any $1/2 < c < 1$ and $n \gg_c 1$. 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). ■

We now prove Lemma 7.4.

Proof of Lemma 7.4. By Markov's inequality, it suffices to show that there exists $p > 0$ such that for a.a. $z \in \mathbb{C}$ a.s.

$$\limsup_{n\to\infty} \int s^{-p} d\nu_{\frac{1}{\sqrt{n}}X_n - zI} < \infty \qquad \text{and} \qquad \limsup_{n\to\infty} \int s^{p} d\nu_{\frac{1}{\sqrt{n}}X_n - zI} < \infty.$$

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

$$\int s^p d\nu_{\frac{1}{\sqrt{n}}X_n + \frac{1}{\sqrt{n}}F_n - zI} \leq 1 + \frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}X_n + \frac{1}{\sqrt{n}}F_n - zI\right)^*\left(\frac{1}{\sqrt{n}}X_n + \frac{1}{\sqrt{n}}F_n - zI\right)$$

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

$$\begin{aligned}\frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}X_n - zI\right)^*\left(\frac{1}{\sqrt{n}}X_n - zI\right) &\leq 1 + \frac{1}{n^2}\sum_{i,j=1}^{n}|x_{ij}|^2 - 2\mathrm{Re}\left(\frac{z}{n^{3/2}}\sum_{k=1}^{n}x_{kk}\right) + |z|^2 \\ &\longrightarrow 2 + |z|^2\end{aligned}$$

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,

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

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

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

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

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

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

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

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 $\frac{1}{\sqrt{n}}(X_n + F_n) - zI$. By Theorem 1.9, we have that for some $A > 0$,

$$\sigma_n > n^{-A} \text{ a.s.}$$

Thus,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^{-p} &\leq \frac{1}{n}\sum_{i=1}^{n-n^{1-\gamma}}\sigma_i^{-p} + \frac{1}{n}\sum_{i=n-n^{1-\gamma}}^{n}\sigma_i^{-p} \\ &\leq \frac{1}{n}\sum_{i=n^{1-\gamma}}^{n-1}\sigma_{n-i}^{-p} + \frac{1}{n}n^{1-\gamma}n^{Ap} \\ &\leq \frac{1}{c_0^p}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{i}\right)^p\right] + n^{Ap-\gamma}\end{aligned}$$

a.s. by Lemma 7.5. 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 $p < \min\{1, \gamma/A\}$. ∎

7.2 Variance Bound

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

Lemma 7.7. 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 **C0** with atom variables (ξ_1, ξ_2). Define

$$R_n := \left(\frac{1}{\sqrt{n}}X_n - zI\right), \qquad H_n(\alpha) := (R_n^* R_n - \alpha I)^{-1},$$

where $\alpha \in \mathbb{C}$ with $\mathrm{Im}(\alpha) \neq 0$. Then

$$\mathbf{E}\left|\frac{1}{n}\mathrm{tr}H_n(\alpha) - \mathbf{E}\left[\frac{1}{n}\mathrm{tr}H_n(\alpha)\right]\right|^4 \leq C\frac{c_\alpha^4}{n^2} \tag{36}$$

uniformly for $z \in \mathbb{C}$ where

$$c_\alpha = \frac{1}{|\mathrm{Im}(\alpha)|} + \frac{|\alpha|}{|\mathrm{Im}(\alpha)|^2}.$$

Moreover, for every fixed α,

$$\frac{1}{n}\mathrm{tr}H_n(\alpha) = \mathbf{E}\left[\frac{1}{n}\mathrm{tr}H_n(\alpha)\right] + O(n^{-1/8}) \text{ a.s.} \tag{37}$$

uniformly for $z \in \mathbb{C}$. □

Proof. Let $\mathbf{E}_{\leq k}$ denote conditional expectation with respect to the σ-algebra generated by $\mathbf{r}_1(X_n), \ldots, \mathbf{r}_k(X_n), \mathbf{c}_1(X_n), \ldots, \mathbf{c}_k(X_n)$. Define

$$Y_k := \mathbf{E}_{\leq k}\frac{1}{n}\mathrm{tr}H_n(\alpha)$$

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

$$\sum_{k=1}^n \alpha_k = \frac{1}{n}\mathrm{tr}H_n(\alpha) - \mathbf{E}\frac{1}{n}\mathrm{tr}H_n(\alpha).$$

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

$$R_{n,k} := \frac{1}{\sqrt{n}}X_{n,k} - zI, \qquad H_{n,k}(\alpha) := \left(R_{n,k}^* R_{n,k} - \alpha I\right)^{-1}.$$

It follows that

$$\mathbf{E}_{\leq k}\frac{1}{n}\mathrm{tr}H_{n,k}(\alpha) = \mathbf{E}_{\leq k-1}\frac{1}{n}\mathrm{tr}H_{n,k}(\alpha)$$

and hence

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

By the resolvent identity,

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

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

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

We then note that

$$\|H_n(\alpha)R_n^* R_n\| \leq \sup_{t\geq 0}\frac{t}{|t-\alpha|} \leq 1 + \sup_{t\geq 0}\frac{|\alpha|}{|t-\alpha|} \leq 1 + \frac{|\alpha|}{|\mathrm{Im}(\alpha)|}$$

since the eigenvalues of $R_n^* R_n$ are non-negative. Similarly,

$$\|H_{n,k}(\alpha) R_{n,k}^* R_{n,k}\| \leq 1 + \frac{|\alpha|}{|\mathrm{Im}(\alpha)|}.$$

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

$$|\mathrm{tr} H_n(\alpha) - \mathrm{tr} H_{n,k}(\alpha)| \leq 8 c_\alpha.$$

Thus we conclude that

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

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

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

The proof of (36) is complete.

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

$$\mathbf{P} \left(\left| \frac{1}{n} \mathrm{tr} H_n(\alpha) - \mathbf{E} \frac{1}{n} \mathrm{tr} H_n(\alpha) \right| > \varepsilon \right) \leq C \frac{c_\alpha^4}{n^2 \varepsilon^4}.$$

The result follows by taking $\varepsilon = n^{-1/8}$ and applying the Borel-Cantelli Lemma. ■

7.3 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

$$\hat{x}_{ij} = \begin{cases} x_{ij} \mathbf{1}_{\{|x_i j| \leq n^\delta\}} - \mathbf{E}[x_{ij} \mathbf{1}_{\{|x_i j| \leq n^\delta\}}], & i \neq j \\ 0, & i = j \end{cases}$$

and

$$\tilde{x}_{ij} = \begin{cases} \frac{\hat{x}_{ij}}{\sqrt{\mathbf{E}|\hat{x}_{ij}|^2}}, & i \neq j \\ 0, & i = j \end{cases}$$

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 μ and ν. We prove the following truncation lemma.

Lemma 7.8. 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,

$$\mathbf{E}[\mathrm{Re}(\tilde{x}_{ij})^k \mathrm{Im}(\tilde{x}_{ij})^l \mathrm{Re}(\tilde{x}_{ji})^m \mathrm{Im}(\tilde{x}_{ji})^p] = \mathbf{E}[\mathrm{Re}(x_{ij})^k \mathrm{Im}(x_{ij})^l \mathrm{Re}(x_{ji})^m \mathrm{Im}(x_{ji})^p] + o(1) \tag{38}$$

uniformly for $i \neq j$ and all non-negative integers k, l, m, p such that $k + l + m + p \leq 2$. □

Proof. By [3, Corollary A.42],

$$L^4(\nu_{H_n}, \nu_{\hat{H}_n}) \leq \frac{2}{n^3}\left[\mathrm{tr}(H_n + \hat{H}_n)\mathrm{tr}\left((X_n - \hat{X}_n)^*(X_n - \hat{X}_n)\right)\right]. \tag{39}$$

By the law of large numbers,

$$\begin{aligned}\frac{1}{n}\mathrm{tr}H_n &= \frac{1}{n^2}\sum_{i,j=1}^{n}|x_{ij}|^2 - 2\mathrm{Re}\left(\frac{z}{n^{3/2}}\sum_{k=1}^{n}x_{kk}\right) + |z|^2 \\ &\longrightarrow 1 + |z|^2\end{aligned}$$

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,

$$\frac{1}{n}\mathrm{tr}\hat{H}_n \longrightarrow 1 + |z|^2$$

a.s. as $n \to \infty$.

For the remaining terms, we note that

$$\begin{aligned}\frac{1}{n^2}\mathrm{tr}\left((X_n - \hat{X}_n)^*(X_n - \hat{X}_n)\right) &\leq \frac{2}{n^2}\sum_{1\leq i<j\leq n}\left[|x_{ij}|^2\mathbf{1}_{\{|x_{ij}|>n^\delta\}} + \mathbf{E}|x_{ij}|^2\mathbf{1}_{\{|x_{ij}|>n^\delta\}}\right] \\ &\quad + \frac{2}{n^2}\sum_{1\leq j<i\leq n}\left[|x_{ij}|^2\mathbf{1}_{\{|x_{ij}|>n^\delta\}} + \mathbf{E}|x_{ij}|^2\mathbf{1}_{\{|x_{ij}|>n^\delta\}}\right] \\ &\quad + \frac{1}{n^2}\sum_{i=1}^{n}|x_{ii}|^2.\end{aligned}$$

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

$$L(\nu_{H_n}, \nu_{\tilde{H}_n}) = o(1) \tag{40}$$

a.s. as $n \to \infty$.

