

# ON THE CONCENTRATION OF RANDOM MULTILINEAR FORMS AND THE UNIVERSALITY OF RANDOM BLOCK MATRICES

HOI H. NGUYEN AND SEAN O'ROURKE

ABSTRACT. The circular law asserts that if  $\mathbf{X}_n$  is a  $n \times n$  matrix with iid complex entries of mean zero and unit variance, then the empirical spectral distribution of  $\frac{1}{\sqrt{n}}\mathbf{X}_n$  converges almost surely to the uniform distribution on the unit disk as  $n$  tends to infinity. Answering a question of Tao, we prove the circular law for a general class of random block matrices with dependent entries. The proof relies on an inverse-type result for the concentration of linear operators and multilinear forms.

## 1. INTRODUCTION

The *eigenvalues* of a  $n \times n$  matrix  $\mathbf{M}$  are the roots in  $\mathbb{C}$  of the characteristic polynomial  $\det(\mathbf{M} - z\mathbf{I})$ , where  $\mathbf{I}$  is the identity matrix. We let  $\lambda_1(\mathbf{M}), \dots, \lambda_n(\mathbf{M})$  denote the eigenvalues of  $\mathbf{M}$ . In this case, the *empirical spectral measure* of  $\mathbf{M}$  is given by

$$\mu_{\mathbf{M}} := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(\mathbf{M})}.$$

The corresponding *empirical spectral distribution* (ESD) is given by

$$F^{\mathbf{M}}(x, y) := \frac{1}{n} \# \{1 \leq i \leq n : \operatorname{Re}(\lambda_i(\mathbf{M})) \leq x, \operatorname{Im}(\lambda_i(\mathbf{M})) \leq y\}.$$

Here  $\#E$  denotes the cardinality of the set  $E$ . If the matrix  $\mathbf{M}$  is Hermitian, then the eigenvalues  $\lambda_1(\mathbf{M}), \dots, \lambda_n(\mathbf{M})$  are real. In this case the ESD is given by

$$F^{\mathbf{M}}(x) := \frac{1}{n} \# \{1 \leq i \leq n : \lambda_i(\mathbf{M}) \leq x\}.$$

Given a random  $n \times n$  matrix  $\mathbf{X}_n$ , an important problem in random matrix theory is to study the limiting distribution of the empirical spectral measure as  $n$  tends to infinity. We consider one of the simplest random matrix ensembles, when the entries of  $\mathbf{X}_n$  are iid copies of the random variable  $\xi$ . We refer to  $\xi$  as the *atom variable* of  $\mathbf{X}_n$ .

When  $\xi$  is a standard complex Gaussian random variable,  $\mathbf{X}_n$  can be viewed as a random matrix drawn from the probability distribution

$$\mathbb{P}(d\mathbf{M}) = \frac{1}{\pi^{n^2}} e^{-\operatorname{tr}(\mathbf{M}\mathbf{M}^*)} d\mathbf{M}$$

on the set of complex  $n \times n$  matrices. Here  $d\mathbf{M}$  denotes the Lebesgue measure on the  $2n^2$  real entries

$$\{\operatorname{Re}(m_{ij}) : 1 \leq i, j \leq n\} \cup \{\operatorname{Im}(m_{ij}) : 1 \leq i, j \leq n\}$$

---

2000 *Mathematics Subject Classification.* 15A52, 15A63, 11B25.  
H. Nguyen is partly supported by research grant DMS-1358648.  
S. O'Rourke is supported by grant AFOSAR-FA-9550-12-1-0083.

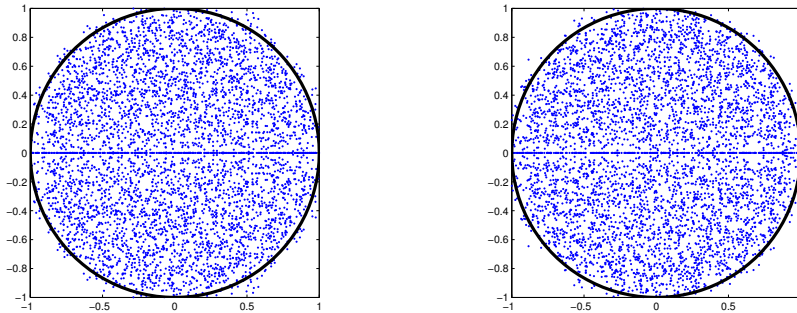


FIGURE 1. The eigenvalues of random matrices with iid entries. The first plot contains the eigenvalues of 50 samples of  $100 \times 100$  random matrices drawn from the real Ginibre ensemble. The second plot contains the eigenvalues of 50 samples of  $100 \times 100$  random matrices whose entries are Bernoulli random variables (i.e. each entry takes values  $\pm 1$  with equal probability). The black circle in each plot is the unit circle of radius one centered at the origin.

of  $\mathbf{M} = (m_{ij})_{i,j=1}^n$ . This is known as the *complex Ginibre ensemble*. The *real Ginibre ensemble* and *quaternionic Ginibre ensemble* are defined analogously.

Following Ginibre [13], one may compute the joint density of the eigenvalues of a random matrix  $\mathbf{X}_n$  drawn from the complex Ginibre ensemble. Mehta [22, 23] used this joint density function to compute the limiting spectral measure of the complex Ginibre ensemble. In particular, he showed that if  $\mathbf{X}_n$  is drawn from the complex Ginibre ensemble, then the ESD of  $\frac{1}{\sqrt{n}}\mathbf{X}_n$  converges to the *circular law*  $F_{\text{circ}}$ , where

$$F_{\text{circ}}(x, y) := \mu_{\text{circ}}(\{z \in \mathbb{C} : \text{Re}(z) \leq x, \text{Im}(z) \leq y\})$$

and  $\mu_{\text{circ}}$  is the uniform probability measure on the unit disk in the complex plane. Edelman [10] verified the same limiting distribution for the real Ginibre ensemble.

For the general (non-Gaussian) case, there is no formula for the joint distribution of the eigenvalues and the problem appears much more difficult. The universality phenomenon in random matrix theory asserts that the spectral behavior of a random matrix does not depend on the distribution of the atom variable  $\xi$  in the limit  $n \rightarrow \infty$ . In other words, one expects that the circular law describes the limiting ESD of a large class of random matrices (not just Gaussian matrices); Figure 1 presents a numerical simulation depicting this universality phenomenon.

In the 1950's, Wigner [41] proved a version of the universality phenomenon for Hermitian random matrices, now known as Wigner matrices. However, the random matrix ensemble described above is not Hermitian. In fact, many of the techniques used to deal with Hermitian random matrices do not apply to non-Hermitian matrices [3, Section 11.1].

An important result was obtained by Girko [14, 15] who related the empirical spectral measure of non-Hermitian matrices to that of Hermitian matrices. Building upon this Hermitization technique, Bai [4, 3] gave the first rigorous proof of the circular law for general (non-Gaussian) distributions under a number of moment and smoothness assumptions on the atom variable  $\xi$ . Important results were obtained more recently by Pan and Zhou [30] and Götze and Tikhomirov [17]. Tao and Vu [36] were able to prove the circular law under the assumption that  $\mathbb{E}|\xi|^{2+\varepsilon} < \infty$ , for some  $\varepsilon > 0$ . Recently, Tao and Vu [33, 38] established the law assuming only that  $\xi$  has finite variance.

For any  $m \times n$  matrix  $\mathbf{M}$ , we denote the Hilbert-Schmidt norm  $\|\mathbf{M}\|_2$  by the formula

$$\|\mathbf{M}\|_2 := \sqrt{\operatorname{tr}(\mathbf{M}\mathbf{M}^*)} = \sqrt{\operatorname{tr}(\mathbf{M}^*\mathbf{M})}. \quad (1.1)$$

**Theorem 1.1** (Tao-Vu, [38]). *Let  $\xi$  be a complex random variable with mean zero and unit variance. For each  $n \geq 1$ , let  $\mathbf{X}_n$  be a  $n \times n$  matrix whose entries are iid copies of  $\xi$ , and let  $\mathbf{N}_n$  be a  $n \times n$  deterministic matrix. If  $\operatorname{rank}(\mathbf{N}_n) = o(n)$  and  $\sup_{n \geq 1} \frac{1}{n^2} \|\mathbf{N}_n\|_2^2 < \infty$ , then the ESD of  $\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n)$  converges almost surely to the circular law  $F_{\text{circ}}$  as  $n \rightarrow \infty$ <sup>1</sup>.*

One of the key steps in proving Theorem 1.1 is controlling the largest and smallest singular values of  $\mathbf{X}_n + \mathbf{N}_n$ . We recall that the *singular values* of a  $m \times n$  matrix  $\mathbf{M}$  are the eigenvalues of  $|\mathbf{M}| := \sqrt{\mathbf{M}^*\mathbf{M}}$ . We let  $\sigma_1(\mathbf{M}) \geq \dots \geq \sigma_n(\mathbf{M}) \geq 0$  denote the singular values of  $\mathbf{M}$ . In particular, the largest and smallest singular values are given by

$$\sigma_1(\mathbf{M}) := \sup_{\|x\|=1} \|\mathbf{M}x\|, \quad \sigma_n(\mathbf{M}) := \inf_{\|x\|=1} \|\mathbf{M}x\|,$$

where  $\|v\|$  denotes the Euclidean norm of the vector  $v$ . We let  $\|\mathbf{M}\| := \sigma_1(\mathbf{M})$  denote the spectral norm of the matrix  $\mathbf{M}$ .

While the behavior of the largest singular value is well studied (e.g. see [1, 32]), bounds for the smallest singular value appear more difficult. Using techniques from additive combinatorics, Tao and Vu established the following bound on the least singular value of  $\mathbf{X}_n + \mathbf{N}_n$ .

**Theorem 1.2** (Tao-Vu, [36]). *Assume that  $\mathbf{X}_n$  is an  $n \times n$  random matrix whose entries are iid copies of a random variable with mean zero and variance one. Assume that  $\mathbf{N}_n$  is a deterministic  $n \times n$  matrix whose entries are bounded by  $n^\alpha$  in absolute value. Then for any  $B > 0$ , there exists  $A > 0$  (depending on  $B$  and  $\alpha$ ) such that*

$$\mathbb{P}(\sigma_n(\mathbf{X}_n + \mathbf{N}_n) \leq n^{-A}) = O(n^{-B}).$$

## 2. UNIVERSALITY OF RANDOM BLOCK MATRICES

The goal of this note is to study a class of random matrices that generalizes the random matrix ensemble discussed above. In particular, we consider random block matrices whose entries are not necessarily independent. We will show that, under some moment assumptions, the limiting ESD of these block matrices is also given by the circular law.

**2.1. Quaternions and matrices of quaternions.** One of the prototypical examples of a block matrix is that of a quaternionic matrix. We now review some preliminary facts on quaternions and matrices of quaternions. Most of these results can be found in the detailed survey by Zhang [42]. Let  $\mathbb{H}$  denote the non-commutative field of quaternions. As a real vectors space  $\mathbb{H}$  admits a basis  $\{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$  with the usual multiplicative table: 1 is the identity element and

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \quad \mathbf{jk} = -\mathbf{kj} = \mathbf{i}, \quad \mathbf{ki} = -\mathbf{ik} = \mathbf{j}.$$

For  $q = q_0 + \mathbf{i}q_1 + \mathbf{j}q_2 + \mathbf{k}q_3 \in \mathbb{H}$ , we have  $q^* := q_0 - q_1\mathbf{i} - q_2\mathbf{j} - q_3\mathbf{k}$ ,  $\operatorname{Re}(q) := q_0$ , and  $\operatorname{Im}(q) := q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$ . Then

$$qq^* = q_0^2 + q_1^2 + q_2^2 + q_3^2,$$

and thus any nonzero quaternion is invertible. Define the norm  $|q| := \sqrt{qq^*}$ . It follows that for any  $q, q' \in \mathbb{H}$ ,  $|qq'| = |q||q'|$ . Real numbers and complex numbers can be thought of as quaternions in the natural way, and one has  $\mathbb{R} \subset \mathbb{C} \subset \mathbb{H}$ . Every quaternion  $q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$  can be written uniquely as  $q = c_1 + c_2\mathbf{j}$  where  $c_1 = q_0 + q_1\mathbf{i}$ ,  $c_2 = q_2 + q_3\mathbf{i}$  are complex numbers.

<sup>1</sup>Here, and throughout the paper, we use asymptotic notation such as  $O, o$  under the assumption that  $n \rightarrow \infty$ . See Section 2.11 for a complete description of our asymptotic notation.

We say that two quaternions  $q, q'$  are *similar* if there exists a nonzero quaternion  $x$  such that  $q = xq'x^{-1}$ . We let  $\mathbb{S}(\mathbb{H})$  denote the group of quaternions with norm one. It follows that  $q, q'$  are similar if and only if there exists  $x \in \mathbb{S}(\mathbb{H})$  with  $q = xq'x^*$ . The following lemma shows that every quaternion is similar to a complex number.