Again using [3, Corollary A.42],

$$L^4(\nu_{\hat{H}_n}, \nu_{\tilde{H}_n}) \leq \frac{2}{n^3}\mathrm{tr}(\hat{H}_n + \tilde{H}_n)\mathrm{tr}\left((\hat{X}_n - \tilde{X}_n)^*(\hat{X}_n - \tilde{X}_n)\right).$$

It then follows that

$$L(\nu_{\hat{H}_n}, \nu_{\tilde{H}_n}) = o(1) \tag{41}$$

a.s. since

$$1 - \sqrt{\mathbf{E}|\hat{x}_{ij}|^2} = o(1)$$

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 $i \neq j$. ■

7.4 Replacement

In 1922, Lindeberg [27] gave an elegant proof of the central limit theorem using a replacement method. Recently, this technique has also been applied to study random matrices with exchangeable or independent entries (see, for example, [5, 47] and references therein). In this sub-section, we prove a comparison lemma for sequences of random matrices that satisfy condition **C0** using this method. We begin with a definition.

Definition 7.9 (Moment matching)**.** Let (ξ_1, ξ_2) and (η_1, η_2) be two random vectors in $\mathbb{C}^2$. We say that (ξ_1, ξ_2) and (η_1, η_2) match to order k if

$$\mathbf{E}[\mathrm{Re}(\xi_1)^i\mathrm{Im}(\xi_1)^j\mathrm{Re}(\xi_2)^l\mathrm{Im}(\xi_2)^m] = \mathbf{E}[\mathrm{Re}(\eta_1)^i\mathrm{Im}(\eta_1)^j\mathrm{Re}(\eta_2)^l\mathrm{Im}(\eta_2)^m]$$

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 7.10. 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 (ξ_1, ξ_2) and (η_1, η_2), respectively. Assume the moments of (ξ_1, ξ_2) and (η_1, η_2) match to order 2. Then for a.a. $z \in \mathbb{C}$ a.s.

$$\nu_{\frac{1}{\sqrt{n}}X_n - zI} - \nu_{\frac{1}{\sqrt{n}}Y_n - zI} \longrightarrow 0$$

as $n \to \infty$. □

We proceed using the Stieltjes transform. For a $n \times n$ matrix A, we define the matrices

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

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

$$\frac{\partial (G_n(A))_{ij}}{\partial \mathrm{Re}(A_{st})} = -\frac{1}{\sqrt{n}}\left[(G_n(A)R_n(A)^*)_{is}(G_n(A))_{tj} + (G_n(A))_{it}(R_n(A)G_n(A))_{sj}\right] \tag{42}$$

and

$$\frac{\partial (G_n(A))_{ij}}{\partial \mathrm{Im}(A_{st})} = -\frac{\sqrt{-1}}{\sqrt{n}}\left[(G_n(A)R_n(A)^*)_{is}(G_n(A))_{tj} - (G_n(A))_{it}(R_n(A)G_n(A))_{sj}\right]. \tag{43}$$

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

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

Using the derivatives above, we write out the power series

$$f(x_1, x_2, x_3, x_4) = f(0,0,0,0) + \sum_{k=1}^{4} \frac{\partial f}{\partial x_k}(0,0,0,0)x_k + \sum_{i,j=1}^{4} \frac{\partial^2 f}{\partial x_i \partial x_j}(0,0,0,0)x_i x_j + \varepsilon \tag{44}$$

where $|\varepsilon| \leq CM(|x_1|^3 + |x_2|^3 + |x_3|^3 + |x_4|^4)$ 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

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

and similarly

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

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

$$\frac{\partial f}{\partial x_k} = O_\alpha\left(\frac{1}{n}\right), \qquad \frac{\partial^2 f}{\partial x_k \partial x_i} = O_\alpha\left(\frac{1}{n}\right), \qquad \frac{\partial^3 f}{\partial x_k \partial x_i \partial x_j} = O_\alpha\left(\frac{1}{n}\right) \tag{45}$$

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 7.10. By Lemma 7.8 and Remark 7.2, it suffices to show that a.s.

$$\nu_{R_n(\tilde{X}_n)^* R_n(\tilde{X}_n)} - \nu_{R_n(\tilde{Y}_n)^* R_n(\tilde{Y}_n)} \longrightarrow 0. \tag{46}$$

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^δ for some $0 < \delta < \frac{1}{2}$.

By [3, Theorem B.9], we can equivalently state (46) as

$$\frac{1}{n}\mathrm{tr} G_n(\tilde{X}_n) - \frac{1}{n}\mathrm{tr} G_n(\tilde{Y}_n) \longrightarrow 0$$

a.s. for each fixed α with $\mathrm{Im}(\alpha) > 0$. However by Lemma 7.7, this reduces to showing that

$$\mathbf{E}\frac{1}{n}\mathrm{tr} G_n(\tilde{X}_n) - \mathbf{E}\frac{1}{n}\mathrm{tr} G_n(\tilde{Y}_n) \longrightarrow 0. \tag{47}$$

We will verify (47) by showing that for each fixed α with $\mathrm{Im}(\alpha) > 0$,

$$\mathbf{E} f\left(\frac{\mathrm{Re}(\xi_1)}{\sqrt{n}}, \frac{\mathrm{Im}(\xi_1)}{\sqrt{n}}, \frac{\mathrm{Re}(\xi_2)}{\sqrt{n}}, \frac{\mathrm{Im}(\xi_2)}{\sqrt{n}}\right) = \mathbf{E} f\left(\frac{\mathrm{Re}(\eta_1)}{\sqrt{n}}, \frac{\mathrm{Im}(\eta_1)}{\sqrt{n}}, \frac{\mathrm{Re}(\eta_2)}{\sqrt{n}}, \frac{\mathrm{Im}(\eta_2)}{\sqrt{n}}\right) + o_\alpha(n^{-2}) \tag{48}$$

where we take A to be any matrix independent of (ξ_1, ξ_2) and (η_1, η_2). Indeed, by allowing a and b to range over all $O(n^2)$ indices, and by the triangle inequality, we obtain

$$\mathbf{E}\frac{1}{n}\mathrm{tr} G_n(\tilde{X}_n) = \mathbf{E}\frac{1}{n}\mathrm{tr} G_n(\tilde{Y}_n) + o_\alpha(1)$$

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 7.8.

Using (44), (45), and the independence assumption from condition **C0**, we obtain

$$\mathbf{E}[f(x_1, x_2, x_2, x_4)] = \mathbf{E}[f(0,0,0,0)] + \sum_{i,j=1}^{4} \mathbf{E}\frac{\partial^2 f}{\partial x_i \partial x_j}(0,0,0,0)\mathbf{E}[x_i x_j] + \mathbf{E}[\varepsilon]$$

where

$$x_1 = \frac{\mathrm{Re}(\xi_1)}{\sqrt{n}}, \qquad x_2 = \frac{\mathrm{Im}(\xi_1)}{\sqrt{n}}, \qquad x_3 = \frac{\mathrm{Re}(\xi_2)}{\sqrt{n}}, \qquad x_4 = \frac{\mathrm{Im}(\xi_2)}{\sqrt{n}},$$

and

$$\mathbf{E}|\varepsilon| = O_\alpha\left(\frac{n^\delta}{n^{2.5}}\right)$$

for some $0 < \delta < 1/2$ from Lemma 7.8.

We repeat the same procedure for (η_1, η_2) and obtain

$$\mathbf{E}[f(y_1, y_2, y_2, y_4)] = \mathbf{E}[f(0,0,0,0)] + \sum_{i,j=1}^{4} \mathbf{E}\frac{\partial^2 f}{\partial y_i \partial y_j}(0,0,0,0)\mathbf{E}[y_i y_j] + O_\alpha\left(\frac{n^\delta}{n^{2.5}}\right)$$

where

$$y_1 = \frac{\mathrm{Re}(\eta_1)}{\sqrt{n}}, \qquad y_2 = \frac{\mathrm{Im}(\eta_1)}{\sqrt{n}}, \qquad y_3 = \frac{\mathrm{Re}(\eta_2)}{\sqrt{n}}, \qquad y_4 = \frac{\mathrm{Im}(\eta_2)}{\sqrt{n}}.$$

By (38), we have that $\mathbf{E}[x_i x_j] = \mathbf{E}[y_i y_j] + o(n^{-1})$ uniformly for $1 \leq i, j \leq 4$. Combining this with (45) yields

$$\mathbf{E}[f(y_1, y_2, y_2, y_4)] = \mathbf{E}[f(x_1, x_2, x_2, x_4)] + o_\alpha(n^{-2})$$

and the proof of Lemma 7.10 is complete.

7.5 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 7.10, 7.4, 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, Theorem A.44], it follows that for a.a. $z \in \mathbb{C}$ a.s.

$$\nu_{\frac{1}{\sqrt{n}}(X_n+F_n)-zI} - \nu_{\frac{1}{\sqrt{n}}X_n-zI} \longrightarrow 0$$

as $n \to \infty$. Since both (ξ_1, ξ_2) and (η_1, η_2) are from the (μ, ρ)-family, then (ξ_1, ξ_2) and (η_1, η_2) match to order 2. Thus by Lemma 7.10, for a.a. $z \in \mathbb{C}$ a.s.

$$\nu_{\frac{1}{\sqrt{n}}X_n-zI} \longrightarrow \nu_z$$

as $n \to \infty$. Therefore, for a.a. $z \in \mathbb{C}$ a.s.

$$\nu_{\frac{1}{\sqrt{n}}(X_n+F_n)-zI} \longrightarrow \nu_z$$

as $n \to \infty$.

Furthermore, by Lemma 7.4, for a.a. $z \in \mathbb{C}$ a.s. log is uniformly integrable for $\{\nu_{\frac{1}{\sqrt{n}}(X_n+F_n)-zI}\}_{n\geq 1}$, $\{\nu_{\frac{1}{\sqrt{n}}X_n-zI}\}_{n\geq 1}$, and $\{\nu_{\frac{1}{\sqrt{n}}Y_n-zI}\}_{n\geq 1}$. The result then follows by Lemma 7.1 and the uniqueness of the logarithmic potential [4, Lemma 4.1]. ∎

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 (ξ_1, ξ_2) and $\rho = \mathbf{E}[\xi_1\xi_2]$ for some $-1 < \rho < 1$. Let $\{Y_n\}_{n\geq 1}$ be the sequence of random matrices that satisfies condition **C0** with atom variables (η_1, η_2) where η_1 and η_2 are jointly Gaussian and $\mathbf{E}[\eta_1\eta_2] = \rho$. In [31, Theorem 5.2], it is shown that

$$\mathbf{E}\nu_{\frac{1}{\sqrt{n}}Y_n-zI} \longrightarrow \nu_z$$

as $n \to \infty$ where the family $\{\nu_z\}_{z\in\mathbb{C}}$ determines the elliptic law with parameter ρ by Lemma 7.1. In fact, using the variance bound in Lemma 7.7 and [3, Theorem B.9] it can be shown that a.s.

$$\nu_{\frac{1}{\sqrt{n}}Y_n-zI} \longrightarrow \nu_z$$

as $n \to \infty$. By Lemma 7.4, for a.a. $z \in \mathbb{C}$ a.s. log is uniformly integrable for $\{\nu_{\frac{1}{\sqrt{n}}Y_n-zI}\}_{n\geq 1}$ and hence by Lemma 7.1, we conclude that

$$\mu_{\frac{1}{\sqrt{n}}Y_n} \longrightarrow \mu_\rho$$

a.s. as $n \to \infty$.

Since both (ξ_1, ξ_2) and (η_1, η_2) are from the $(1, \rho)$-family, the proof of the theorem is complete by an application of Lemma 7.3. ∎

7.6 Proof of Theorem 1.8

In the proof of Theorem 1.5 above, we relied on the previous results in [31], where the entries are assumed to be real. In order to prove Theorem 1.8, we first need to study the complex Gaussian case.

Lemma 7.11. 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 (ξ_1, ξ_2) from the (μ, ρ)-family, where $\mathrm{Re}(\xi_1), \mathrm{Im}(\xi_1), \mathrm{Re}(\xi_2), \mathrm{Im}(\xi_2)$ are jointly Gaussian. Then for a.a. $z \in \mathbb{C}$ a.s.

$$\nu_{\frac{1}{\sqrt{n}}X_n-zI} \longrightarrow \nu_z$$

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

Let us assume Lemma 7.11 for now and complete the proof of Theorem 1.8.

Proof of Theorem 1.8. Let $\{X_n\}_{n\geq 1}$ be a sequence of complex random matrices that satisfy condition **C0** with atom variables (ξ_1, ξ_2) from the (μ, ρ)-family. Let $\{Y_n\}_{n\geq 1}$ be the sequence of complex random matrices that satisfy condition **C0** with atom variables (η_1, η_2) from the (μ, ρ)-family, where $\mathrm{Re}(\eta_1), \mathrm{Im}(\eta_1), \mathrm{Re}(\eta_2), \mathrm{Im}(\eta_2)$ are jointly Gaussian. By Lemma 7.11, for a.a. $z \in \mathbb{C}$ a.s.

$$\nu_{\frac{1}{\sqrt{n}}Y_n - zI} \longrightarrow v_z$$

as $n \to \infty$ where the family $\{v_z\}_{z\in\mathbb{C}}$ determines the elliptic law with parameter ρ by Lemma 7.1. Moreover, log is uniformly integrable for $\{\nu_{\frac{1}{\sqrt{n}}Y_n - zI}\}_{n\geq 1}$ by Lemma 7.4. Therefore, by Lemma 7.1, it follows that a.s.

$$\mu_{\frac{1}{\sqrt{n}}Y_n} \longrightarrow \mu_\rho$$

as $n \to \infty$. The proof of Theorem 1.8 is now complete by Lemma 7.3. ∎

All that remains is to prove Lemma 7.11. Let $\{X_n\}_{n\geq 1}$ be the sequence of random matrices defined in Lemma 7.11 with jointly Gaussian off-diagonal entries. We follow [31] and introduce the following notation.

For $n \times n$ matrices A and B, we define the $2n \times 2n$ bock matrices

$$V := \begin{bmatrix} \frac{1}{\sqrt{n}}A & 0 \\ 0 & \frac{1}{\sqrt{n}}B^* \end{bmatrix}, \qquad J_z := \begin{bmatrix} 0 & zI \\ \bar{z}I & 0 \end{bmatrix}$$

and set

$$V(z) := VJ_1 - J_z,$$

where

$$J_1 := \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}.$$

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

$$\begin{aligned} \frac{\partial R_{ab}}{\partial \mathrm{Re}(A_{cd})} &= -\frac{1}{\sqrt{n}} R_{ac} R_{d+n,b}, \\ \frac{\partial R_{ab}}{\partial \mathrm{Im}(A_{cd})} &= -\frac{\sqrt{-1}}{\sqrt{n}} R_{ac} R_{d+n,b}, \\ \frac{\partial R_{ab}}{\partial \mathrm{Re}(B_{cd})} &= -\frac{1}{\sqrt{n}} R_{a,d+n} R_{cb}, \\ \frac{\partial R_{ab}}{\partial \mathrm{Im}(B_{cd})} &= \frac{\sqrt{-1}}{\sqrt{n}} R_{a,d+n} R_{cb}, \end{aligned}$$

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 [37]. That is, if $Y = \{\xi_i\}_{i=1}^p$ is a real random Gaussian vector such that

$$\mathbf{E}[\xi_j] = 0, \qquad \mathbf{E}[\xi_j \xi_k] = C_{jk}$$

for $j, k = 1, 2, \ldots, p$ and if $\Phi : \mathbb{R}^p \to \mathbb{C}$ has bounded partial derivatives, then

$$\mathbf{E}[\xi_j \Phi] = \sum_{k=1}^{p} C_{jk} \mathbf{E}[(\nabla \Phi)_k].$$

Using the partial derivatives above and the Gaussian decoupling formula, we obtain

$$\begin{aligned} \mathbf{E}[R_{ab} x_{cd}] = &-\frac{1}{\sqrt{n}} \mathbf{E}[R_{a,d+n} R_{cb}] - \frac{\rho}{\sqrt{n}} \mathbf{E}[R_{ad} R_{c+n,b}] \\ &+ \frac{1-2\mu}{\sqrt{n}} \mathbf{E}[R_{a,c} R_{d+n,b}] + \frac{(1-2\mu)\rho}{\sqrt{n}} \mathbf{E}[R_{a,c+n} R_{db}] \end{aligned} \tag{49}$$

and

$$\mathbf{E}[R_{ab}\bar{x}_{cd}] = -\frac{1}{\sqrt{n}}\mathbf{E}[R_{ac}R_{d+n,b}] - \frac{\rho}{\sqrt{n}}\mathbf{E}[R_{a,c+n}R_{db}] + \frac{(1-2\mu)\rho}{\sqrt{n}}\mathbf{E}[R_{a,d}R_{c+n,b}] + \frac{1-2\mu}{\sqrt{n}}\mathbf{E}[R_{a,d+n}R_{cb}] \tag{50}$$

for $1 \leq c, d \leq n$, $c \neq d$, and $1 \leq a, b \leq 2n$.

Following [31], we define the functions

$$s_n := s_n(\alpha, z) = \frac{1}{2n}\mathbf{E}[\mathrm{tr}R] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[R_{ii}] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[R_{i+n,i+n}]$$

and

$$t_n := t_n(\alpha, z) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[R_{i+n,i}], \qquad u_n := u_n(\alpha, z) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[R_{i,i+n}].$$

We now fix $z, \alpha \in \mathbb{C}$ with $\mathrm{Im}(\alpha) > 0$. 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 7.12.

$$\mathrm{Var}\left(\frac{1}{2n}\mathrm{tr}R\right) = O_{\alpha,z}\left(\frac{1}{n}\right), \tag{51}$$

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}R_{i+n,i}\right) = O_{\alpha,z}\left(\frac{1}{n}\right), \tag{52}$$

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}R_{i,i+n}\right) = O_{\alpha,z}\left(\frac{1}{n}\right). \tag{53}$$

□

Proof. We begin by noting that

$$\frac{1}{n}\sum_{i=1}^{n}R_{i+n,i} = \frac{1}{n}\mathrm{tr}(P_2RP_1)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}R_{i,i+n} = \frac{1}{n}\mathrm{tr}(P_1RP_1^*)$$

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

$$\mathrm{Var}\left(\frac{1}{n}\mathrm{tr}(PRQ)\right) = O_{\alpha,z}\left(\frac{1}{n}\right)$$

for arbitrary partial isometries P and Q.