**Lemma 2.2** ([42]). *If  $q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k} \in \mathbb{H}$ , then  $q$  and  $\operatorname{Re}(q) + |\operatorname{Im}(q)|\mathbf{i}$  are similar.*

Let  $\mathbf{M}$  be a  $n \times n$  matrix with quaternion entries. Then  $\lambda \in \mathbb{H}$  is called a *right eigenvalue* of  $\mathbf{M}$  if there exists a nonzero vector  $X \in \mathbb{H}^n$  such that  $\mathbf{M}X = X\lambda$ . If  $\lambda$  is a right eigenvalue of  $\mathbf{M}$ , one finds that  $q\lambda q^{-1}$  is also a right eigenvalue of  $\mathbf{M}$  for any nonzero quaternion  $q$ . Hence the right spectrum of  $\mathbf{M}$  is either infinite or contained in  $\mathbb{R}$ . From Lemma 2.2, we restrict our attention to complex right eigenvalues. We consider the (unique) decomposition  $\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2\mathbf{j}$ . Then for any  $\lambda \in \mathbb{C}$  and  $X = Y + Z\mathbf{j}$  with  $Y, Z \in \mathbb{C}^n$ , the following are equivalent:

- (i)  $\mathbf{M}X = X\lambda$ ,
- (ii)  $\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ -\overline{\mathbf{M}}_2 & \overline{\mathbf{M}}_1 \end{bmatrix} \begin{bmatrix} Y \\ -\overline{Z} \end{bmatrix} = \lambda \begin{bmatrix} Y \\ -\overline{Z} \end{bmatrix}$ ,
- (iii)  $\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ -\overline{\mathbf{M}}_2 & \overline{\mathbf{M}}_1 \end{bmatrix} \begin{bmatrix} Z \\ \overline{Y} \end{bmatrix} = \overline{\lambda} \begin{bmatrix} Z \\ \overline{Y} \end{bmatrix}$ .

Thus, the right spectrum of  $\mathbf{M}$ , when restricted to complex numbers, is given by the  $2n$  eigenvalues of the complex matrix

$$\begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ -\overline{\mathbf{M}}_2 & \overline{\mathbf{M}}_1 \end{bmatrix}.$$

Moreover, the complex eigenvalues appear as conjugate pairs. The whole set of right eigenvalues of  $\mathbf{M}$  is then the union of all similarity classes of the complex right eigenvalues of  $\mathbf{M}$ .

**2.3. Random quaternionic matrices.** Let  $\xi$  be a real random variable with mean zero and variance  $1/4$ . We study the right eigenvalues of random quaternion matrices whose entries are iid copies of  $\xi_0 + \xi_1\mathbf{i} + \xi_2\mathbf{j} + \xi_3\mathbf{k}$ , where  $\xi_0, \xi_1, \xi_2, \xi_3$  are iid copies of  $\xi$ . From the discussion above, we find that this is equivalent to studying the eigenvalues of random complex block matrices. Indeed, the problem reduces to studying the eigenvalues of the  $2n \times 2n$  matrix

$$\mathbf{X}_n = \begin{bmatrix} \mathbf{A}_n & \mathbf{B}_n \\ -\overline{\mathbf{B}}_n & \overline{\mathbf{A}}_n \end{bmatrix}, \quad (2.1)$$

where  $\mathbf{A}_n, \mathbf{B}_n$  are independent  $n \times n$  complex matrices whose entries are iid copies of  $\xi_0 + \xi_1\mathbf{i}$ . We note, however, that the entries of  $\mathbf{X}_n$  are not independent. Thus, Theorem 1.1 cannot be applied to the block matrix  $\mathbf{X}_n$ .

In the case that  $\xi$  is Gaussian (i.e. the quaternionic Ginibre ensemble), the circular law was established by Benaych-Georges and Chapon [5] using logarithmic potential theory. We will verify the circular law for random quaternionic matrices when the atom variable  $\xi$  is non-Gaussian.

**Theorem 2.4** (Universality for quaternion random matrices). *Let  $\xi$  be a complex random variable with mean zero and variance  $1/2$ , and suppose  $\mathbb{E}[\xi^2] = 0$  and  $\mathbb{E}|\xi|^{2+\eta} < \infty$  for some  $\eta > 0$ . For each  $n \geq 1$ , let  $\mathbf{A}_n, \mathbf{B}_n$  be independent  $n \times n$  matrices whose entries are iid copies of  $\xi$ , and let  $\mathbf{X}_n$  be the  $2n \times 2n$  matrix defined in (2.1). For each  $n \geq 1$ , let  $\mathbf{N}_n$  be a deterministic  $2n \times 2n$  matrix, and suppose the sequence  $\{\mathbf{N}_n\}_{n \geq 1}$  satisfies  $\operatorname{rank}(\mathbf{N}_n) = O(n^{1-\varepsilon})$  and  $\sup_{n \geq 1} \frac{1}{n^2} \|\mathbf{N}_n\|_2^2 < \infty$ , for some  $\varepsilon > 0$ . Then the ESD of  $\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n)$  converges almost surely to the circular law  $F_{\text{circ}}$  as  $n \rightarrow \infty$ .*

**2.5. Random block matrices.** More generally, we will study random block matrices of the form

$$\mathbf{X}_n = \begin{bmatrix} \mathbf{A}_n & \mathbf{B}_n \\ \mathbf{C}_n & \mathbf{D}_n \end{bmatrix}, \quad (2.2)$$

where  $\mathbf{A}_n = (a_{ij})_{i,j=1}^n$ ,  $\mathbf{B}_n = (b_{ij})_{i,j=1}^n$ ,  $\mathbf{C}_n = (c_{ij})_{i,j=1}^n$ ,  $\mathbf{D}_n = (d_{ij})_{i,j=1}^n$ , and

$$\{(a_{ij}, b_{ij}, c_{ij}, d_{ij}) : 1 \leq i, j \leq n\}$$

is a collection of iid copies of the random vector  $(\xi_1, \xi_2, \xi_3, \xi_4)$ . Here the random variables  $\xi_1, \xi_2, \xi_3, \xi_4$  are not required to be independent.

This ensemble of block matrices was proposed by Tao at the AIM Workshop on Random Matrices as a matrix model with dependent entries in which the circular law is still expected to hold<sup>2</sup>. We will prove the circular law for this ensemble of random block matrices under some moment assumptions on the atom variables  $\xi_1, \xi_2, \xi_3, \xi_4$ .

The matrix  $\mathbf{X}_n$  in (2.2) can be viewed as a  $2 \times 2$  block matrix. More generally, we will study  $d \times d$  block matrices for any  $d \geq 2$ . We begin with the following definition.

**Definition 2.6** (Random block matrices with dependent entries; Condition **C0**). Let  $d \geq 2$ . Let  $(\xi_{st})_{s,t=1}^d$  be a complex random matrix where each entry  $\xi_{st}$  has mean zero and variance  $1/d$ . For each  $s, t \in \{1, \dots, d\}$ , let  $\{x_{st;ij}\}_{i,j \geq 1}$  be an infinite double array of complex random variables all defined on the same probability space. For each  $n \geq 1$  and all  $s, t \in \{1, \dots, d\}$ , define the  $n \times n$  random matrix  $\mathbf{X}_{n,st} := (x_{st;ij})_{i,j=1}^n$ . Define the  $dn \times dn$  random block matrix

$$\mathbf{X}_n := \begin{bmatrix} \mathbf{X}_{n,11} & \dots & \mathbf{X}_{n,1d} \\ \vdots & \ddots & \vdots \\ \mathbf{X}_{n,d1} & \dots & \mathbf{X}_{n,dd} \end{bmatrix} = (\mathbf{X}_{n,st})_{s,t=1}^d.$$

We say the sequence of matrices  $\{\mathbf{X}_n\}_{n \geq 1}$  satisfies condition **C0** with parameter  $d$  and atom variables  $(\xi_{st})_{s,t=1}^d$  if the following conditions hold:

- (i)  $\{(x_{st;ij})_{s,t=1}^d : 1 \leq i, j\}$  is a collection of iid copies of  $(\xi_{st})_{s,t=1}^d$ ,
- (ii) We have  $\mathbb{E}[\xi_{st}\overline{\xi_{uv}}] = 0$  for all  $(s, t) \neq (u, v)$ .

In Theorem 2.7 below, we establish the circular law for a class of random block matrices that satisfy condition **C0**. In particular, Theorem 2.4 is a corollary of the following theorem in the case that  $d = 2$ .

**Theorem 2.7** (Universality for random block matrices). *Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of random matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume that*

$$\max_{1 \leq s, t \leq d} \mathbb{E}|\xi_{st}|^{2+\eta} < \infty$$

for some  $\eta > 0$ . For each  $n \geq 1$ , let  $\mathbf{N}_n$  be a deterministic  $dn \times dn$  matrix, and suppose the sequence  $\{\mathbf{N}_n\}_{n \geq 1}$  satisfies  $\text{rank}(\mathbf{N}_n) = O(n^{1-\varepsilon})$  and  $\sup_{n \geq 1} \frac{1}{n^2} \|\mathbf{N}_n\|_2^2 < \infty$  for some  $\varepsilon > 0$ . Then the ESD of  $\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n)$  converges almost surely to the circular law  $F_{\text{circ}}$  as  $n \rightarrow \infty$ .

In Definition 2.6, we require the atom variables  $(\xi_{st})_{s,t=1}^d$  to be uncorrelated. In this note, we will not deal with the correlated case. However, when there is a correlation among the atom variables,

<sup>2</sup>See <http://www.aimath.org/WWN/randommatrices/randommatrices.pdf>.

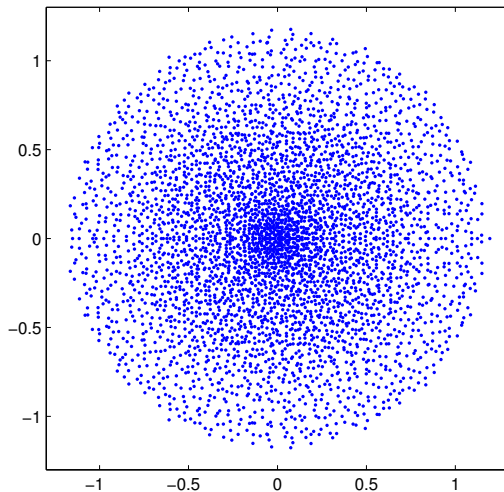


FIGURE 2. The eigenvalue plot of  $\frac{1}{\sqrt{2n}}\mathbf{X}_n$ , when  $n = 2000$ ,  $\mathbf{X}_n$  is defined in (2.3), and  $\mathbf{A}_n, \mathbf{B}_n$  are independent  $n \times n$  random matrices drawn from the real Ginibre ensemble. The eigenvalues appear to concentrate near the origin.

we do not always expect the circular law to be the limiting distribution. In Figure 2, we plot the eigenvalues of  $\frac{1}{\sqrt{2n}}\mathbf{X}_n$  in the case that

$$\mathbf{X}_n = \begin{bmatrix} \mathbf{A}_n & \mathbf{A}_n \\ \mathbf{A}_n & \mathbf{B}_n \end{bmatrix}, \quad (2.3)$$

where  $\mathbf{A}_n, \mathbf{B}_n$  are independent  $n \times n$  random matrices drawn from the real Ginibre ensemble. In particular,  $\mathbf{X}_n$  does not satisfy condition (ii) of Definition 2.6. Figure 2 predicts that more of the eigenvalues will concentrate near the origin, and so we do not believe the limiting distribution will be uniform on the unit disk.

**2.8. Least singular value bound.** One of the key ingredients in the proof of Theorem 2.7 is a bound on the least singular value of random matrices  $\{\mathbf{X}_n\}_{n \geq 1}$  that satisfy condition **C0**. In particular, we establish the following result.

**Theorem 2.9** (Least singular value bound). *Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of random matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume that*

$$\max_{1 \leq s, t \leq d} \mathbb{E}|\xi_{st}|^{2+\eta} < \infty$$

for some  $\eta > 0$ . For each  $n \geq 1$ , let  $\mathbf{N}_n$  be a deterministic  $dn \times dn$  matrix whose entries are bounded by  $n^\alpha$  for some  $\alpha > 0$ . Then, for every  $B > 0$ , there exist  $A > 0$  (depending only on  $d, B, \alpha$ ) such that

$$\mathbb{P}(\sigma_{dn}(\mathbf{X}_n + \mathbf{N}_n) \leq n^{-A}) = O(n^{-B}).$$

**2.10. Overview.** The proof of Theorem 2.9 requires studying an inverse Littlewood-Offord problem for random multilinear forms. To this end, we introduce the Littlewood-Offord problem and random multilinear forms in Section 3. Sections 4, 5, and 6 contain the proof of Theorem 2.9. Finally, we prove Theorem 2.7 in Section 7. A number of auxiliary results are contained in the appendix.

**2.11. Notation.** We use asymptotic notation (such as  $O, o, \Omega, \asymp$ ) under the assumption that  $n \rightarrow \infty$ . We use  $X \ll Y, Y \gg X, Y = \Omega(X)$ , or  $X = O(Y)$  to denote the bound  $X \leq CY$  for all sufficiently large  $n$  and for some constant  $C$ . Notations such as  $X \ll_k Y$  and  $X = O_k(Y)$  mean that the hidden constant  $C$  depends on another constant  $k$ .  $X = o(Y)$  or  $Y = \omega(X)$  means that  $X/Y \rightarrow 0$  as  $n \rightarrow \infty$ .

We let  $\|\mathbf{M}\|_2$  denote the Hilbert-Schmidt norm of  $\mathbf{M}$  (defined in (1.1)), and let  $\|\mathbf{M}\|$  denote the spectral norm of  $\mathbf{M}$ . For a vector  $\mathbf{v}$ , we let  $\|\mathbf{v}\| = \|\mathbf{v}\|_2$  denote the Euclidean norm of  $\mathbf{v}$ .

We let  $\mathbf{I}_n$  denote the  $n \times n$  identity matrix. Often we will just write  $\mathbf{I}$  for the identity matrix when the size can be deduced from the context. Similarly, we let  $\mathbf{0}$  denote the zero matrix.

For an event  $E$ , we let  $\mathbf{1}_E$  denote the indicator function of the event  $E$ . 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.

**Acknowledgments.** The authors are grateful to T. Tao and the anonymous referees for valuable comments and suggestions.

### 3. THE LITTLEWOOD-OFFORD PROBLEM AND RANDOM MULTILINEAR FORMS

In this section, we introduce the Littlewood-Offord problem and some anti-concentration results for random multilinear forms, which will be used to prove Theorem 2.9.

**3.1. The Littlewood-Offord problem.** Consider  $\xi$  a real random variable with mean zero and unit variance. A large portion of classical probability theory is devoted to studying random sums  $S_\xi(A) := \sum_{i=1}^n a_i x_i$ , where  $A = \{a_1, \dots, a_n\}$  is a multiset of complex vectors in  $\mathbb{C}^d$  and  $x_1, \dots, x_n$  are iid copies of  $\xi$ . The *Littlewood-Offord problem* is to estimate the *small ball probability*

$$\rho_{\beta, \xi}(A) := \sup_{z \in \mathbb{C}^d} \mathbb{P}(\|S_\xi(A) - z\| \leq \beta).$$

In particular, if  $\rho_{\beta, \xi}(A)$  is small, then the random sum  $S_\xi(A)$  is well spread. Conversely, if  $\rho_{\beta, \xi}(A)$  is large, then the random sum concentrates near a point.

A classical result of Littlewood and Offord [21], which was strengthened by Erdős [12], gives an estimate for the small ball probability when  $\xi$  is a Bernoulli random variable (takes values  $\pm 1$  each with probability  $1/2$ ) and  $d = 1$ .

**Theorem 3.2** (Erdős, [12]). *Let  $\xi$  be a Bernoulli random variable. If the complex numbers  $a_i$  satisfy  $|a_i| \geq 1$  for all  $i$ , then*

$$\rho_{1, \xi}(A) = O(n^{-1/2}).$$

The reader is invited to consult [28] and references therein for further extensions of this result. Motivated by inverse theorems from additive combinatorics, Tao and Vu [34] consider the following phenomenon:

If  $\rho_{\beta, \xi}(A)$  is large, then most of the elements of  $A$  are additively correlated.

In order to introduce the precise result, we recall the notion of a *generalized arithmetic progression* (GAP). A set  $Q \subset \mathbb{C}^d$  is a *GAP of rank  $r$*  if it can be expressed in the form

$$Q = \{g_0 + k_1 g_1 + \cdots + k_r g_r : K_i \leq k_i \leq K'_i, k_i \in \mathbb{Z} \text{ for all } 1 \leq i \leq r\}$$

for some  $g_0, \dots, g_r \in \mathbb{C}^d$ , and some integers  $K_1, \dots, K_r, K'_1, \dots, K'_r$ .

The vectors  $g_i$  are the *generators* of  $Q$ , the numbers  $K'_i$  and  $K_i$  are the *dimensions* of  $Q$ , and  $\text{Vol}(Q) := |B|$  is the *volume* of  $Q$ . We say that  $Q$  is *proper* if  $|Q| = \text{Vol}(Q)$ . If  $g_0 = 0$  and  $-K_i = K'_i$  for all  $i \geq 1$ , we say that  $Q$  is *symmetric*.

Consider a proper symmetric GAP  $Q = \{\sum_{i=1}^r k_i g_i : -K_i \leq k_i \leq K_i\}$  of rank  $r = O(1)$  and size  $N = n^{O(1)}$  in  $\mathbb{C}$ . Assume that  $\xi$  has Bernoulli distribution and for each  $a_i$  there exists  $q_i \in Q$  such that  $|a_i - q_i| \leq \delta$ . Then, because the random sum  $\sum_i q_i x_i$  takes values in the GAP  $nQ := \{\sum_{i=1}^r k_i g_i : -nK_i \leq k_i \leq nK_i\}$ , a GAP of size  $|nQ| \leq n^r N = n^{O(1)}$ , the pigeon-hole principle implies that  $\sum_i q_i x_i$  takes some value in  $nQ$  with probability  $n^{-O(1)}$ . Thus we have

$$\rho_{n\delta, \xi}(A) = n^{-O(1)}. \quad (3.1)$$

This example shows that if  $A$  is *close* to a *GAP* of rank  $O(1)$  and size  $n^{O(1)}$ , then  $A$  has large small ball probability. It was shown by Tao and Vu in [33, 34, 36, 37] that this is essentially the only example which has large small ball probability. We recite here an explicit version from [26] which will be used later on.

We say that a vector  $a$  is  $\delta$ -*close* to a set  $Q$  if there exists  $q \in Q$  such that  $\|a - q\| \leq \delta$ .

**Theorem 3.3** (Inverse Littlewood-Offord theorem for linear forms, [26]). *Let  $0 < \varepsilon < 1$  and  $B > 0$ . Let  $\beta > 0$  be a parameter that may depend on  $n$ . Suppose that  $\sum_i \|a_i\|^2 = 1$  and*

$$\rho := \rho_{\beta, \xi}(A) \geq n^{-B},$$

where  $x_1, \dots, x_n$  are iid copies of a random variable  $\xi$  having bounded  $(2 + \eta)$ -moment. Then, for any number  $n'$  between  $n^\varepsilon$  and  $n$ , there exists a proper symmetric GAP  $Q = \{\sum_{i=1}^r k_i g_i : |k_i| \leq K_i\}$  such that

- At least  $n - n'$  elements of  $a_i$  are  $\beta$ -close to  $Q$ .
- $Q$  has small rank,  $r = O_{B, \varepsilon}(1)$ , and small size

$$|Q| \leq \max\{O_{B, \varepsilon}\left(\frac{\rho^{-1}}{\sqrt{n'}}\right), 1\}.$$

- There is a non-zero integer  $p = O_{B, \varepsilon}(\sqrt{n'})$  such that all generators  $g_i$  of  $Q$  have the form  $g_i = (g_{i1}, \dots, g_{id})$ , where  $g_{ij} = \beta \frac{p_{ij}}{p}$  with  $p_{ij} \in \mathbb{Z}$  and  $|p_{ij}| = O_{B, \varepsilon}(\beta^{-1} \sqrt{n'})$ .

**3.4. Random multilinear forms.** One can view the sum  $S_\xi(A) = a_1 x_1 + \cdots + a_n x_n$  as a linear function of the random variables  $x_1, \dots, x_n$ . It is natural to study general polynomials of higher degree.

Let  $D$  be a fixed positive integer. Let  $x_{1i_1}, x_{2i_2}, \dots, x_{Di_D}$ ,  $1 \leq i_1, \dots, i_D \leq n$ , be iid copies of a random variable  $\xi$ , and let  $A = (a_{i_1 i_2 \dots i_D})_{1 \leq i_1, \dots, i_D \leq n}$  be an  $n^D$ -array of complex numbers. We define the  $D$ -*multilinear concentration probability* of  $A$  by

$$\rho_{\beta, \xi}(A) := \sup_{a \in \mathbb{C}, L_{D-1}} \mathbb{P}\left(\sum_{1 \leq i_1, \dots, i_D \leq n} a_{i_1 i_2 \dots i_D} x_{1i_1} x_{2i_2} \cdots x_{Di_D} + L_{D-1}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D) \in B(a, \beta)\right),$$

where  $\mathbf{x}_i = (x_{i1}, \dots, x_{in})$  and  $L_{D-1}(\mathbf{x}_1, \dots, \mathbf{x}_D)$  is any  $(D-1)$ -multilinear form of  $(\mathbf{x}_1, \dots, \mathbf{x}_D)$ .



We would like to characterize  $A$  with large  $\rho_{\beta,\xi}(A)$ . The following examples serve as good candidates.

**Example 3.5.** *In what follows  $\xi$  has Bernoulli distribution and for each  $a_{i_1 i_2 \dots i_D}$  there exists  $q_{i_1 i_2 \dots i_D}$  such that  $|a_{i_1 i_2 \dots i_D} - q_{i_1 i_2 \dots i_D}| \leq \delta$ .*

- (1) *Let  $Q$  be a proper symmetric GAP of rank  $r = O(1)$  and size  $n^{O(1)}$ . Assume that the approximated values  $q_{i_1 i_2 \dots i_D}$  belong to  $Q$ . Then, the pigeon-hole principle implies that  $\sum_{i_1, \dots, i_D} q_{i_1 i_2 \dots i_D} x_{1i_1} \dots x_{Di_D}$  takes some value in  $n^2 Q$  with probability  $n^{-O(1)}$ . Passing back to  $a_{i_1 i_2 \dots i_D}$ , we obtain  $\rho_{n^2 \delta, \xi}(A) = n^{-O(1)}$ .*
- (2) *Assume that  $q_{i_1 i_2 \dots i_D}$  can be written as  $q_{i_1 i_2 \dots i_D} = k_{i_1} b_{i_1 i_2 \dots i_D} + l_{i_2} b_{i_1 \bar{i}_2 \dots i_D} + \dots + m_{i_D} b_{i_1 i_2 \dots \bar{i}_D}$ , where  $b_{i_1 i_2 \dots i_D}, \dots, b_{i_1 i_2 \dots \bar{i}_D}$  are arbitrary sequences in  $\mathbb{R}^d$  without indices  $i_1, \dots, i_D$  respectively, and  $k_{i_1}, l_{i_2}, \dots, m_{i_D}$  are integers bounded by  $n^{O(1)}$  such that*

$$\mathbb{P}_{\mathbf{x}_1} \left( \sum_{i_1} k_{i_1} x_{1i_1} = 0 \right) = n^{-O(1)}, \dots,$$

$$\mathbb{P}_{\mathbf{x}_D} \left( \sum_{i_D} m_{i_D} x_{Di_D} = 0 \right) = n^{-O(1)}.$$

Then, as  $\sum_{i_1, i_2, \dots, i_D} q_{i_1 i_2 \dots i_D} x_{1i_1} x_{2i_2} \dots x_{Di_D}$  factors out, we have

$$\mathbb{P} \left( \sum_{i_1, i_2, \dots, i_D} q_{i_1 i_2 \dots i_D} x_{1i_1} x_{2i_2} \dots x_{Di_D} = \mathbf{0} \right) = n^{-O(1)}.$$

Passing back to  $a_{i_1 i_2 \dots i_D}$ , we hence obtain  $\rho_{n^2 \delta, \xi}(A) = n^{-O(1)}$ .

- (3) *Assume that  $q_{i_1 i_2 \dots i_D} = q'_{i_1 i_2 \dots i_D} + q''_{i_1 i_2 \dots i_D}$ , where  $q'_{i_1 i_2 \dots i_D} \in Q$ , a proper symmetric GAP of rank  $O(1)$  and size  $n^{O(1)}$ , and  $q''_{i_1 i_2 \dots i_D}$  is a sum of a few forms from (2) in such a way that the linear factors are zero with high probability. As such, we have*

$$\sup_{q \in n^2 Q} \mathbb{P}_{\mathbf{x}_1, \dots, \mathbf{x}_D} \left( \sum_{i_1, \dots, i_D} q_{i_1 i_2 \dots i_D} x_{1i_1} x_{2i_2} \dots x_{Di_D} = q \right) = n^{-O(1)}.$$

Hence we also have  $\rho_{n^2 \delta, \xi}(A) = n^{-O(1)}$  in this case.

The above examples demonstrate that if the  $a_{i_1 i_2 \dots i_D}$  can be decomposed into additive and algebraic structural parts, then  $\rho_{\xi, \beta}(A)$  is large. We conjecture that these are essentially the only cases that have large concentration probability.

**Conjecture 3.6.** *Assume that  $\rho_{\xi, \beta}(A) \geq n^{-B}$  for some generic  $\xi$  and small  $\beta$ , then most of the elements of  $A$  can be  $\beta$ -approximated by a set of  $q_{i_1 i_2 \dots i_D}$  from (3) of Example 3.5.*

Due to its nature, we believe that any justification of Conjecture 3.6 would be highly technical. In this note we prove a weak version of it as follows.

**Theorem 3.7** (Weak inverse-type theorem for multilinear forms). *Let  $0 < \varepsilon < 1$  and  $C > 0$ . Let  $\beta > 0$  be a parameter that may depend on  $n$ . Assume that*

$$\rho = \sup_{a, L_{D-1}} \mathbb{P}_{\mathbf{x}_1, \dots, \mathbf{x}_D} \left( \sum_{i_1, i_2, \dots, i_D} a_{i_1 i_2 \dots i_D} x_{1i_1} x_{2i_2} \dots x_{Di_D} - L_{D-1}(\mathbf{x}_1, \dots, \mathbf{x}_D) \in B(a, \beta) \right) \geq n^{-C},$$

where  $\mathbf{x}_1 = (x_{11}, \dots, x_{1n}), \dots, \mathbf{x}_D = (x_{D1}, \dots, x_{Dn})$ , and  $x_{1i_1}, \dots, x_{Di_D}$  are iid copies of a random variable  $\xi$  having bounded  $(2 + \eta)$ -moment. Then there exist index sets  $I_1, I_1^0$  with  $|I_1| = n - n^\varepsilon$  and  $|I_1^0| = O_{C, \varepsilon}(1)$  such that for any  $i_1 \in I_1$ , there exist index sets  $I_2, I_2^0$  depending on  $i_1$  with  $|I_2| = n - n^\varepsilon$  and  $|I_2^0| = O_{C, \varepsilon}(1)$  such that ... there exist index sets  $I_{D-1}, I_{D-1}^0$  depending on  $i_1, \dots, i_{D-2}$  with  $|I_{D-1}| = n - n^\varepsilon$  and  $|I_{D-1}^0| = O_{C, \varepsilon}(1)$  such that the following holds: for any

$i_{D-1} \in I_{D-1}$ , there exist integers  $k_{j_1 \dots j_{D-1}}$ , where each index  $j_k$  with  $1 \leq k \leq D-1$  either takes value  $i_k$  or belongs to the thin sets  $I_k^0$ , such that  $k_{j_1 \dots j_{D-1}} = n^{O_{C,d,\varepsilon}(1)}$  and  $k_{i_1, \dots, i_{D-1}} \neq 0$ , as well as

$$\mathbb{P}_{\mathbf{x}_D} \left( \left| \sum_{1 \leq i_D \leq n} \sum_{j_1 \dots j_{D-1}} k_{j_1 \dots j_{D-1}} a_{j_1 \dots j_{D-1} i_D} x_{i_D} \right| \leq \beta n^{O_{C,\varepsilon}(1)} \right) \geq n^{-O_{C,\varepsilon}(1)}.$$

Notice that while in Example 3.5 most of the  $D-1$  dimensional arrays of  $A$  have structure, Theorem 3.7 just asserts that most of the 1-dimension arrays  $a_{j_1 \dots j_{D-1} i_D}$ ,  $1 \leq i_D \leq n$ , with fixed  $j_1, \dots, j_{D-1}$ , have structure.