Let $\mathbf{E}_{\leq k}$ denote conditional expectation with respect to the σ-algebra generated by the random vectors

$$\mathbf{r}_1(X_n), \ldots, \mathbf{r}_k(X_n), \mathbf{c}_1(X_n), \ldots, \mathbf{c}_k(X_n).$$

Define

$$Y_k := \mathbf{E}_{\leq k}\frac{1}{n}\mathrm{tr}(PRQ)$$

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

$$\sum_{k=1}^{2n}\alpha_k = \frac{1}{n}\mathrm{tr}(PRQ) - \mathbf{E}\frac{1}{n}\mathrm{tr}(PRQ).$$

Thus we need to show that

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

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

$$R_k := [V_k J_1 - J_z - \alpha I]^{-1}.$$

Since

$$\mathbf{E}_{\leq k}\frac{1}{n}\mathrm{tr}(PR_kQ) = \mathbf{E}_{\leq k-1}\frac{1}{n}\mathrm{tr}(PR_kQ)$$

it follows that

$$\alpha_k = \mathbf{E}_{\leq k}\left[\frac{1}{n}\mathrm{tr}(PRQ) - \frac{1}{n}\mathrm{tr}(PR_kQ)\right] - \mathbf{E}_{\leq k-1}\left[\frac{1}{n}(PRQ) - \frac{1}{n}\mathrm{tr}PR_kQ)\right].$$

Because $V_k - V$ is at most rank 4, we have that

$$|\mathrm{tr}(PRQ) - \mathrm{tr}(PR_kQ)| \leq 4\|R((V_kJ_1 - J_z) - (VJ_1 - J_z))R_k\|.$$

We now claim that

$$\|R((V_kJ_1 - J_z) - (VJ_1 - J_z))R_k\| = O_{\alpha,z}(1) \tag{54}$$

uniformly in k. Indeed,

$$\|R(VJ_1 - J_z)R_k\| \leq \frac{1}{|\mathrm{Im}(\alpha)|}\|R(VJ_1 - J_z)\| \leq \frac{1}{|\mathrm{Im}(\alpha)|}\sup_{t\in\mathbb{R}}\frac{|t|}{|t-\alpha|} = O_{\alpha,z}(1)$$

since $VJ_1 - J_z$ is Hermitian. Similarly,

$$\|R(V_kJ_1 - J_z)R_k\| = O_{\alpha,z}(1)$$

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

$$\mathbf{E}\left|\sum_{k=1}^{2n}\alpha_k\right|^2 \leq C\mathbf{E}\sum_{k=1}^{2n}|\alpha_k|^2 = O_{\alpha,z}\left(\frac{1}{n}\right).$$

■

We are now ready to prove Lemma 7.11.

Proof of Lemma 7.11. Fix $z, \alpha \in \mathbb{C}$ with $\mathrm{Im}(\alpha) > 0$. By the resolvent identity,

$$1 + \alpha s_n = \frac{1}{2n}\mathbf{E}\mathrm{tr}(RVJ_1) - \frac{\bar{z}}{2}u_n - \frac{z}{2}t_n.$$

We decompose,

$$\frac{1}{2n}\mathbf{E}\mathrm{tr}(RVJ_1) = \frac{1}{2}A_1 + \frac{1}{2}A_2$$

where

$$A_1 = \frac{1}{n}\mathbf{E}\sum_{i=1}^{n}(RVJ_1)_{ii} \qquad \text{and} \qquad A_2 = \frac{1}{n}\mathbf{E}\sum_{i=1}^{n}(RVJ_1)_{i+n,i+n}.$$

By (50) and Lemma 7.12, we have

$$\begin{aligned} A_1 &= \frac{1}{n^{3/2}}\mathbf{E}\sum_{i,j=1}^{n} R_{i,j+n}\bar{x}_{ij} \\ &= -\frac{1}{n^2}\mathbf{E}\sum_{i,j=1}^{n}[R_{ii}R_{j+n,j+n} + \rho R_{i,i+n}R_{j,j+n} - (1-2\mu)\rho R_{i,j}R_{i+n,j+n} \\ &\qquad - (1-2\mu)R_{i,j+n}R_{i,j+n}] + o_{\alpha,z}(1) \\ &= -s_n^2 - \rho u_n^2 + o_{\alpha,z}(1). \end{aligned}$$

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

$$\left|\frac{1}{n^2}\sum_{i,j=1}^{n} R_{i,j}R_{i+n,j+n}\right| = \left|\frac{1}{n^2}\mathrm{tr}(R_1^{\mathrm{T}}R_4)\right| \le \frac{1}{n}\|R\|^2 \le \frac{1}{n|\mathrm{Im}(\alpha)|^2}$$

and

$$\left|\frac{1}{n^2}\sum_{i,j=1}^{n} R_{i,j+n}R_{i,j+n}\right| = \left|\frac{1}{n^2}\mathrm{tr}(R_2^{\mathrm{T}}R_2)\right| \le \frac{1}{n|\mathrm{Im}(\alpha)|^2}.$$

Similarly (using (49) and Lemma 7.12),

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

Thus

$$1 + \alpha s_n = -s_n^2 - \frac{\rho}{2}u_n^2 - \frac{\rho}{2}t_n^2 - \frac{\bar{z}}{2}u_n - \frac{z}{2}t_n + o_{\alpha,z}(1). \tag{55}$$

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

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

where

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

We repeat almost exactly the same procedure as above (using (50) and Lemma 7.12) and obtain

$$\begin{aligned} A_3 &= -\frac{1}{n^2}\mathbf{E}\sum_{i,j=1}^{n}[R_{i+n,i}R_{j+n,j+n} + \rho R_{i+n,i+n}R_{j,j+n} \\ &\qquad - (1-2\mu)\rho R_{i+n,j}R_{i+n,j+n} - (1-2\mu)R_{i+n,j+n}R_{i,j+n}] + o_{\alpha,z}(1) \\ &= -t_n s_n - \rho s_n u_n + o_{\alpha,z}(1). \end{aligned}$$

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:

$$\begin{aligned} 1 + \alpha s_n &= -s_n^2 - \frac{\rho}{2}u_n^2 - \frac{\rho}{2}t_n^2 - \frac{\bar{z}}{2}u_n - \frac{z}{2}t_n + o_{\alpha,z}(1) \\ \alpha t_n &= -t_n s_n - \rho s_n u_n - \bar{z}s_n + o_{\alpha,z}(1) \\ \alpha u_n &= -u_n s_n - \rho s_n t_n - z s_n + o_{\alpha,z}(1). \end{aligned} \tag{56}$$

We note that the system of equations above does not depend on μ. In addition to the case above, we also consider the case when $\mu = 1$. This corresponds to the real Gaussian case studied in [31]. Repeating the same calculations as above, we obtain the following system of equations in the real Gaussian case:

$$\begin{aligned} 1 + \alpha \hat{s}_n &= -\hat{s}_n^2 - \frac{\rho}{2}\hat{u}_n^2 - \frac{\rho}{2}\hat{t}_n^2 - \frac{\bar{z}}{2}\hat{u}_n - \frac{z}{2}\hat{t}_n + o_{\alpha,z}(1) \\ \alpha \hat{t}_n &= -\hat{t}_n \hat{s}_n - \rho \hat{s}_n \hat{u}_n - \bar{z}\hat{s}_n + o_{\alpha,z}(1) \\ \alpha \hat{u}_n &= -\hat{u}_n \hat{s}_n - \rho \hat{s}_n \hat{t}_n - z\hat{s}_n + o_{\alpha,z}(1). \end{aligned} \tag{57}$$

One can also check that this system matches (5.4), (5.5), and (5.6) from [31]. In [31], it is shown that for every $\alpha, z \in \mathbb{C}$ with $\mathrm{Im}(\alpha) > 0$,

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

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

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

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

By Lemma 7.12 and [3, Lemma B.9], it suffices to show that for every $\alpha, z \in \mathbb{C}$ with $\mathrm{Im}(\alpha) > 0$,

$$\lim_{n\to\infty} s_n = s_0 \tag{58}$$

in order to complete the proof of Lemma 7.11.

Since $\|R\| \le \frac{1}{\mathrm{Im}(\alpha)}$, it follows that $|s_n|, |t_n|, |u_n| \le \frac{1}{\mathrm{Im}(\alpha)}$. Similarly, $|\hat{s}_n|, |\hat{t}_n|, |\hat{u}_n| \le \frac{1}{\mathrm{Im}(\alpha)}$. 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 $\mathrm{Im}(\alpha)$ sufficiently large.

Taking $\mathrm{Im}(\alpha) > 1$, we subtract the last two equations of (56) and (57) to obtain

$$\begin{aligned} |t_n - \hat{t}_n| &\le \frac{1}{\mathrm{Im}(\alpha)^2 - 1}|u_n - \hat{u}_n| + \frac{1 + \rho + \mathrm{Im}(\alpha)|z|}{\mathrm{Im}(\alpha)^2 - 1}|s_n - \hat{s}_n| + o_{\alpha,z}(1) \\ |u_n - \hat{u}_n| &\le \frac{1}{\mathrm{Im}(\alpha)^2 - 1}|t_n - \hat{t}_n| + \frac{1 + \rho + \mathrm{Im}(\alpha)|z|}{\mathrm{Im}(\alpha)^2 - 1}|s_n - \hat{s}_n| + o_{\alpha,z}(1). \end{aligned}$$

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

$$\begin{aligned} |t_n - \hat{t}_n| &\le \frac{1}{100}|u_n - \hat{u}_n| + \frac{1}{100}|s_n - \hat{s}_n| + o_{\alpha,z}(1) \\ |u_n - \hat{u}_n| &\le \frac{1}{100}|t_n - \hat{t}_n| + \frac{1}{100}|s_n - \hat{s}_n| + o_{\alpha,z}(1) \end{aligned}$$

and hence

$$\begin{aligned} |t_n - \hat{t}_n| &\le \frac{2}{99}|s_n - \hat{s}_n| + o_{\alpha,z}(1) \\ |u_n - \hat{u}_n| &\le \frac{2}{99}|s_n - \hat{s}_n| + o_{\alpha,z}(1). \end{aligned}$$

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

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

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

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

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

and the proof of Lemma 7.11 is complete. ∎

Acknowledgements

The authors would like to thank B. Khoruzhenko for pointing out reference [20], and for providing many valuable comments and suggestions. The authors are grateful to the anonymous referees for many valuable suggestions and corrections.

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

$$\sup_a \mathbf{P}\left(\left|\sum_i (a_i x_i + b_i x_i') - a\right| \le r\right)$$
$$\le \exp(\pi r^2) \int_{\mathbf{C}} \exp\left(-\sum_{i=1}^n \mathbf{E}_{\xi_1,\xi_2,\xi_1',\xi_2'} \|\mathrm{Re}(2(\xi_1-\xi_1')a_i t + 2(\xi_2-\xi_2')b_i t)\|_{\mathbf{R}/\mathbf{Z}}^2 - \pi|t|^2\right) dt,$$

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

Proof. (of Claim A.1) First of all, we have

$$\mathbf{P}\left(\sum_{i=1}^n (a_i x_i + b_i x_i') \in B(a,r)\right) = \mathbf{P}\left(\left|\sum_{i=1}^n a_i x_i + b_i x_i' - a\right|^2 \le r^2\right)$$
$$= \mathbf{P}\left(\exp(-\pi|\sum_{i=1}^n a_i x_i + b_i x_i' - a|^2) \ge \exp(-\pi r^2)\right)$$
$$\le \exp(\pi r^2)\mathbf{E}\exp\left(-\pi\left|\sum_{i=1}^n a_i x_i + b_i x_i' - a\right|^2\right).$$

Note that for any $z \in \mathbf{R}^2$, $\exp(-\pi|z|^2) = \int_{\mathbf{C}} e(zt)\exp(-\pi|t|^2)dt$, where $e(u) := \exp(2\pi\sqrt{-1}\mathrm{Re}(u))$. Thus,

$$\mathbf{P}\left(\sum_{i=1}^n a_i x_i + b_i x_i' \in B(a,r)\right) \le \exp(\pi r^2)\int_{\mathbf{C}} \mathbf{E}e\left(\left(\sum_{i=1}^n a_i x_i + b_i x_i'\right)t\right) e(-at)\exp(-\pi|t|^2)dt.$$

Next, because of independence we have $|\mathbf{E}e((\sum_{i=1}^n a_i x_i + b_i x_i')t)| = \prod_{i=1}^n |\mathbf{E}e(x_i a_i t + x_i' b_i t)|$, and so

$$|\mathbf{E}e(x_i a_i t + x_i' b_i t)| \le |\mathbf{E}e(x_i a_i t + x_i' b_i t)|^2/2 + 1/2$$
$$= \mathbf{E}_{\xi_1,\xi_2,\xi_1',\xi_2'} e\Big((\xi_1-\xi_1')a_i t + (\xi_2-\xi_2')b_i t\Big)/2 + 1/2$$
$$= \mathbf{E}_{\xi_1,\xi_2,\xi_1',\xi_2'} \cos\Big(2\pi\mathrm{Re}\big((\xi_1-\xi_1')a_i t + (\xi_2-\xi_2')b_i t\big)\Big)/2 + 1/2$$
$$\le \exp\Big(-\mathbf{E}_{\xi_1,\xi_2,\xi_1',\xi_2'} \|\mathrm{Re}\big(2(\xi_1-\xi_2)a_i t + 2(\xi_1'-\xi_2')b_i t\big)\|_{\mathbf{R}/\mathbf{Z}}^2\Big),$$

where the random vector (ξ_1', ξ_2') is an identical independent copy of (ξ_1, ξ_2), and in the last inequality we estimated crudely $|\cos \pi z| \le 1 - \sin^2(\pi z)/2 \le 1 - 2\|z\|_{\mathbf{R}/\mathbf{Z}}^2 < \exp(-\|z\|_{\mathbf{R}/\mathbf{Z}}^2)$. ■

Observe that, as (ξ_1, ξ_2) belongs to a given (μ, ρ)-family, so does the pair $(\omega_1, \omega_2) := ((\xi_1 - \xi_1')/2, (\xi_2 - \xi_2')/2)$. Intuitively, for $\mathbf{E}|\psi_1|^2 = \mathbf{E}|\psi_2|^2 = 1$ and $|\rho| = |\mathbf{E}[\psi_1\psi_2]| < 1$, these two random variables are essentially not multiples of each other. We summarize this useful fact as a claim below.

Claim A.2. Assume that (ω_1, ω_2) belongs to a given (μ, ρ)-family. Then there exist positive numbers α, δ, c_0, C_0 and two Lebesgue-measurable sets R_1 and R_2 in the set $\{(x, y) \in \mathbf{C}^2, c_0 < |x|, |y| < C_0\}$ such that $\mathbf{P}((\omega_1, \omega_2) \in R_1), \mathbf{P}((\omega_1, \omega_2) \in R_2) \ge \delta$ 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 ϵ_0 be a sufficiently small positive constant to be chosen. There exist positive numbers c_0, C_0 depending on ω_1, ω_2 and on ϵ_0 such that the truncated random variables $\psi_1 := \omega_1 \mathbf{1}_{c_0<|\omega_1|<C_0}, \psi_2 := \omega_2 \mathbf{1}_{c_0<|\omega_2|<C_0}$ satisfy the following

1. $1 - \epsilon_0 \leq \mathbf{E}|\psi_1|^2, \mathbf{E}|\psi_2|^2 \leq 1 + \epsilon_0$,
2. $|\rho| - \epsilon_0 \leq |\mathbf{E}[\psi_1\psi_2]| \leq |\rho| + \epsilon_0$.

Observe that it suffices to justify the claim for the truncated pair (ψ_1, ψ_2). Set k to be a sufficiently large integer. We divide the square $Q := \{z \in \mathbf{C}, |\mathrm{Im}(z)|, |\mathrm{Re}(z)| \leq C_0/c_0\}$ into k^2 closed smaller squares $Q_1, \dots, Q_{k^2}$ of size $2C_0/kc_0$ each, and then divide the region $R := \{(x, y) \in \mathbf{C}^2, c_0 < |x|, |y| < C_0\}$ into k^2 closed regions $R_i, i = 1, \dots, k^2$ 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 $\delta > 0$ (depending on c_0, C_0, k), there are squares Q_{i_0}, Q_{j_0} that are not adjacent (i.e. sharing a common edge) and that $\mathbf{P}(\psi_1/\psi_2 \in Q_{i_0}) \geq \delta$ and $\mathbf{P}(\psi_1/\psi_2 \in Q_{j_0}) \geq \delta$. Indeed, assuming otherwise, $\mathbf{P}(\psi_1/\psi_2 \in Q_i) < \delta$ 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

$$\mathbf{P}(\psi_1/\psi_2 \in Q') \geq 1 - (k^2 - 9)\delta.$$

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 ψ_2/ψ_1 instead). Clearly,

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

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 $(|c| - 6C_0/kc_0)\mathbf{E}|\psi_1|^2 - C_0^2(k^2 - 9)\delta$, which is at least $(1 - 6C_0/kc_0)(1 - \epsilon_0) - C_0^2(k^2 - 9)\delta$ because $|c| \geq 1$ and $\mathbf{E}|\psi_1|^2 \geq 1 - \epsilon_0$ from item 1 above. Finally, by choosing k to be large enough (depending on ϵ_0, c_0, C_0) and then δ to be small enough (depending on ϵ_0, C_0 and k), we obtain a lower bound $1 - 2\epsilon_0$ for $|\mathbf{E}[\psi_1\psi_2]|$. This is impossible as from item 2 we have $|\mathbf{E}[\psi_1\psi_2]| \leq |\rho| + \epsilon_0 < 1 - 2\epsilon_0$.