For the rest of this section we give a proof of Theorem 3.7. Our argument heavily relies on the following simple fact about generalized arithmetic progressions of small rank.

**Fact 3.8.** *Assume that  $q_1, \dots, q_{r+1}$  are elements of a GAP of rank  $r$  and of cardinality  $n^C$ , then there exists integer coefficients  $\alpha_1, \dots, \alpha_r$  with  $|\alpha_i| \leq n^{rC}$  and such that*

$$\sum_i \alpha_i q_i = 0.$$

**3.9. Proof of Theorem 3.7.** Without loss of generality, we assume that  $\xi$  has discrete distribution. The continuous case can be easily extended by a standard limiting argument. We begin by applying Theorem 3.3.

**Lemma 3.10.** *Let  $\varepsilon < 1$  and  $C$  be positive constants. Assume that  $\rho_{\xi, \beta}(A) = \rho \geq n^{-C}$ . Then the following holds with probability at least  $\frac{3\rho}{4}$  with respect to  $\mathbf{x}_2, \dots, \mathbf{x}_D$ . There exist a proper symmetric GAP  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$  of rank  $O_{C,\varepsilon}(1)$  and size  $O_{C,\varepsilon}(1/\rho)$  and a set  $I_{\mathbf{x}_2, \dots, \mathbf{x}_D}$  of  $n - n^\varepsilon$  indices such that for each  $i \in I_{\mathbf{x}_2, \dots, \mathbf{x}_D}$ , there exists  $q_i \in Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$  so that*

$$\left| \sum_{i_2, \dots, i_D} a_{i i_2 \dots i_D} x_{2i_2} \dots x_{Di_D} - q_i \right| \leq \beta.$$

*Proof.* (of Lemma 3.10) For short we write

$$\sum_{i_1, i_2, \dots, i_D} a_{i_1 i_2 \dots i_D} x_{1i_1} x_{2i_2} \dots x_{Di_D} = \sum_{i=1}^n x_{1i} B_i(\mathbf{x}_2, \dots, \mathbf{x}_D),$$

where

$$B_i(\mathbf{x}_2, \dots, \mathbf{x}_D) := \sum_{i_2, \dots, i_D} a_{i i_2 \dots i_D} x_{2i_2} \dots x_{Di_D}.$$

We call a vector tuple  $(\mathbf{x}_2, \dots, \mathbf{x}_D)$  *good* if

$$\mathbb{P}_{\mathbf{x}_1} \left( \left| \sum_{i=1}^n x_{1i} B_i(\mathbf{x}_2, \dots, \mathbf{x}_D) \right| \in B(a, \beta) \right) \geq \rho/4.$$

We call  $\mathbf{x}_2, \dots, \mathbf{x}_D$  *bad* otherwise. Let  $G$  be the collection of good tuples.

First, we estimate the probability  $p$  of randomly chosen vectors  $(\mathbf{x}_2, \dots, \mathbf{x}_D)$  being bad by an averaging method.

$$\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} \mathbb{P}_{\mathbf{x}_1} \left( \sum_{i=1}^n x_{1i} B_i(\mathbf{x}_2, \dots, \mathbf{x}_D) \in B(a, \beta) \right) = \rho$$

$$p\rho/4 + 1 - p \geq \rho.$$

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

Thus, the probability of randomly chosen vectors  $(\mathbf{x}_2, \dots, \mathbf{x}_D)$  being good is at least

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

Next, we consider good vectors  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G$ . By definition, we have

$$\mathbb{P}_{\mathbf{x}_1} \left( \sum_{i=1}^n x_{1i} B_i(\mathbf{x}_2, \dots, \mathbf{x}_D) \in B(a, \beta) \right) \geq \rho/4.$$

Observe that this is a high concentration of a linear form of  $x_{1i}$ . A direct application of Theorem 3.3 to the sequence  $B_i(\mathbf{x}_2, \dots, \mathbf{x}_D)$ ,  $i = 1, \dots, n$  yields the desired result.  $\square$

By a useful property of GAP containment (see for instance [35, Section 8] and [25, Theorem 6.1]), we may assume that the  $q_i(\mathbf{x}_2, \dots, \mathbf{x}_D)$  span  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$ . From now on we fix such a  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$  for each  $\mathbf{x}_2, \dots, \mathbf{x}_D$ . Let  $G$  be the collection of good vectors  $(\mathbf{x}_2, \dots, \mathbf{x}_D)$ . Thus,

$$\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} ((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G) \geq 3\rho/4. \quad (3.2)$$

Now we state our crucial lemma for the proof of Theorem 3.7.

**Lemma 3.11.** *There exists an index set  $I$  of size at least  $n - 2n^\varepsilon$ , an index set  $I_0$  of size  $O_{C, \varepsilon}(1)$ , and an integer  $k \neq 0$  with  $|k| \leq n^{O_{C, \varepsilon}(1)}$  such that for any index  $i$  from  $I$ , there are numbers  $k_{ii_0} \in \mathbb{Z}$ ,  $i_0 \in I_0$ , all bounded by  $n^{O_{C, \varepsilon}(1)}$ , such that*

$$\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} \left( kB_i(\mathbf{x}_2, \dots, \mathbf{x}_D) + \sum_{i_0 \in I_0} k_{ii_0} B_{i_0}(\mathbf{x}_2, \dots, \mathbf{x}_D) \in B(a, \beta) \right) = \rho/n^{O_{C, \varepsilon}(1)}.$$

*Proof.* (of Lemma 3.11) For each  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G$ , we choose from  $I_{\mathbf{x}_2, \dots, \mathbf{x}_D}$   $s$  indices  $i_{(1, \mathbf{x}_2, \dots, \mathbf{x}_D)}, \dots, i_{(s, \mathbf{x}_2, \dots, \mathbf{x}_D)}$  such that  $q_{i_{(j, \mathbf{x}_2, \dots, \mathbf{x}_D)}}(\mathbf{x}_2, \dots, \mathbf{x}_D)$  span  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$ , where  $s$  is the rank of  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$ . We note that  $s = O_{C, \varepsilon}(1)$  for all  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G$ .

Consider the tuples  $(i_{(1, \mathbf{x}_2, \dots, \mathbf{x}_D)}, \dots, i_{(s, \mathbf{x}_2, \dots, \mathbf{x}_D)})$  for all  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G$ . Because there are  $\sum_s O_{C, \varepsilon, \mu}(n^s) = n^{O_{C, \varepsilon, \mu}(1)}$  possibilities these tuples can take, there exists a tuple, say  $(1, \dots, r)$  (by rearranging the rows of  $A$  if needed), such that  $(i_{(1, \mathbf{x}_2, \dots, \mathbf{x}_D)}, \dots, i_{(s, \mathbf{x}_2, \dots, \mathbf{x}_D)}) = (1, \dots, r)$  for all  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G'$ , where  $G'$  is a subset of  $G$  satisfying

$$\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} ((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G') \geq \mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} ((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G) / n^{O_{C, \varepsilon}(1)} = \rho/n^{O_{C, \varepsilon}(1)}. \quad (3.3)$$

For each  $1 \leq i \leq r$ , we express  $q_i(\mathbf{x}_2, \dots, \mathbf{x}_D)$  in terms of the generators of  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$  for each  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G'$ ,

$$q_i(\mathbf{x}_2, \dots, \mathbf{x}_D) = c_{i1}(\mathbf{x}_2, \dots, \mathbf{x}_D)g_1(\mathbf{x}_2, \dots, \mathbf{x}_D) + \dots + c_{ir}(\mathbf{x}_2, \dots, \mathbf{x}_D)g_r(\mathbf{x}_2, \dots, \mathbf{x}_D),$$

where  $c_{i1}(\mathbf{x}_2, \dots, \mathbf{x}_D), \dots, c_{ir}(\mathbf{x}_2, \dots, \mathbf{x}_D)$  are integers bounded by  $n^{O_{C, \varepsilon}(1)}$ , and  $g_i(\mathbf{x}_2, \dots, \mathbf{x}_D)$  are the generators of  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$ .

We show that there are many  $(\mathbf{x}_2, \dots, \mathbf{x}_D)$  that correspond to the same coefficients  $c_{i_1 i_2}$ .

**Claim 3.12.** *There exists a (“dense”) subset  $G'' \subset G'$  such that the following holds*

- $\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} ((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G'') \geq \mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} ((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G') / n^{O_{C, \varepsilon}(1)} \geq \rho/n^{O_{C, \varepsilon}(1)}$ ;
- (common tuples) there exist  $r$  tuples  $(c_{11}, \dots, c_{1r}), \dots, (c_{r1}, \dots, c_{rr})$ , whose components are integers bounded by  $n^{O_{C, \varepsilon, \mu}(1)}$ , such that the following hold for all  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G''$ :
  - (1)  $q_i(\mathbf{x}_2, \dots, \mathbf{x}_D) = c_{i1}g_1(\mathbf{x}_2, \dots, \mathbf{x}_D) + \dots + c_{ir}g_r(\mathbf{x}_2, \dots, \mathbf{x}_D)$ , for  $i = 1, \dots, r$ ;
  - (2) The vectors  $(c_{11}, \dots, c_{1r}), \dots, (c_{r1}, \dots, c_{rr})$  span  $\mathbb{Z}^{\text{rank}(Q_{\mathbf{x}_2, \dots, \mathbf{x}_D})}$ .

*Proof.* (of Claim 3.12) Consider the collection  $\mathcal{C}$  of the coefficient-tuples

$$\mathcal{C} := \left\{ \left( (c_{11}(\mathbf{x}_2, \dots, \mathbf{x}_D), \dots, c_{1r}(\mathbf{x}_2, \dots, \mathbf{x}_D)); \dots; (c_{r1}(\dots), \dots, c_{rr}(\dots)) \right), (\mathbf{x}_2, \dots, \mathbf{x}_D) \in G' \right\}.$$

Because the number of possibilities these tuples can take is at most  $(n^{O_{C,\varepsilon}(1)})^{r^2} = n^{O_{C,\varepsilon}(1)}$ , by the pigeon-hole principle there exists a coefficient-tuple, say  $((c_{11}, \dots, c_{1r}), \dots, (c_{r1}, \dots, c_{rr})) \in \mathcal{C}$ , such that

$$\begin{aligned} & \left( (c_{11}(\mathbf{x}_2, \dots, \mathbf{x}_D), \dots, c_{1r}(\mathbf{x}_2, \dots, \mathbf{x}_D)); \dots; (c_{r1}(\mathbf{x}_2, \dots, \mathbf{x}_D), \dots, c_{rr}(\mathbf{x}_2, \dots, \mathbf{x}_D)) \right) \\ &= \left( (c_{11}, \dots, c_{1r}), \dots, (c_{r1}, \dots, c_{rr}) \right) \end{aligned}$$

for all  $(\mathbf{x}_2, \dots, \mathbf{x}_D)$  from a subset  $G''$  of  $G'$  which satisfies

$$\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D}((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G'') \geq \mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D}((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G') / n^{O_{C,\varepsilon}(1)} \geq \rho / n^{O_{C,\varepsilon}(1)}. \quad (3.4)$$

□

Now we focus on the elements of  $G''$ . Because  $|I_{\mathbf{x}_2, \dots, \mathbf{x}_D}| \geq n - n^\varepsilon$  for each  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G''$ , we obtain the following.

**Claim 3.13.** *There is a set  $I$  of size  $n - 3n^\varepsilon$  such that  $I \cap \{1, \dots, r\} = \emptyset$  and for each  $i \in I$  we have*

$$\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D}(i \in I_{\mathbf{x}_2, \dots, \mathbf{x}_D}, (\mathbf{x}_2, \dots, \mathbf{x}_D) \in G'') \geq \mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D}((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G'') / 2. \quad (3.5)$$

*Proof.* (of Claim 3.13) The result easily follows by an elementary averaging argument, □

**Lemma 3.11: proof conclusion.** Now we fix an arbitrary index  $i$  from  $I$ . We concentrate on those  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G''$  where the index  $i$  belongs to  $I_{\mathbf{x}_2, \dots, \mathbf{x}_D}$ . Because  $q_i(\mathbf{x}_2, \dots, \mathbf{x}_D) \in Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$ , we can write

$$q_i(\mathbf{x}_2, \dots, \mathbf{x}_D) = c_1(\mathbf{x}_2, \dots, \mathbf{x}_D)g_1(\mathbf{x}_2, \dots, \mathbf{x}_D) + \dots + c_r(\mathbf{x}_2, \dots, \mathbf{x}_D)g_r(\mathbf{x}_2, \dots, \mathbf{x}_D),$$

where  $c_1(\mathbf{x}_2, \dots, \mathbf{x}_D), \dots, c_r(\mathbf{x}_2, \dots, \mathbf{x}_D)$  are integers bounded by  $n^{O_{C,\varepsilon}(1)}$ .

For short, we denote by  $\mathbf{v}_{i, \mathbf{x}_2, \dots, \mathbf{x}_D}$  the vector  $(c_1(\mathbf{x}_2, \dots, \mathbf{x}_D), \dots, c_r(\mathbf{x}_2, \dots, \mathbf{x}_D))$ , we also use the shorthand  $\mathbf{v}_j$  for the vectors  $(c_{j1}, \dots, c_{jr})$  obtained from Claim 3.12.

Because  $Q_{\mathbf{x}_2, \dots, \mathbf{x}_D}$  is spanned by  $q_1(\mathbf{x}_2, \dots, \mathbf{x}_D), \dots, q_r(\mathbf{x}_2, \dots, \mathbf{x}_D)$ , we must have  $k := \det(\mathbf{v}_1, \dots, \mathbf{v}_r) \neq 0$  and that

$$\begin{aligned} & kq_i(\mathbf{x}_2, \dots, \mathbf{x}_D) + \det(\mathbf{v}_{i,y}, \mathbf{v}_2, \dots, \mathbf{v}_r)q_1(\mathbf{x}_2, \dots, \mathbf{x}_D) + \dots \\ & + \det(\mathbf{v}_{i,y}, \mathbf{v}_1, \dots, \mathbf{v}_{r-1})q_r(\mathbf{x}_2, \dots, \mathbf{x}_D) = 0. \end{aligned}$$

Furthermore, because each coefficient of the identity above is bounded by  $n^{O_{C,\varepsilon,\mu}(1)}$ , there exists a subset  $G''_i$  of  $G''$  such that all  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G''_i$  correspond to the same identity, and

$$\begin{aligned} \mathbb{P}_{(\mathbf{x}_2, \dots, \mathbf{x}_D)}((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G''_i) &\geq (\mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D}((\mathbf{x}_2, \dots, \mathbf{x}_D) \in G'') / 2) / (n^{O_{C,\varepsilon}(1)})^r \\ &\geq \rho / n^{O_{C,\varepsilon}(1)}. \end{aligned}$$

In other words, there exist integers  $k_1, \dots, k_r$ , all bounded by  $n^{O_{C,\varepsilon}(1)}$ , such that

$$kq_i(\mathbf{x}_2, \dots, \mathbf{x}_D) + k_1q_1(\mathbf{x}_2, \dots, \mathbf{x}_D) + \dots + k_rq_r(\mathbf{x}_2, \dots, \mathbf{x}_D) = 0$$

for all  $(\mathbf{x}_2, \dots, \mathbf{x}_D) \in G''_i$ .

Note that  $k$  is independent of the choice of  $i$  and  $(\mathbf{x}_2, \dots, \mathbf{x}_D)$ . By passing from  $q_i$  to  $B_i$  by approximation, we thus complete the proof of Lemma 3.11. □

We are now ready to complete the proof of our inverse result.

**Theorem 3.7: proof conclusion.** From Lemma 3.11, for any fixed  $i \in I$ , we consider the following  $(D - 1)$ -multilinear form

$$\begin{aligned} B'_i(\mathbf{x}_2, \dots, \mathbf{x}_D) &:= kB_i(\mathbf{x}_2, \dots, \mathbf{x}_D) + \sum_{i_0 \in I_0} k_{ii_0} B_{i_0}(\mathbf{x}_2, \dots, \mathbf{x}_D) \\ &= (ka_{ii_2 \dots i_D} + \sum_{i_0} k_{ii_0} a_{i_0 i_2 \dots i_D}) x_{2i_2} \dots x_{Di_D} \\ &:= \sum_{i_2, \dots, i_D} b'_{i_2 \dots i_D} x_{2i_2} \dots x_{Di_D}. \end{aligned}$$

By the conclusion of Lemma 3.11, we have  $\sup_a \mathbb{P}_{\mathbf{x}_2, \dots, \mathbf{x}_D} (B'_i(\mathbf{x}_2, \dots, \mathbf{x}_D) \in B(a, \beta)) \geq \rho/n^{O_{C, \varepsilon}(1)}$ . Thus Lemma 3.11 is applicable again for this new  $(D - 1)$ -multilinear form. By iterating the process  $D - 1$  times, we obtain the conclusion of Theorem 3.7.

#### 4. SINGULARITY OF BLOCK MATRICES: THE APPROACH TO PROVE THEOREM 2.9

As the singular values do not change under row and column permutations, for the sake of convenience, we will restrict our analysis to matrices of the form  $\mathbf{M}_n = \mathbf{X}_n + \mathbf{N}_n$ , where  $\mathbf{N}_n$  is any deterministic matrix of polynomially bounded norm and  $\mathbf{X}_n$  is a  $dn \times dn$  matrix whose  $ij$ -th block takes the form

$$\begin{pmatrix} x_{11;ij} & x_{12;ij} & \dots & x_{1d;ij} \\ \dots & \dots & \dots & \dots \\ x_{d1;ij} & x_{d2;ij} & \dots & x_{dd;ij} \end{pmatrix},$$

where  $(x_{11;ij}, \dots, x_{dd;ij}), 1 \leq i, j \leq n$ , are iid copies of  $(\xi_{11}, \dots, \xi_{dd})$  which satisfy the following conditions from Definition 2.6 and Theorem 2.9:

$$\begin{aligned} \mathbb{E}\xi_{st} &= 0, \mathbb{E}|\xi_{st}|^2 = 1, \mathbb{E}|\xi_{st}|^{2+\eta} < \infty \text{ for some } \eta > 0 \\ \mathbb{E}[\xi_{st}\overline{\xi_{uv}}] &= 0 \text{ for all } (s, t) \neq (u, v). \end{aligned} \tag{4.1}$$

We now restate Theorem 2.9 as follows.

**Theorem 4.1.** *For any  $B > 0$ , there exists  $A > 0$  depending on  $B$  and  $\alpha$  such that*

$$\mathbb{P}(\sigma_{dn}(\mathbf{M}_n) \leq n^{-A}) \leq n^{-B}.$$

In the sequel we sketch the proof of Theorem 4.1. In general, our approach will resemble that of [24, 25, 31, 36, 40] where the main ingredient is an inverse-type argument. However, as our matrix now consists of large blocks of correlated entries, we need to elaborate more on the algebraic and technical side. For the sake of simplicity, we will prove our result under the following condition.

**Assumption 4.2.** *With probability one,  $|x_{11;ij}| \leq n^{B+1} \wedge \dots \wedge |x_{dd;ij}| \leq n^{B+1}$  for all  $i, j$ .*

In what follows we assume that  $\mathbf{M}_n$  has full rank. This is the main case to consider as most random matrices are non-singular with very high probability. The case that  $\mathbf{M}_n$  is singular can be deduced by a standard argument (see for instance [27, Appendix A]).

Assume that  $\sigma_{nd}(\mathbf{M}_n) \leq n^{-A}$ . Thus  $\mathbf{M}_n \mathbf{x} = \mathbf{y}$  for some  $\|\mathbf{x}\| = 1$  and  $\|\mathbf{y}\| \leq n^{-A}$ . Let  $\mathbf{C} = (c_{i,j}(\mathbf{M}_n)), 1 \leq i, j \leq dn$ , be the matrix of cofactors of  $\mathbf{M}_n$ . By definition,  $\mathbf{C}\mathbf{y} = \det(\mathbf{M}_n) \cdot \mathbf{x}$ , and thus we have  $\|\mathbf{C}\mathbf{y}\| = |\det(\mathbf{M}_n)|$ .

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

$$|c_{1,1}(\mathbf{M}_n)y_1 + \dots + c_{1,dn}(\mathbf{M}_n)y_{dn}| \geq |\det(\mathbf{M}_n)|/(dn)^{1/2}. \quad (4.2)$$

**Claim 4.3.** *Let  $\mathbf{M}_{n-1}$  be the matrix obtained from  $\mathbf{M}_n$  by removing its first  $d$  rows, and  $c_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1}), 1 \leq i_1, \dots, i_d \leq nd$  be the sign determinant of the minor obtained from  $\mathbf{M}_{n-1}$  by removing its  $i_1, \dots, i_d$ -th columns. We have*

$$\sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} |c_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1})|^2 \geq n^{2A-O(B+\alpha)} |\det(\mathbf{M}_n)|^2 \quad (4.3)$$

*Proof.* (of Claim 4.3) As  $\|\mathbf{y}\| \leq n^{-A}$ , it follows from (4.2) that

$$\sum_{i_1=1}^{dn} |c_{1, i_1}(\mathbf{M}_n)|^2 \geq n^{2A-2} |\det(\mathbf{M}_n)|^2. \quad (4.4)$$

Next, as each cofactor  $c_{1, i_1}(\mathbf{M}_n)$ , as a sign determinant of a  $(dn-1) \times (dn-1)$  block of  $\mathbf{M}_n$ , can be expressed as

$$c_{1, i_1}(\mathbf{M}_n) = (-1)^{(1+i_1)\dots+(d+i_d)} \sum_{1 \leq i_2, \dots, i_d \leq dn} m_{2i_2} \dots m_{di_d} c_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1}).$$

The claim then follows by applying Cauchy-Schwarz inequality together with Condition 4.2 and the upper bound  $n^\alpha$  on the entries of  $\mathbf{N}_n$ .  $\square$

By Claim 4.3, in order to prove Theorem 4.1 it suffices to justify the following result.

**Theorem 4.4.** *For any  $B > 0$ , there exists  $A > 0$  such that*

$$\mathbb{P}\left(\left(\sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} |c_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1})|^2\right)^{1/2} \geq n^A |\det(\mathbf{M}_n)|\right) \leq n^{-B}.$$

Next, express  $\det(\mathbf{M}_n)$  as a  $d$ -multilinear form of its first  $d$  rows

$$\det(\mathbf{M}_n) = \sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} c_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1}) m_{1i_1} \dots m_{di_d}.$$

With  $c := (\sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} |c_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1})|^2)^{1/2}$  and  $a_{i_1 \dots i_d} := c_{i_1 \dots i_d}(\mathbf{M}_{n-1})/c$ ,

$$\frac{1}{c} \det(\mathbf{M}_n) = \sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} a_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1}) m_{1i_1} \dots m_{di_d}. \quad (4.5)$$

Heuristically, conditioning on  $\mathbf{M}_{n-1}$ , the  $d$ -multilinear form in the RHS of (4.5) is comparable to 1 in absolute value with probability extremely close to one. Thus the assumption  $\mathbb{P}(|\det(\mathbf{M}_n)|/c \leq n^{-A}) \geq n^{-B}$  of Theorem 4.4, with an appropriately large value  $A$ , must yield a high cancellation of the multilinear form. Based on this observation, our rough approach will consist of two main steps.

- *Step 1.* Assume that for an appropriately large value  $A > 0$  we have

$$\mathbb{P}_{x_{11}, \dots, x_{dn}} \left( \left| \sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} a_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1}) m_{1i_1} \dots m_{di_d} \right| \leq n^{-A} |\mathbf{M}_{n-1}| \right) \geq n^{-B}.$$

Then the normalized cofactors  $a_{i_1 \dots i_d}$  of  $\mathbf{M}_{n-1}$  must satisfy a very special property.

- *Step 2.* The probability, with respect to  $\mathbf{M}_{n-1}$ , that the  $a_{i_1 \dots i_d}$  satisfy this special property is negligible.

Although the setting of Step 1 is identical to our inverse problem discussed in Section 3, the dependencies of the entries make the problem substantially harder. We will remove these dependencies using a series of decoupling tricks to arrive at a conclusion as useful as Theorem 3.7.

**Theorem 4.5** (Step 1). *Let  $0 < \varepsilon < 1$  be given constant. Assume that*

$$\sup_a \mathbb{P}_{x_{11;11}, \dots, x_{dd;1n}} \left( \left| \sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} a_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1}) m_{1i_1} \dots m_{di_d} - a \right| \leq n^{-A} \right) \geq n^{-B}.$$

for some sufficiently large integer  $A$ , where  $a_{i,j} = c_{i,j}(\mathbf{M}_{n-1})/c$ . Then there exists  $k = O(d)$  indices  $i_1 < \dots < i_k$  and a complex vector  $\mathbf{u} = (u_1, \dots, u_{nd})$  which satisfies the following properties.

- (orthogonality)  $\|\mathbf{u}\|_2 \asymp 1$  and  $|\langle \mathbf{u}_1, \mathbf{r}_i^{(1)}(\mathbf{M}_{n-1}) \rangle + \langle \mathbf{u}_2, \mathbf{r}_i^{(2)}(\mathbf{M}_{n-1}) \rangle| \leq n^{-A/2 + O_{B,\varepsilon}(1)}$  for  $n - O_{B,\varepsilon}(1)$  rows  $\mathbf{r}_i$  of  $\mathbf{M}_{n-1}$ , where  $\mathbf{u}_1$  and  $\mathbf{r}_i^{(1)}$  are the subvectors corresponding to the components indexed by  $i_1, \dots, i_k$  of  $\mathbf{u}$  and  $\mathbf{r}_i$  respectively, and  $\mathbf{u}_2$  and  $\mathbf{r}_i^{(2)}$  are the subvectors corresponding to the remaining components of  $\mathbf{u}$  and  $\mathbf{r}_i$  respectively;
- (additive structure) there exists a generalized arithmetic progression  $Q$  of rank  $O_{B,\varepsilon}(1)$  and size  $n^{O_{B,\varepsilon}(1)}$  that contains at least  $dn - 2n^\varepsilon$  components  $u_i$ ;
- (controlled form) all the components  $u_i$ , and all the generators of the generalized arithmetic progression are rational complex numbers of the form  $\frac{p}{q} + \sqrt{-1} \frac{p'}{q'}$ , where  $|p|, |q|, |p'|, |q'| \leq n^{A/2 + O_{B,\varepsilon}(1)}$ .