In summary, we have obtained two closed sub-regions R_{i_0}, R_{j_0} of R such that the corresponding squares Q_{i_0} and Q_{j_0} are not adjacent and that both $\mathbf{P}(\psi_1/\psi_2 \in Q_{i_0})$ and $\mathbf{P}(\psi_1/\psi_2 \in Q_{j_0})$ are greater than δ. By definition, as Q_{i_0} and Q_{j_0} are not adjacent, we have $|a/b - c/d| \geq 2C_0/kc_0$ as long as $(a, b) \in R_{i_0}$ and $(c, d) \in R_{j_0}$, completing the proof. ∎

We now apply Claim A.1 and A.2 to prove Theorem 3.2. Our method here follows [35] 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 and z' the random variables $2(\xi_1 - \xi_2)$ and $2(\xi'_1 - \xi'_2)$ respectively, where (ξ'_1, ξ'_2) is an identical independent copy of (ξ_1, ξ_2). By definition, we have

$$\gamma = \sup_a \mathbf{P}\left(\left|\sum_i (a_i x_i + b_i x_i) - a\right| \leq \beta\right) = \sup_a \mathbf{P}\left(\left|\sum_i (a'_i x_i + b'_i x'_i) - a\right| \leq 1\right) = n^{-O(1)}.$$

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

$$\begin{aligned}\frac{\gamma}{2} &\leq \int_{|t|\leq M} \exp\left(-\sum_{i=1}^n \mathbf{E}_{\xi_1,\xi_2,\xi'_1,\xi'_2} \left\|\mathrm{Re}\Big(2(\xi_1 - \xi'_1)a'_i t + 2(\xi_2 - \xi'_2)b'_i t)\Big)\right\|^2_{\mathbf{R}/\mathbf{Z}} - \pi|t|^2\right) dt \\ &= \int_{|t|\leq M} \exp\left(-\sum_{i=1}^n \mathbf{E}_{z,z'} \left\|\mathrm{Re}\Big(za'_i t + z'b'_i t\Big)\right\|^2_{\mathbf{R}/\mathbf{Z}} - \pi|t|^2\right) dt. \qquad (59)\end{aligned}$$

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

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

Then it follows from (59) that $\sum_{m\leq M} \mu(S_m) \exp(-\frac{m}{2} + 1) \geq \gamma$, where $\mu(\cdot)$ denotes the Lebesgue measure of a measurable set. Hence there exists $m \leq M$ such that $\mu(S_m) \geq \gamma \exp(\frac{m}{4} - 2)$.

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

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

Consider $t_1,t_2 \in B(x_0,1/2)\cap S_m$. By the Cauchy-Schwarz inequality (note that $\mathbf{E}_{z,z'}\|\mathrm{Re}(za_i't+z'b_i't)\|_{\mathbf{R/Z}}^2$ satisfies the triangle inequality in t) we have

$$\begin{aligned}&\sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i'(t_1-t_2)+z'b_i'(t_1-t_2)\Big)\right\|_{\mathbf{R/Z}}^2\\ &\leq 2\left(\sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i't_1+z'b_i't_1\Big)\right\|_{\mathbf{R/Z}}^2+\sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i't_2+z'b_i't_2\Big)\right\|_{\mathbf{R/Z}}^2\right)\leq 4m.\end{aligned}$$

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

$$T := \left\{t\in B(0,1): \sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i't+z'b_i't\Big)\right\|_{\mathbf{R/Z}}^2 \leq 4m\right\},$$

then

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

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

$$B_0 := \left\{k_1/N+\sqrt{-1}k_2/N : k_1,k_2\in\mathbf{Z}, -N\leq k_1,k_2\leq N\right\}.$$

We consider all the shifted boxes $z+B_0$, where $(\mathrm{Re}(z),\mathrm{Im}(z))\in[0,1/N]^2$. 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 $t\in(z_0+B_0)\cap T$ we have

$$\begin{aligned}&\sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i'(t-t_0)+z'b_i'(t-t_0)\Big)\right\|_{\mathbf{R/Z}}^2\\ &\leq 2\sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i't+z'b_i't\Big)\right\|_{\mathbf{R/Z}}^2\\ &+2\sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i't_0+z'b_i't_0\Big)\right\|_{\mathbf{R/Z}}^2\leq 16m.\end{aligned}$$

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

$$\sum_{i=1}^n \mathbf{E}_{z,z'}\left\|\mathrm{Re}\Big(za_i's+z'b_i's\Big)\right\|_{\mathbf{R/Z}}^2\leq 16m.$$

Double counting and separation. By definition of S, we have

$$\mathbf{E}_{z,z'}\sum_{s\in S}\sum_{i=1}^n \left\|\mathrm{Re}\Big(za_i's+z'b_i's\Big)\right\|_{\mathbf{R/Z}}^2\leq 16m|S|.$$

Notice that, for $z=2(\xi_1-\xi_1')$ and $z'=2(\xi_2-\xi_2')$, $(z/4,z'/4)$ belongs to the (μ,ρ)-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

$$\sum_{s\in S}\sum_{i=1}^n \left\|\mathrm{Re}\Big((4c_1a_i'+4c_2b_i')s\Big)\right\|_{\mathbf{R/Z}}^2\leq 16\delta^{-1}m|S|$$

and

$$\sum_{s\in S}\sum_{i=1}^{n}\Big\|\mathrm{Re}\Big((4c_1'a_i'+4c_2'b_i')s\Big)\Big\|_{\mathbf{R}/\mathbf{Z}}^2\le 16\delta^{-1}m|S|.$$

From now on, for brevity, we denote by v_i the complex number $4c_1a_i'+4c_2b_i'$ for $1\le i\le n$, and by v_{n+i} the complex number $4c_1'a_i'+4c_2'b_i'$ for $1\le i\le n$. We then have

$$\sum_{s\in S}\sum_{i=1}^{2n}\|\mathrm{Re}(v_is)\|_{\mathbf{R}/\mathbf{Z}}^2\le 32\delta^{-1}m|S|.$$

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

$$\sum_{s\in S}\sum_{i=1}^{2n}\|\langle v_i,s\rangle\|_{\mathbf{R}/\mathbf{Z}}^2\le 32\delta^{-1}m|S|.$$

Let n' be any number between n^ϵ and $2n$. We say that an index $1\le i\le 2n$ is *bad* if

$$\sum_{s\in S}\|\langle v_i,s\rangle\|_{\mathbf{R}/\mathbf{Z}}^2\ge\frac{32\delta^{-1}m|S|}{n'}.$$

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 (counting multiplicities). 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

$$\sum_{s\in S}\|\langle s,v_i\rangle\|_{\mathbf{R}/\mathbf{Z}}^2\le 32\delta^{-1}m|S|/n'.$$

Set $k:=\Big\lfloor\sqrt{\frac{n'}{2048\pi^2\delta^{-1}m}}\Big\rfloor$ and let $V_k:=k(V\cup\{0\})$, the Minkowski sum of k copies of $V\cup\{0\}$. By the Cauchy-Schwarz inequality, for any $v\in V_k$ we have

$$\sum_{s\in S}2\pi^2\|\langle s,v\rangle\|_{\mathbf{R}/\mathbf{Z}}^2\le\frac{|S|}{2},$$

which implies

$$\sum_{s\in S}\cos(2\pi\langle s,v\rangle)\ge\frac{|S|}{2}.$$

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 $\cos(2\pi\langle s,x\rangle)\ge 1/2$ and $\sin(2\pi\langle s,x\rangle)\le 1/12$. Thus for any $x\in C(0,\frac{1}{512})$,

$$\sum_{s\in S}\cos\left(2\pi\langle s,(v+x)\rangle\right)\ge\frac{|S|}{4}-\frac{|S|}{12}=\frac{|S|}{6}.$$

On the other hand one can easily check that

$$\begin{aligned}\int_{x\in[0,N]^2}\left(\sum_{s\in S}\cos(2\pi\langle s,x\rangle)\right)^2dx&\le\sum_{s_1,s_2\in S}\int_{x\in[0,N]^2}\exp\left(2\pi\sqrt{-1}\langle s_1-s_2,x\rangle\right)dx\\&\ll|S|N^2.\end{aligned}$$

Hence we deduce the following

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

Now using 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

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

Thus, we have obtained the following

$$\mu\left(k(V\cup\{0\}) + C\left(0,\frac{1}{512}\right)\right) \ll \gamma^{-1}\exp\left(-\frac{m}{4}+2\right)m. \tag{60}$$

Let $D := 2048\times 16\times\delta^{-1} = \Theta(\delta^{-1})$. We approximate each vector v of V by a closest vector in $(\frac{\mathbf{Z}}{Dk})^2$,

$$\left\|v - \frac{u}{Dk}\right\|_2 \le \frac{\sqrt{2}}{Dk}, \text{ with } u\in\mathbf{Z}^2.$$

Let U be the union of the collection of all such u with $\{0\}$. Since $\sum_{v\in V}\|v\|_2^2 = O(\beta^{-2})$, we have

$$\sum_{u\in U}\|u\|_2^2 = O_{\delta^{-1}}(k^2\beta^{-2}). \tag{61}$$

It follows from (60) that

$$\begin{aligned}|k(U+C_0(0,1))| &= O\left(\gamma^{-1}(Dk)^2\exp(-\frac{m}{4}+2)m\right)\\ &= O\left(\gamma^{-1}k^2\exp(-\frac{m}{4}+2)m\right),\end{aligned} \tag{62}$$

where $C_0(0,r)$ is the discrete cube $C_0(0,r) = \{(x,y)\in\mathbf{Z}^2, |x|,|y|\le r\}$.

We next pause to recall some useful results from [35] and [41]. For any integer $t\ge 1$, we say that a symmetric GAP $Q = \{k_1g_1+\cdots+k_rg_r, k_i\in\mathbf{Z}, |k_i|\le N_i\}$ is *t-proper* if the GAP $tQ = \{k_1g_1+\cdots+k_rg_r, k_i\in\mathbf{Z}, |k_i|\le tN_i\}$ is proper.