In the second step, we show that the probability that  $\mathbf{M}_{n-1}$  has the above properties is negligible.

**Theorem 4.6** (Step 2). *With respect to  $\mathbf{M}_{n-1}$ , the probability that there exists a vector  $\mathbf{u}$  as in Theorem 4.5 is  $\exp(-\Omega(n))$ .*

## 5. SINGULARITY OF BLOCK MATRICES: PROOF THEOREM 4.5

Recall that in the inverse step, Theorem 4.5, we assumed a high concentration of a multilinear form on a small ball of radius  $n^{-A}$ . As the entries in each block are dependent, we are not able to apply Theorem 3.7 yet. In what follows we present two main steps to remove these dependencies.

**5.1. Dependency removal I: general linear forms.** First, it will be useful to study the concentration of the linear form

$$\sum_{1 \leq i \leq n} (a_{11;i} x_{11;i} + \dots + a_{dd;i} x_{dd;i}),$$

where  $(x_{11;i}, \dots, x_{dd;i})$  are iid copies of  $(x_{11}, \dots, x_{dd})$  satisfying (4.1). Intuitively, as the covariance of  $(x_{11}, \dots, x_{dd})$  is non-singular, the random variables  $(x_{11;i}, \dots, x_{dd;i})$  are not totally dependent on each other. (See Appendix A for a more precise statement.) This fact may suggest a way to apply Theorem 3.3 with respect to  $(x_{11;1}, \dots, x_{11;n})$  while holding  $x_{12;1}, \dots, x_{dd;n}$  fixed and vice versa. In what follows  $(x_{1;i}, \dots, x_{D;i})$  plays the role of  $(x_{11;i}, \dots, x_{dd;i})$ .

**Theorem 5.2** (Inverse Littlewood-Offord theorem for mixing linear forms). *Let  $0 < \varepsilon < 1, B > 0$  be given, and  $D$  be a positive integer. Let  $\beta > 0$  be an arbitrary real number that may depend on  $n$ . Suppose that  $a_{1;i}, \dots, a_{D;i} \in \mathbb{C}$  such that  $\sum_{i=1}^n \sum_{1 \leq j \leq D} |a_{j;i}|^2 = 1$  and*

$$\sup_a \mathbb{P}_{x_{1;1}, \dots, x_{D;n}} \left( \left| \sum_{i=1}^n (a_{1;i} x_{1;i} + \dots + a_{D;i} x_{D;i}) - a \right| \leq \beta \right) = \gamma \geq n^{-B},$$

where  $(x_{1;i}, \dots, x_{D;i}), 1 \leq i \leq n$  are iid copies of  $(x_1, \dots, x_D)$  from (4.1). Then there exist positive constants  $\alpha, c_0, C_0$  depending only on the distribution of  $(x_1, \dots, x_D)$  and  $D$  tuples  $(\eta_{k1}, \eta_{k2}, \dots, \eta_{kD}), 1 \leq k \leq D$  of complex numbers such that

- $|\eta_{ij}|$  are bounded from below and above by  $c_0$  and  $C_0$  respectively,
- The least singular value of the matrix  $(\eta_{ij})$  is at least  $\alpha$ ,
- for any number  $n'$  between  $n^\varepsilon$  and  $n$ , there exists a proper symmetric GAP  $Q = \{\sum_{i=1}^r k_i g_i : k_i \in \mathbb{Z}, |k_i| \leq L_i\} \subset \mathbb{C}$  whose parameters satisfy (i) and (ii) of Theorem 3.3 and for at least  $n - n'$  indices  $i$ , we have  $\eta_{k_1 a_{1;i}} + \dots + \eta_{k_D c_i a_{D;i}}, 1 \leq k \leq D$  are  $\beta$ -close to  $Q$ .

As Theorem 5.2 can be shown by using the method of [26], we skip its proof and refer the reader to Appendix A for a proof of a somewhat more general result (Theorem 6.1 below). We now introduce a useful corollary.

**Corollary 5.3.** *Assume as in Theorem 5.2. Then there exists a proper symmetric GAP  $Q = \{\sum_{i=1}^r k_i g_i : k_i \in \mathbb{Z}, |k_i| \leq L_i\} \subset \mathbb{C}$  whose parameters satisfy (i) and (ii) of Theorem 5.2 and an index set  $I$  of size at least  $n - n'$  such that with  $(\gamma_{ij})$  being the inverse matrix of  $(\eta_{ij})$ : the sequence  $a_{k;i}, i \in I, 1 \leq k \leq D$  are  $O(\beta)$ -close to the GAP  $P = P_1 + \dots + P_D$ , where  $P_k = \gamma_{k1} \cdot Q + \gamma_{k2} \cdot Q + \dots + \gamma_{kD} \cdot Q$ .*

**5.4. Dependency removal II: decoupling.** We now work with the multilinear form appearing in Theorem 4.5. Our goal is to show the following.

**Theorem 5.5.** *Let  $0 < \varepsilon < 1$  and  $B > 0$ . Let  $\beta > 0$  be a parameter that may depend on  $n$ . Assume that*

$$\mathbb{P}_{x_{11;11}, \dots, x_{dd;1n}} \left( \left| \sum_{1 \leq i_1, i_2, \dots, i_d \leq dn} a_{i_1 i_2 \dots i_d}(\mathbf{M}_{n-1}) m_{1i_1} \dots m_{di_d} \right| \leq n^{-A} |\mathbf{M}_{n-1}| \right) \geq n^{-B}.$$

*Then there exist index sets  $I_1, I_1^0$  with  $|I_1| = dn - n^\varepsilon$  and  $|I_1^0| = O_{C,\varepsilon}(1)$  such that for any  $i_1 \in I_1$ , there exist index sets  $I_2, I_2^0$  depending on  $i_1$  with  $|I_2| = dn - n^\varepsilon$  and  $|I_2^0| = O_{C,\varepsilon}(1)$  such that  $\dots$  there exist index sets  $I_{d-1}, I_{d-1}^0$  depending on  $i_1, \dots, i_{d-2}$  with  $|I_{d-1}| = n - n^\varepsilon$  and  $|I_{d-1}^0| = O_{C,\varepsilon}(1)$  such that the following holds: for any  $i_{d-1} \in I_{d-1}$ , there exist integers  $k_{j_1 \dots j_{d-1}}$ , where each index  $j_k$  with  $1 \leq k \leq d-1$  either takes value  $i_k$  or belongs to the thin sets  $I_k^0$ , such that  $k_{j_1 \dots j_{d-1}} = n^{O_{C,d,\varepsilon}(1)}$  and  $k_{i_1, \dots, i_{d-1}} \neq 0$ , as well as*

$$\mathbb{P}_{\mathbf{x}_d} \left( \left| \sum_{1 \leq i_d \leq n} \sum_{j_1 \dots j_{d-1}} k_{j_1 \dots j_{d-1}} a_{j_1 \dots j_{d-1} i_d} x_{i_d} \right| \leq \beta n^{O_{C,\varepsilon}(1)} \right) \geq n^{-O_{C,\varepsilon}(1)}.$$

Thus Theorem 5.5 asserts that as long as the entries in each block are not totally dependent, the conclusion of Theorem 3.7 still holds as if the matrix entries were mutually independent.

In what follows we introduce the main supporting lemmas to prove Theorem 5.5. By definition, we can rewrite this form as

$$\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d} \det[\mathbf{c}_{i_1}, \dots, \mathbf{c}_{i_d}],$$

where  $\det[\mathbf{c}_{i_1}, \dots, \mathbf{c}_{i_d}]$  is the determinant of the  $d \times d$  block generated by the  $i_1$ -th,  $\dots$   $i_d$ -th columns of the matrix of the first  $d$  rows of  $M_n$ .

Let  $\mathcal{U} := \{U_1, \dots, U_d\}$  be an ordered random partition of  $[n]$ . These index sets will serve as the collection of blocks (among the  $n$  blocks of size  $d \times d$  of the matrix generated by the first  $d$  rows) to be partitioned. We denote by  $B(U_i)$  the collection of indices generated by  $U_i$ , that is

$$B(U_i) := \cup_{l \in U_i} \{(d-1)l + 1, \dots, dl\}. \quad (5.1)$$

Given any partition  $\mathcal{U}$ , we easily obtain the following lemma by a series applications of the Cauchy-Schwarz inequality.



**Lemma 5.6** (Decoupling lemma). *Assume that*

$$\rho = \sup_a \mathbb{P}_{x_{11;11}, \dots, x_{dd;1n}} \left( \left| \sum_{1 \leq i_1, \dots, i_d \leq dn} a_{i_1 i_2 \dots i_d} \det[\mathbf{c}_{i_1}, \dots, \mathbf{c}_{i_d}] - a \right| \leq \beta \right) \geq n^{-B},$$

Then,

$$\mathbb{P}_{x'_{11;11}, \dots, x'_{dd;1n}} \left( \left| \sum_{i_1 \in B(U_1), \dots, i_d \in B(U_d)} a_{i_1 i_2 \dots i_d} \det[\mathbf{c}_{i_1}, \dots, \mathbf{c}_{i_d}] - a \right| = O_B(\beta \sqrt{\log n}) \right) = \Omega(\rho^{2d}), \quad (5.2)$$

where  $(x_{11;11}', \dots, x_{dd;11}')$ ;  $\dots$ ;  $(x_{11;1n}', \dots, x_{dd;1n}')$  are iid copies of the vector  $(x_{11} - x'_{11}, \dots, x_{dd} - x'_{dd})$ , and where  $(x'_{11}, \dots, x'_{dd})$  is an independent copy of  $(x_{11}, \dots, x_{dd})$ .

As the proof of Lemma 5.6 is standard, we refer the reader to [8, 25, 40]. As the columns  $\mathbf{c}_{i_1}, \dots, \mathbf{c}_{i_d}$  are independent, we will be able to obtain an analogue of Lemma 3.11 as follows.

**Lemma 5.7.** *There exist index sets  $I_0(U_1)$  with  $|I_0(U_1)| = O(1)$  and  $I(U_1) \subset B(U_1)$  with  $|I(U_1)| \geq d|U_1| - n^\varepsilon$  and an integer  $k \neq 0, k = n^{O_{B,\varepsilon}(1)}$  such that for any  $i \in I(U_1)$ , there exist integers  $k_{ii_0} = n^{O_{B,\varepsilon}(1)}$  such that*

$$\mathbb{P}_{x_{rs;1j}, j \in B(U_2) \cup \dots \cup B(U_d)} \left( \left| k \sum_{i_2 \in B(U_2), \dots, i_d \in B(U_d)} a_{i, i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right. \right. \\ \left. \left. - \sum_{i_0 \in I_0} k_{ii_0} \sum_{i_2 \in B(U_2), \dots, i_d \in B(U_d)} a_{i_0, i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right| = O(\beta n^{O_{B,\varepsilon}(1)}) \right) \geq \rho^{2d} / n^{O_{B,\varepsilon}(1)},$$

where  $\mathbf{c}_j^{\bar{1}}$  is the  $j$ -th column  $\mathbf{c}_j$  without its first component.

*Proof.* (of Lemma 5.7) As usual, it suffices to assume  $\xi$  to have discrete distribution. For each  $l \in U_1$ , let  $B_l = \{(l-1)d+1, \dots, ld\}$  be the  $l$ -th block. By the determinant expansion, we have

$$\sum_{i_1 \in B_l, i_2 \in B(U_2), \dots, i_d \in B(U_d)} a_{i_1 i_2 \dots i_d} \det[\mathbf{c}_{i_1}, \dots, \mathbf{c}_{i_d}] \\ = \sum_{i=1}^n x_{1i;1l} \sum_{i_2, \dots, i_d} (-1)^{1+i} a_{((l-1)d+i) i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \\ + \dots \\ + \sum_{i=1}^n x_{di;1l} \sum_{i_2, \dots, i_d} (-1)^{d+i} a_{((l-1)d+i) i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{d}}, \dots, \mathbf{c}_{i_d}^{\bar{d}}].$$

By summing over  $l \in U_1$  and by applying Lemma 5.6 and Corollary 5.3 for the random variables  $x_{rs;1j}, j \in B(U_1)$ , with high probability with respect to the random variables indexing from  $\bar{U}_1$  (i.e.  $x_{rs;1j}, j \in \bar{U}_1$ ), most of the coefficients

$$\sum_{i_2, \dots, i_d} a_{((l-1)d+1) i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}], \dots, \sum_{i_2, \dots, i_d} a_{((ld) i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}]$$

belong to a GAP of rank  $O_{C,\varepsilon}(1)$  and size  $n^{O_{C,\varepsilon}(1)}$ . From here, to conclude Lemma 5.7, we just follow the proof of Lemma 3.11 verbatim.  $\square$

**5.8. Randomization.** Roughly speaking, by iterating Lemma 5.7 to the new  $(d-1)$ -linear form of the random variables restricted by  $\bar{U}_1$  and so on, we will be able to deduce an analogue of Theorem 3.7 with the dependence upon  $U_1, \dots, U_d$ . One might then try to randomize  $U_1, \dots, U_d$  to obtain Theorem 5.5. However, the randomization of  $U_1, \dots, U_d$  altogether may pose a highly technical

difficulty. To avoid this hurdle we will try to randomize one pair at a time before each iteration of Lemma 5.7.

Assume that  $(U_{12}, U_3, \dots, U_d)$  is a  $(d-1)$  ordered partition of  $[n]$  in which each partition has order  $\Theta(n)$ . Fixing this partition, we next partition  $U_{12}$  into  $U_1, U_2$  randomly. For the new  $d$  partition  $(U_1, \dots, U_d)$  we then apply Lemma 5.7. As a result, the index  $i_2$  in the conclusion belongs to  $B(U_2)$ . We will show that by randomizing  $U_1$ , one may recover the result for  $i_2$  now an element of  $B(U_{12})$ . Let us first extend Lemma 5.7 as follows.

**Lemma 5.9.** *There exist subsets  $I_0(U_1)$  and  $I(U_1)$  of  $B(U_{12})$  with size  $O(1)$  and  $d|U_{12}| - n^\varepsilon$  respectively and an integer  $k \neq 0, k = n^{O_{B,\varepsilon}(1)}$  such that for any  $i \in I(U_1)$ , there exist integers  $k_{ii_0} = n^{O_{B,\varepsilon}(1)}$  such that the following holds for all  $i \in I(U_1)$  :*

$$\begin{aligned} & \mathbb{P}_{x_{11,11}, \dots, x_{dd,1n}} \left( \left| k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{ii_2 \dots i_d}(U_1) \det[\mathbf{c}_{i_2}^\dagger, \dots, \mathbf{c}_{i_d}^\dagger] \right. \right. \\ & \left. \left. - \sum_{i_0 \in I_0} k_{ii_0} \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i_0 i_2 \dots i_d}(U_1) \det[\mathbf{c}_{i_2}^\dagger, \dots, \mathbf{c}_{i_d}^\dagger] \right| = O(\beta n^{O_{B,\varepsilon}(1)}) \right) \geq \rho^{2d} / n^{O_{B,\varepsilon}(1)}, \end{aligned}$$

where

$$a_{ii_2 \dots i_d}(U_1) := \begin{cases} a_{ii_2 \dots i_d} & \text{if } i \in B(U_1), i_2 \in B(U_2) \\ a_{i_2 i \dots i_d} & \text{if } i \in B(U_2), i_2 \in B(U_1) \\ 0 & \text{otherwise.} \end{cases}$$

In comparison with Lemma 5.7, the probability in Lemma 5.9 is now with respect to all random variables of the first  $d$  rows of  $M_n$ . Also,  $i_2$  now runs over all the indices restricted by  $U_{12}$ . The entries  $a_{ii_2 \dots i_d}(U_1)$ , without the indices  $i_3, \dots, i_d$ , could be viewed as entries of a symmetric matrix.

*Proof.* (of Lemma 5.9) We first fix the random variables restricted by  $\bar{U}_{12}$  for which the conclusion of Lemma 5.6 holds with respect to the random variables restricted by  $U_{12}$ . Similarly to the proof of Lemma 5.7, the following holds with high probability with respect to  $x_{rs;1i}, i \in B(U_2)$ : there exist subsets  $I_0(U_1)$  and  $I(U_1)$  of  $B(U_1)$  with size  $O(1)$  and  $d|U_1| - n^\varepsilon$  respectively such that the following holds for all  $i \in I(U_1)$  :

$$\begin{aligned} & \mathbb{P}_{x_{rs;1i_2}, i_2 \in B(U_2)} \left( \left| k' \sum_{i_2 \in B(U_2), \dots, i_d \in B(U_d)} a_{ii_2 \dots i_d} \det[\mathbf{c}_{i_2}^\dagger, \dots, \mathbf{c}_{i_d}^\dagger] \right. \right. \\ & \left. \left. - \sum_{i_0 \in I_0} k'_{ii_0} \sum_{i_2 \in B(U_2), \dots, i_d \in B(U_d)} a_{i_0 i_2 \dots i_d} \det[\mathbf{c}_{i_2}^\dagger, \dots, \mathbf{c}_{i_d}^\dagger] \right| = O(\beta n^{O_{B,\varepsilon}(1)}) \right) \\ & \geq \rho^{2d} / n^{O_{B,\varepsilon}(1)}. \end{aligned} \tag{5.3}$$

By switching the role of  $U_1$  and  $U_2$ , there also exist subsets  $I_0(U_2)$  and  $I(U_2)$  of  $B(U_2)$  with size  $O(1)$  and  $d|U_2| - n^\varepsilon$  respectively such that the following holds for all  $i \in I(U_2)$  :

$$\begin{aligned} & \mathbb{P}_{x_{rs;1i_1}, i_1 \in B(U_1)} \left( \left| k'' \sum_{i_1 \in B(U_1), i_3 \in B(U_3), \dots, i_d \in B(U_d)} a_{i_1 i_3 \dots i_d} \det[\mathbf{c}_{i_1}^\dagger, \mathbf{c}_{i_3}^\dagger, \dots, \mathbf{c}_{i_d}^\dagger] \right. \right. \\ & \left. \left. - \sum_{i_0 \in I_0} k''_{ii_0} \sum_{i_1 \in B(U_1), i_3 \in B(U_3), \dots, i_d \in B(U_d)} a_{i_0 i_1 i_3 \dots i_d} \det[\mathbf{c}_{i_1}^\dagger, \mathbf{c}_{i_3}^\dagger, \dots, \mathbf{c}_{i_d}^\dagger] \right| = O(\beta n^{O_{B,\varepsilon}(1)}) \right) \\ & \geq \rho^{2d} / n^{O_{B,\varepsilon}(1)}. \end{aligned} \tag{5.4}$$

Now, by the definition of  $a_{ii_2 \dots i_d}(U_1)$ , with  $I = I(U_1) \cup I(U_2)$  and  $I_0 = I_0(U_1) \cup I_0(U_2)$ , we can rewrite both of the events in (5.3) and (5.4) in the following form

$$\begin{aligned}
& k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{ii_2 \dots i_d}(U_1) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \\
& - \sum_{i_0 \in I_0} k_{ii_0} \sum_{i_2 \in B(U_2), \dots, i_d \in B(U_d)} a_{i_0 i_2 \dots i_d}(U_1) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] = O(\beta n^{O_{B,\varepsilon}(1)}).
\end{aligned}$$

The conclusion of Lemma 5.9 then follows from (5.3) and (5.4), noting that  $\{x_{rs;1i_1}, i_1 \in U_1\}$  and  $\{x_{rs;1i_2}, i_2 \in U_2\}$  are independent.  $\square$

Now we randomize  $U_1$  to obtain the following main result of the subsection.

**Lemma 5.10** (Randomization). *There exist subsets  $I_0(U_{12})$  and  $I(U_{12})$  of  $B(U_{12})$  with size  $O(1)$  and  $d|U_{12}| - n^\varepsilon$  respectively such that the following holds for all  $i \in I$ :*

$$\begin{aligned}
& \mathbb{P}_{x_{rs;1i}, i \in \bar{B}(U_{12}); x_{rs;1i'}, i \in B(U_{12})} \left( \left| k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i, i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}'}, \mathbf{c}_{i_3}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right. \right. \\
& \left. \left. - \sum_{i_0 \in I_0} k_{ii_0} \sum_{i_2 \in B(U_2), \dots, i_d \in B(U_d)} a_{i_0, i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}'}, \mathbf{c}_{i_3}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right| = O(\beta n^{O_{B,\varepsilon}(1)}) \right) \geq \rho^{2d}/n^{O_{B,\varepsilon}(1)},
\end{aligned}$$

where  $x_{rs}^{(i)'} := \eta_i x_{rs}^{(i)}$  with  $\eta_i$  iid Bernoulli random variables of parameter  $1/2$ , and  $\mathbf{c}_{i_2}^{\bar{1}'} := \eta_{i_2} \mathbf{c}_{i_2}^{\bar{1}}$  in the determinants.

*Proof.* (of Theorem 5.10) Note that Lemma 5.9 holds for any choice  $U_1 \subset U_{12}$ . As  $I_0(U_1) \subset [n]^{O_{B,\varepsilon}(1)}$  and  $k(U_1) \leq n^{O_{B,\varepsilon}(1)}$ , there are only  $n^{O_{B,\varepsilon}(1)}$  possibilities that the tuple  $(I_0(U_1), k(U_1))$  can take. Thus, there exists a tuple  $(I_0, k)$  such that  $I_0(U_1) = I_0$  and  $k(U_1) = k$  for  $2^{|U_{12}|}/n^{O_{B,\varepsilon}(1)}$  different sets  $U_1$ . Let us denote this set of  $U_1$  by  $\mathcal{S}$ ; we have

$$|\mathcal{S}| \geq 2^{|U_{12}|}/n^{O_{B,\varepsilon}(1)}.$$

Next, let  $I$  be the collection of all  $i \in B(U_{12})$  which belong to at least  $|\mathcal{S}|/2$  index sets  $I(U_1)$ . Then,

$$\begin{aligned}
|I||\mathcal{S}| + (d|U_{12}| - |I|)|\mathcal{S}|/2 & \geq (d|U_{12}| - n^\varepsilon)|\mathcal{S}| \\
|I| & \geq |U_{12}| - 2n^\varepsilon.
\end{aligned}$$

From now on we fix an  $i \in I$ . Consider the tuples  $(k_{ii_0}(U_1), i_0 \in I_0)$  over all  $U_1$  where  $i \in I(U_1)$ . Because there are only  $n^{O_{B,\varepsilon}(1)}$  possibilities such tuples can take, there must be a tuple, say  $(k_{ii_0}, i_0 \in I_0)$ , such that  $(k_{ii_0}(U_1), i_0 \in I_0) = (k_{ii_0}, i_0 \in I_0)$  for at least  $|\mathcal{S}|/2n^{O_{B,\varepsilon}(1)} = 2^n/n^{O_{B,\varepsilon}(1)}$  sets  $U_1$ . Without loss of generality, we assume that  $i \in U_1$  for at least half of those sets. Let  $\mathcal{U}$  denote the collection of such  $U_1$ , and for each  $U_1 \in \mathcal{U}$  we let  $\mathbf{u} = (u_1, \dots, u_{|B(U_{12})|}) \in \mathbb{R}^{|B(U_{12})|}$  be its characteristic vector, i.e.  $u_i = 1$  if  $i \in B(U_1)$ , and  $u_i = 0$  otherwise.

By the definition of  $a_{ii_2 \dots i_d}(U_1)$ , as  $i \in U_1$ , we can write

$$\begin{aligned}
& k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i, i_2 \dots i_d}(U_1) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \\
& - \sum_{i_0 \in I_0} k_{ii_0} \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i_0, i_2 \dots i_d}(U_1) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \\
& = k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i, i_2 \dots i_d}(1 - u_{i_2}) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \\
& - \sum_{i_0 \in I_0} k_{ii_0} \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i_0, i_2 \dots i_d}(1 - u_{i_2}) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}].
\end{aligned}$$

Recall that  $|\mathcal{U}| = 2^{|U_{12}|}/n^{O_{B,\varepsilon}(1)}$ . Hence, by Lemma 5.9, we have the following joint probability

$$\begin{aligned} & \mathbb{P}_{x_{11;11}, \dots, x_{dd;1d}} \mathbb{P}_{U_1} \left( \left| k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i, i_2 \dots i_d} (1 - u_{i_2}) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right. \right. \\ & \quad \left. \left. - \sum_{i_0 \in I_0} k_{i i_0} \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i_0 i_2 \dots i_d} (1 - u_{i_2}) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right| = O(\beta n^{O_{B,\varepsilon}(1)}) \right) \\ & = n^{-O_{B,\varepsilon}(1)}. \end{aligned}$$

By applying the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} n^{-O_{B,\varepsilon}(1)} & \leq \left[ \mathbb{E}_{x_{rs;1i}, 1 \leq i \leq n, 1 \leq r, s \leq d} \mathbb{E}_{U_1}(\dots) \right]^2 \leq \mathbb{E}_{x_{11;11}, \dots, x_{dd;1d}} \left[ \mathbb{E}_{U_1}(\dots) \right]^2 = \mathbb{E}_{x_{11;11}, \dots, x_{dd;1d}} \left[ \mathbb{E}_{\mathbf{u}}(\dots) \right]^2 \\ & \leq \mathbb{E}_{x_{rs;1i}, 1 \leq i \leq n, 1 \leq r, s \leq d} \mathbb{E}_{\mathbf{u}, \mathbf{u}'} \left( k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i, i_2 \dots i_d} (u_{i_2} - u'_{i_2}) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right. \\ & \quad \left. - \sum_{i_0 \in I_0} k_{i i_0} \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i_0 i_2 \dots i_d} (u_{i_2} - u'_{i_2}) \det[\mathbf{c}_{i_2}^{\bar{1}}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right) = O(\beta n^{O_{B,\varepsilon}(1)}) \\ & = \mathbb{E}_{x_{rs;1i}, i \notin B(U_{12}); x_{rs;1i'}, i \in B(U_{12}), 1 \leq r, s \leq d} \left( k \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i, i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}'}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right. \\ & \quad \left. - \sum_{i_0 \in I_0} k_{i i_0} \sum_{i_2 \in B(U_{12}), \dots, i_d \in B(U_d)} a_{i_0 i_2 \dots i_d} \det[\mathbf{c}_{i_2}^{\bar{1}'}, \dots, \mathbf{c}_{i_d}^{\bar{1}}] \right) = O(\beta n^{O_{B,\varepsilon}(1)}), \end{aligned} \quad (5.5)$$

where  $x_{rs}^{(i)'} := (u_i - u'_i)x_{rs}^{(i)}$  and in the determinant formulas the column  $\mathbf{c}_{i_2}^{\bar{1}'}$  stands for  $(u_{i_2} - u'_{i_2})\mathbf{c}_{i_2}^{\bar{1}}$ . Also, in the first estimate we used the elementary property that for any function  $f$ ,

$$\begin{aligned} & \mathbb{E}_{\mathbf{u}, \mathbf{u}'} \left( \|f(\mathbf{u})\|_2 = O(\beta n^{O_{B,\varepsilon}(1)}), \|f(\mathbf{u}')\|_2 = O(\beta n^{O_{B,\varepsilon}(1)}) \right) \\ & \leq \mathbb{E}_{\mathbf{u}, \mathbf{u}'} \left( \|f(\mathbf{u}) - f(\mathbf{u}')\|_2 = O(\beta n^{O_{B,\varepsilon}(1)}) \right). \end{aligned}$$

The proof is complete by noting that  $k$  and  $I_0$  are independent of the choice of  $i$ .  $\square$

*Proof.* (of Theorem 5.5) Note that by Lemma 5.10, we just need to deal with a  $(d-1)$ -multilinear form of the rows  $\mathbf{r}_2, \dots, \mathbf{r}_d$ . Our next step is to apply this machinery again when fixing  $U_{123} = U_{12} \cup U_3$  and letting  $U_{12}$  be chosen as a random subset of  $U_{123}$ . By iterating the machinery  $d-1$  times similarly to the proof of Theorem 3.7, we obtain the result as claimed.  $\square$

We now conclude this section by proving the inverse step of Section 4.