Lemma A.3. [41, Theorem 1.21, also Theorem 1.17] Let $A>0$ be a constant. Assume that X is a subset of integers such that $|lX|\le l^A|X|$ for some number $l\ge 2$. Then lX is contained in a symmetric 2-proper GAP Q of rank $r = O_A(1)$, and of cardinality $O_A(|lX|)$. □

Lemma A.4. [35, Lemma A.2] Assume that $0\in X$ and that $P = \{\sum_{i=1}^r x_ia_i : |x_i|\le N_i\}$ is a symmetric 2-proper GAP that contains kX. Then $X\subset\{\sum_{i=1}^r x_ia_i : |x_i|\le 2N_i/k\}$. □

Thus by (62) and Lemma A.3, there exists a 2-proper symmetric GAP $R = \{\sum_{i=1}^r x_ig_i : |x_i|\le M_i\}$ that contains $k(U+C_0(0,1))$ and

$$r = O(1) \text{ and } |R| = O(\gamma^{-1}k^2\exp(-\frac{m}{4}+2)m). \tag{63}$$

Furthermore, by Lemma A.4

$$U + C_0(0,1)\subset P := \left\{\sum_{i=1}^r x_ig_i : |x_i|\le 2M_i/k\right\}.$$

We remark that as $k(U+C_0(0,1))$ is a dense copy of R, there exist $m_1, m_2 = O(1)$ such that the dilated GAP $m_1\cdot R$ can be contained in the set $m_2k(U+C_0(0,1))$ (see for instance [40, Lemma B.3]). Using (61), we conclude that all the generators g_i of P are bounded,

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

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

Case 1: $r\ge 3$. Recall that $|P| = O(\gamma^{-1}k^{(2-r)}\exp(-\frac{m}{4}+2)m) = O(\gamma^{-1}/\sqrt{n'})$. 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 $O(\frac{\beta\sqrt{\log n'}}{\sqrt{n'}})$-close to Q.)

Case 2: $r = 2$. Because the unit vectors $e_1 = (1, 0), e_2 = (0, 1)$ belong to $P = \{\sum_{i=1}^{2} x_i g_i : |x_i| \leq 2M_i/k\} \subset \mathbf{Z}^2$, the set of generators g_1, g_2 forms a basis with unit determinant in $\mathbf{R}^2$. In this case we will be making use of R. Note that by definition $C_0(0, k) \subset R$.

Let r_0 be the smallest positive number such that

$$|6R \cap C_0(0, r_0 k)| \geq 50|R|/k. \tag{64}$$

By definition,

$$|6R \cap C_0(0, r_0 k/2)| < 50|R|/k. \tag{65}$$

Now let $p \in P$ be an arbitrary element of P, then one has

$$\|p\|_\infty \leq 2r_0 k.$$

Indeed, assume otherwise, then the sets $jp + (6R \cap C_0(0, r_0 k)), -k \leq j \leq k$ are disjointly sitting inside $2R + 6R = 8R$; thus

$$2k|6R \cap C_0(0, r_0 k)| \leq |8R| < 100|R|,$$

where we used the fact that R has rank 2 in the last estimate, a contradiction against (64).

For later use, we record a useful fact as follows.

Fact A.5. For any $z_0 \in \mathbf{R}^2$ we have

$$|3R \cap (z_0 + C_0(0, r_0 k/4))| \leq 50|R|/k.$$

□

Proof. (of Fact A.5) By the elementary bound $|X| \leq |X - X|$ and by (65),

$$\begin{aligned} |3R \cap (z_0 + C_0(0, r_0 k/4))| &\leq |[3R \cap (z_0 + C_0(0, r_0 k/4))] - [3R \cap (z_0 + C_0(0, r_0 k/4)]| \\ &\leq |6R \cap C_0(0, r_0 k/2)| \\ &\leq 50|R|/k. \end{aligned}$$

■

We next consider two subcases.

Subcase 1. Suppose $r_0 < 10$. Then as $P \subset C_0(0, 2r_0 k) \subset C_0(0, 20k)$, it is easy to find a GAP S of size $O(1)$ which is k-close to P (and hence k-close to U because $U \subset P$).

Subcase 2. Suppose $r_0 \geq 10$. Note that $P \subset R \cap C_0(0, 2r_0 k)$. With room to spare, let Z be the intersection of $2R \cap C_0(0, (2r_0 + 1)k)$ with the lattice $\Gamma := \{(ki, kj), i, j \in \mathbf{Z}\}$. Note that Z in non-empty. We next state some nice properties about Z.

Claim A.6. We have

1. P (and hence U) are $O(k)$-close to Z.
2. There is a GAP S of small rank and size $O(|Z|)$ that contains Z.
3. The size of Z is $O(\gamma^{-1}/\sqrt{n'})$.

□

Proof. (of Claim A.6) For part 1, note that

$$P + C_0(0, k) \subset [R \cap C_0(0, 2r_0 k)] + C_0(0, k) \subset 2R \cap C_0(0, (2r_0 + 1)k).$$

Thus $(P + C_0(0, k)) \cap \Gamma \subset Z$. In other words, for every element $p \in P$, there exists $z \in Z$ such that $\|z - p\|_2 \leq \sqrt{2}k$.

For 2, because the generators g_1, g_2 of Q form a basis with unit determinant in $\mathbf{R}^2$, we can view Z as the intersection between Γ and the symmetric convex body in $\mathbf{R}^2$ that corresponds to $\mathrm{conv}(2R \cap C_0(0, (2r_0 + 1)k))$. Thus Z can be contained in a GAP S of rank 2 and size $|S| = O(|Z|)$ by [50, Lemma 3.36].

For 3, note that the sets $z + (R \cap C_0(0, k/4)), z \in Z$ are disjointly lying inside $3R \cap C_0(0, 3r_0k)$, and so

$$|Z| \leq \frac{|3R \cap C_0(0, 3r_0k)|}{|R \cap C_0(0, k/4)|} = \frac{|3R \cap C_0(0, 3r_0k)|}{|C_0(0, k/4)|} \leq \frac{O(|R|/k)}{k^2} = O(\gamma^{-1} \exp(-\frac{m}{4} + 2)m/k) = O(\gamma^{-1}/\sqrt{n'}),$$

where in the third estimate we decomposed $C_0(0, 3r_0k)$ into disjoint copies of $C_0(0, r_0k/4)$ and applied Fact A.5, and we used the bound in (63) for $|R|$ in the second to last estimate. ■

Therefore, in both subcases P is $O(k)$-close a GAP S of small rank and size $O(1 + \gamma^{-1}/\sqrt{n'})$. To obtain Q as concluded in Theorem 3.2 we just set

$$Q := \frac{\beta}{Dk} \cdot S.$$

■

B **Proof of Lemma 4.2**

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

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

By Markov's inequality we have

$$\begin{aligned}
&\mathbf{P}_{\mathbf{x}, \mathbf{x}'} \big(|\sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a| \leq 1 \big) \\
&= \mathbf{P} \big(\exp(-\frac{\pi}{2} |\sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a|^2 \geq \exp(-\frac{\pi}{2})) \big) \\
&\leq \exp(\frac{\pi}{2}) \mathbf{E}_{\mathbf{x}, \mathbf{x}'} \exp \big(-\frac{\pi}{2} |\sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a|^2 \big) \\
&\leq \exp(\frac{\pi}{2}) \int_{\mathbf{C}} \big| \mathbf{E}_{\mathbf{x}, \mathbf{x}'} e[(\sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i) \cdot t] \big| \exp(-\frac{\pi}{2} |t|^2) dt \\
&\leq \exp(\frac{\pi}{2}) (\sqrt{2\pi})^2 \int_{\mathbf{C}} \big| \mathbf{E}_{\mathbf{x}, \mathbf{x}'} e[(\sum_{i,j} a'_{ij} x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i)) \cdot t] \big| \exp(-\frac{\pi}{2} |t|^2) / (\sqrt{2\pi})^2 dt
\end{aligned}$$

where in the fourth equation we used the identity $\exp(-\frac{\pi}{2}|x|^2) = \int_{\mathbf{C}} e(xt) \exp(-\frac{\pi}{2}|t|^2) dt$.

Consider $\mathbf{x} = (x_1, \ldots, x_n)$ as $(\mathbf{x}_U, \mathbf{x}_{\bar{U}})$ and $\mathbf{x}' = (x'_1, \ldots, x'_n)$ as $(\mathbf{x}'_U, \mathbf{x}'_{\bar{U}})$, where $\mathbf{x}_U, \mathbf{x}'_U$ and $\mathbf{x}_{\bar{U}}, \mathbf{x}'_{\bar{U}}$ are the vectors corresponding to $i \in U$ and $i \notin U$ respectively. After a series applications of the identity

$\int_{\mathbf{C}} \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt = 1$ and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
&\Big[\int_{\mathbf{C}} \Big|\mathbf{E}_{\mathbf{x},\mathbf{x}'} e\big((\sum_{i,j} a'_{ij}x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i)\cdot t\big)\Big| \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt\Big]^4 \\
&\le \Big[\int_{\mathbf{C}} \Big|\mathbf{E}_{\mathbf{x},\mathbf{x}'} e\big((\sum_{i,j} a'_{ij}x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i))\cdot t\big)\Big|^2 \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt\Big]^2 \\
&\le \Big[\int_{\mathbf{C}} \mathbf{E}_{\mathbf{x}_U,\mathbf{x}'_U}\Big|\mathbf{E}_{\mathbf{x}_{\bar U},\mathbf{x}'_{\bar U}} e\big((\sum_{i,j} a'_{ij}x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i))\cdot t\big)\Big|^2 \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt\Big]^2 \\
&= \Big[\int_{\mathbf{C}} \mathbf{E}_{\mathbf{x}_U,\mathbf{x}'_U}\mathbf{E}_{\mathbf{x}_{\bar U},\mathbf{x}'_{\bar U},\mathbf{y}_{\bar U},\mathbf{y}'_{\bar U}} e\Big(\big(\sum_{i\in U,j\in\bar U} a'_{ij}x_i(x'_j - y'_j) + \sum_{i\in\bar U,j\in U} a'_{ij}(x_i - y_i)x'_j + \sum_{j\in\bar U} b_j(x_j - y_j) \\
&+ \sum_{j\in\bar U} b'_j(x'_j - y'_j) + \sum_{i\in\bar U,j\in\bar U} a'_{ij}(x_i x'_j - y_i y'_j)\big)\cdot t\Big) \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt\Big]^2 \\
&\le \int_{\mathbf{C}} \mathbf{E}_{\mathbf{x}_{\bar U},\mathbf{x}'_{\bar U},\mathbf{y}_{\bar U},\mathbf{y}'_{\bar U}}\Big|\mathbf{E}_{\mathbf{x}_U,\mathbf{x}'_U} e\Big(\big(\sum_{i\in U,j\in\bar U} a'_{ij}x_i(x'_j - y'_j) + \sum_{i\in\bar U,j\in U} a'_{ij}(x_i - y_i)x'_j) + \sum_{j\in\bar U} b_j(x_j - y_j) \\
&+ \sum_{j\in\bar U} b'_j(x'_j - y'_j) + \sum_{i\in\bar U,j\in\bar U} a'_{ij}(x_i x'_j - y_i y'_j)\big)\cdot t\Big) \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt\Big|^2 \\
&= \int_{\mathbf{C}} \mathbf{E}_{\mathbf{x}_U,\mathbf{y}_U,\mathbf{x}_{\bar U},\mathbf{y}_{\bar U},\mathbf{x}'_U,\mathbf{y}'_U,\mathbf{x}'_{\bar U},\mathbf{y}'_{\bar U}} e\Big(\big(\sum_{i\in U,j\in\bar U} a'_{ij}(x_i - y_i)(x'_j - y'_j) \\
&+ \sum_{i\in\bar U,j\in U} a'_{ij}(x_i - y_i)(x'_j - y'_j)\big)\cdot t\Big) \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt \\
&= \int_{\mathbf{C}} \mathbf{E}_{\mathbf{v},\mathbf{w}} e\Big(\big(\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\big)\cdot t\Big) \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt \\
&= (1/\sqrt{2\pi})^2 \mathbf{E}_{\mathbf{v},\mathbf{w}} \exp(-\frac{\pi}{2}|\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|^2), \qquad (66)
\end{aligned}$$

where $(\mathbf{y}_U, \mathbf{y}'_U)$ and $(\mathbf{y}_{\bar U}, \mathbf{y}'_{\bar U})$ are independent identical copies of $(\mathbf{x}_U, \mathbf{x}'_U)$ and $(\mathbf{x}_{\bar U}, \mathbf{x}'_{\bar U})$ respectively, and $\mathbf{v} := \mathbf{x} - \mathbf{y}, \mathbf{w} := \mathbf{x}' - \mathbf{y}'$.

Thus

$$\begin{aligned}
\gamma^4 &= \Big(\mathbf{P}_{\mathbf{x},\mathbf{x}'}(|\sum_{i,j} a'_{ij}x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i - a| \le 1)\Big)^4 \\
&\le \exp(4\pi)(2\pi)^4\Big(\int_{\mathbf{C}} \Big|\mathbf{E}_{\mathbf{x},\mathbf{x}'} e\big[(\sum_{i,j} a'_{ij}x_i x'_j + \sum_i b_i x_i + \sum_i b'_i x'_i))\cdot t\big]\Big| \exp(-\frac{\pi}{2}|t|^2)/(\sqrt{2\pi})^2 dt\Big)^4 \\
&\le \exp(4\pi)(2\pi)^3 \mathbf{E}_{\mathbf{v},\mathbf{w}} \exp(-\frac{\pi}{2}|\sum_{i\in U,j\in\bar U} a'_{ij}v_i w_j + \sum_{i\in\bar U,j\in U} a'_{ij}v_i w_j|^2).
\end{aligned}$$

Because $\gamma \ge n^{-B}$, the inequality above implies that

$$\mathbf{P}_{\mathbf{v},\mathbf{w}}\Big(|\sum_{i\in U,j\in\bar U} a'_{ij}v_i w_j + \sum_{i\in\bar U,j\in U} a'_{ij}v_i w_j| = O_B(\sqrt{\log n})\Big) \ge \frac{1}{2}\gamma^4/((2\pi)^3\exp(4\pi)).$$

Scaling back to a_{ij}, we thus obtain

$$\mathbf{P}_{\mathbf{v},\mathbf{w}}\Big(|\sum_{i\in U,j\in\bar U} a_{ij}v_i w_j + \sum_{i\in\bar U,j\in U} a_{ij}v_i w_j| = O_B(\beta\sqrt{\log n})\Big) \ge \frac{1}{2}\gamma^4/((2\pi)^3\exp(4\pi)),$$

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] S Chatterjee, *A generalization of the Lindeberg principle*, Ann. Probab. Volume 34, Number 6 (2006), 2061–2440.

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

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

[8] 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).

[9] L. Erdős, *Universality of Wigner random matrices: a survey of recent results*, `arxiv.org/abs/1004.0861`.

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

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

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

[13] V. L. Girko, *The strong circular law, twenty years later*, II. Random Oper. Stochastic Equations 12 (2004), no. 3, 255-312.

[14] V. L. Girko, *Elliptic law: ten years later I*, Random Operators and Stochastic Equations, V. 3, N. 3, (1995), 257-302.

[15] V. L. Girko, *The Elliptic law: ten years later II*, Random Operators and Stochastic Equations, V. 3, N.4, (1995), 377-398.

[16] V. L. Girko, *Strong elliptic law*, Random Operators and Stochastic Equations, V.5, N.3, (1997), 269-306.

[17] V. L. Girko, *The strong elliptic law, twenty years later*, I. Random Oper. Stochastic Equations 12 (2006), no. 1, 59-102.

[18] 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.

[19] F. Götze and A. N. Tikhomirov, *The circular law for random matrices*, Ann. Prob., 38 (2010), no. 4, 1444-1491.

[20] I. Goldsheid and B. A. Khoruzhenko, *The Thouless formula for random non-Hermitian Jacobi matrices*, Israel J. Math., 148 (2005), 331-346.

[21] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab.5 (2000) 119-136.

[22] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. 8 (1977), no. 3-4, 197-211.

[23] K. Johansson, *From Gumbel to Tracy-Widom*, Probab. Theory Related Fields, 138 (2007), no. 1-2, p. 75–112.

[24] 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).

[25] D. Kleitman, *On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors*, Advances in Math. 5 (1970), 155-157.

[26] M. Ledoux, *Complex Hermite polynomials: from the semi-circular law to the circular law*, Commun. Stoch. Anal. 2 (2008), no. 1, p. 27–32.

[27] J. W. Lindeberg, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, Math. Z. 15 (1922), 211-225.

[28] 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.

[29] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Acad. Press. (1967).

[30] M. L. Mehta, *Random Matrices*, third edition. Elsevier/Academic Press, Amsterdam, (2004).

[31] A. Naumov, *Elliptic law for real random matrices*, `arxiv.org/abs/1201.1639`.

[32] H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Mathematics Journal Vol. 161, 4 (2012), 545-586.

[33] H. Nguyen, *A continuous variant of the inverse Littlewood-Offord problem for quadratic forms*, to appear in Contribution to Discrete Mathematics.

[34] H. Nguyen, *On the least singular value of random symmetric matrices*, Electron. J. Probab.,17 (2012), no. 53 1-19 .

[35] H. Nguyen and V. Vu, *Optimal Littlewood-Offord theorems*, Advances in Math., Vol. 226 6 (2011), 5298-5319.

[36] G. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, Journal of Multivariate Analysis, 101 (2010), 645-656.

[37] L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, Mathematical Surveys and Monographs, American Mathematical Society, 2011.

[38] M. Rudelson and R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600-633.

[39] M. Talagrand *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. No. 81 (1995), 73-205.

[40] T. Tao, *Freiman's theorem in solvable groups*, Contribution in Discrete Mathematics, 5 (2010), no. 2, 137–184.

[41] T. Tao and V. Vu, *John-type theorems for generalized arithmetic progressions and iterated sumsets*, Advances in Mathematics, 219 (2008), no. 2, 428-449.

[42] 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.

[43] 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.

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

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

[46] 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.

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

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

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

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

[51] R. Vershynin, *Invertibility of symmetric random matrices*, to appear in Random Structures & Algorithms.

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