**5.11. Proof of Theorem 4.5.** We first apply Theorem 5.5 to obtain

$$\mathbb{P}_{\mathbf{x}_d} \left( \sum_{1 \leq i_d \leq dn} \sum_{j_1 \dots j_{d-1}} k_{j_1 \dots j_{d-1}} a_{j_1 \dots j_{d-1} i_d} x_{i_d} \in B(a, \beta) \right) \geq \rho^{2d}/n^{O_{C,\varepsilon}(1)}. \quad (5.6)$$

Set  $K_0, \dots, K_d$  to be a sequence of thresholds with  $K_i := n^{-A/2+2id}$ . We consider two cases.

**Case 1.** (*degenerate case*)

*Subcase 1.1.* Assume that for all  $i \in I_1$ ,

$$\sum_{i_2, \dots, i_d} |k a_{i i_2 \dots i_d} - \sum k_{i i_0} a_{i_0 i_2 \dots i_d}|^2 \leq K_0^2 = n^{-A}. \quad (5.7)$$

As  $\sum_{i_1, i_2, \dots, i_d} |a_{i_1 i_2 \dots i_d}|^2 = 1$ , there exists  $i_2, \dots, i_d$  such that

$$\sum_{i_1} |a_{i_1 i_2 \dots i_d}|^2 \geq 1/(dn)^{d-1} \geq n^{-O_d(1)}. \quad (5.8)$$

We next fix these indices  $i_2, \dots, i_d$ . It follows from (5.7) that  $|ka_{ii_2 \dots i_d} - \sum k_{ii_0} a_{i_0 i_2 \dots i_d}| \leq K_0$  for any  $i \in I_1$ . Set

$$v_i := \frac{1}{k} \sum k_{ii_0} a_{i_0 i_2 \dots i_d}.$$

It is clear that the set of  $v_i$ 's is a GAP of rank  $|I_0| = O_{B, \varepsilon}(1)$  and size  $n^{O_{B, \varepsilon}(1)}$ . Also, by definition, with  $\mathbf{v} = (v_i, i \in I_1)$

$$\|\mathbf{v} - (a_{ii_2 \dots i_d})_{1 \leq i \leq dn}\| \leq n^{1/2} K_0 = n^{-A/2+1/2}.$$

On the other hand, as the vector  $(a_{ii_2 \dots i_d})_{1 \leq i \leq dn}$  is orthogonal to any row  $\mathbf{r}_j(\mathbf{M}_{n-1})$  of  $\mathbf{M}_{n-1}$ , we have

$$|\langle \mathbf{v}, \mathbf{r}_j(\mathbf{M}_{n-1}) \rangle| \leq n^{-A/2+O_{B, \varepsilon}(1)}.$$

Recall that by the approximation and by (5.7),  $\|\mathbf{v}\| \geq n^{-O_d(1)}$ . Thus by letting  $\mathbf{u} := \mathbf{v}/\|\mathbf{v}\|$ , we have

$$|\langle \mathbf{u}, \mathbf{r}_j(\mathbf{M}_{n-1}) \rangle| \leq n^{-A/2+O_{B, \varepsilon, d}(1)}.$$

It is clear that  $\mathbf{u}$  satisfies all the conditions of Theorem 4.5, we are done with this subcase.

*Subcase 1.2.* From now on we assume that there exists  $i \in I_1$  such that

$$\sum_{i_2, \dots, i_d} |ka_{ii_2 \dots i_d} - \sum k_{ii_0} a_{i_0 i_2 \dots i_d}|^2 \geq K_1^2 = n^{-A}. \quad (5.9)$$

Fixing  $i$ , we apply Theorem 5.5 for the index  $i_2$ , and reconsidering Subcases 1.1 with the new threshold  $K_1 := n^{-A/2-2d}$ . By iterating the process for at most  $d-1$  steps, we will end up with either Subcase 1.1 (and hence done) or with the following non-degenerate case.

**Case 2.** (*non-degenerate case*). There exist a collection  $J_1, \dots, J_{d-1}$  of the indices  $j_1, \dots, j_{d-1}$  such that  $|J_i| = O_{B, \varepsilon}(1)$  and

$$\sum_{1 \leq i \leq dn} \left| \sum_{j_1 \in J_1, \dots, j_{d-1} \in J_{d-1}} k_{j_1 \dots j_{d-1}} a_{j_1 \dots j_{d-1} i} \right|^2 \geq K_d^2 = n^{-A+2d^2}.$$

Notice that for each fixed  $i_1, \dots, i_{d-1}$  the vector  $(a_{i_1, \dots, i_{d-1}, i}, 1 \leq i \leq dn, i \neq i_1, \dots, i_{d-1})$  is orthogonal to any  $\mathbf{r}_j^{(i_1, \dots, i_{d-1})}(\mathbf{M}_{n-1})$ , the  $j$ -th row of  $\mathbf{M}_{n-1}$  without components  $i_1, \dots, i_{d-1}$ . By adding zeros to the missing components  $i_1, \dots, i_{d-1}$  if needed, we see that the  $\mathbb{R}^{nd}$  vector  $(a_{i_1, \dots, i_{d-1}, i}, 1 \leq i \leq dn)$  is now orthogonal to  $\mathbf{r}_j(\mathbf{M}_{n-1})$ .

Set  $J = J_1 \cup \dots \cup J_{d-1}$ , we thus have

$$\sum_{i \notin J, 1 \leq i \leq dn} a_{j_1, \dots, j_{d-1}, i} \mathbf{r}_j(i) = - \sum_{i \in J, 1 \leq i \leq dn} a_{j_1, \dots, j_{d-1}, i} \mathbf{r}_j(i),$$

where the entries  $a_{j_1, \dots, j_{d-1}, i}$  are set to be zero if the indices are not distinct.

Now we set  $\mathbf{w}_1 := (w_i)_{i \notin J}$  and  $\mathbf{w}_2 := (w_i)_{i \in J}$ , where  $w_i := k_{j_1 \dots j_{d-1}} a_{i_1 \dots i_{d-1} i}$ . Then

$$\langle \mathbf{w}_1, \mathbf{r}_j^{(J)}(\mathbf{M}_{n-1}) \rangle = - \langle \mathbf{w}_2, \mathbf{r}_j^{(J)}(\mathbf{M}_{n-1}) \rangle.$$

Set  $\mathbf{v} := \mathbf{w}/\|\mathbf{w}\|$ . Theorem 3.3 applied to (5.6) implies that  $\mathbf{v}$  can be approximated by a vector  $\mathbf{u}$  as follows.

- $|u_i - v_i| \leq n^{-A/2+O_{B, \varepsilon, d}(1)}$  for all  $i$ .

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

Note that, by the approximation above,  $\|\mathbf{u}\| \asymp 1$  and  $|\langle \mathbf{u}_1, \mathbf{r}_j^{(\bar{J})}(\mathbf{M}_{n-1}) \rangle + \langle \mathbf{u}_2, \mathbf{r}_j^{(J)}(\mathbf{M}_{n-1}) \rangle| \leq n^{-A/2+O_{B,\varepsilon}(1)}$  for all row vectors of  $\mathbf{M}_{n-1}$ .

## 6. SINGULARITY OF BLOCK MATRICES: PROOF SKETCH FOR THEOREM 4.6

Our first ingredient is the following variant of Theorem 3.3 in which random variables are replaced by random matrices and  $a_i$  are replaced by vectors.

**Theorem 6.1** (Inverse Littlewood-Offord for sequence of random operators). *Let  $\{\mathbf{u}^{(i)} = (\mathbf{u}_1^{(i)}, \dots, \mathbf{u}_d^{(i)})\}$ ,  $1 \leq i \leq n$  be a sequence of  $n$  vectors in  $\mathbb{C}^d$  such that the following concentration-type holds with high probability*

$$\sup_{\mathbf{u} \in \mathbb{C}^d} \mathbb{P}_{X^{(1)}, \dots, X^{(n)}} \left( \sum_{1 \leq i \leq n} X^{(i)} \mathbf{u}^{(i)} \in B(\mathbf{u}, \beta) \right) = \gamma = n^{-O(1)},$$

where  $X^{(1)}, \dots, X^{(n)}$  are iid block matrices whose entries are copies of  $(x_{11}, \dots, x_{dd})$  from (4.1). Then there exist a positive constant  $\delta$  and  $d^2$  numbers  $c_{11}, \dots, c_{dd}$  such that the least singular value  $\sigma_d$  (largest singular value  $\sigma_1$ ) of the matrix  $(c_{ij})_{1 \leq i, j \leq d}$  is at least  $\delta$  (at most  $\delta^{-1}$ ) and for any number  $n'$  between  $n^\varepsilon$  and  $n$ , there exists a proper symmetric GAP  $Q = \{\sum_{i=1}^r k_i g_i : |k_i| \leq K_i\} \subset \mathbb{C}^d$  such that

- At least  $n - n'$  elements of  $V := \{(c_{11}\mathbf{u}_1^{(i)} + \dots + c_{1d}\mathbf{u}_d^{(i)}), \dots, (c_{d1}\mathbf{u}_1^{(i)} + \dots + c_{dd}\mathbf{u}_d^{(i)})\}$ ,  $1 \leq i \leq n$  are  $\beta$ -close to  $Q$ .
- $Q$  has small rank,  $r = O_{B,\varepsilon}(1)$ , and small size

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

- There is a non-zero integer  $p = O_{B,\varepsilon}(\sqrt{n'})$  such that all steps  $g_i$  of  $Q$  have the form  $g_i = (g_{i1}, \dots, g_{id})$ , where  $g_{ij} = \beta \frac{p_{ij}}{p}$  with  $p_{ij} \in \mathbb{Z}$  and  $|p_{ij}| = O_{B,\varepsilon}(\beta^{-1}\sqrt{n'})$ .

In application,  $X^{(1)}, \dots, X^{(n)}$  will be the  $d \times d$  blocks of  $\mathbf{M}_{n-1}$ . It is crucial to notice that, as most of the elements of  $V$  are  $\beta$ -close to  $Q$ , and as the matrix  $(c_{ij})$  is far from being degenerate, it follows from Theorem 6.1 that most of the individual components  $\mathbf{u}_j^{(i)}$  are also close to another GAP of small rank and size (see Corollary 5.3). We will present the proof of Theorem 6.1 in Appendix A by following the treatment from [26]. For the rest of this section we sketch the proof of Theorem 4.6 following [24] and [37].

Let  $\mathcal{N}$  be the number of such structural vectors  $\mathbf{u}$  from Theorem 4.5. Because each GAP is determined by its generators and dimensions, the number of  $Q$ 's is bounded by

$$(n^{2A+O_{B,\varepsilon}(1)})^{O_{B,\varepsilon}(1)} (n^{O_{B,\varepsilon}(1)})^{O_{B,\varepsilon}(1)} = n^{O_{A,B,\varepsilon}(1)}.$$

Next, for a given  $Q$ , there are at most  $n^{O_{B,\varepsilon}(n)}$  ways to choose the  $nd - 2n^\varepsilon$  components  $u_i$  that  $Q$  contains, and  $n^{O_{A,B,\varepsilon}(n^\varepsilon)}$  ways to choose the remaining components from the set  $\{\frac{p}{q} + i\frac{p'}{q'}, |p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\varepsilon}(1)}\}$ . Hence, we obtain the key bound

$$\mathcal{N} \leq n^{O_{A,B,\varepsilon}(1)} n^{O_{B,\varepsilon}(n)} n^{O_{A,B,\varepsilon}(n^\varepsilon)} = n^{O_{B,\varepsilon}(n)}. \quad (6.1)$$

From now on, by conditioning on  $\mathbf{u}_1$  and on the entries of  $\mathbf{M}_{n-1}$  corresponding to the index  $i_1, \dots, i_d$  of  $\mathbf{u}_1$ , without loss of generality we assume that  $\mathbf{u}_1$  vanishes. Set  $\beta_0 := n^{-A/2+O_{B,\varepsilon}(1)}$ , the bound obtained from the conclusion of Theorem 4.5. We will denote the blocks of  $\mathbf{M}_{n-1}$  by  $X_j^{(i)}$  with  $1 \leq i \leq n$  and  $1 \leq j \leq n-1$ . For a given vector  $\mathbf{u}$ , we define  $\mathbb{P}_{\beta_0}(\mathbf{u})$  as follows

$$\mathbb{P}_{\beta_0}(\mathbf{u}) := \mathbb{P}\left(\sum_{1 \leq i \leq n} X_j^{(i)} \mathbf{u}^{(i)} \in B(0, \beta_0) \text{ for at least } (n-1) - O_{B,\varepsilon}(1) \text{ indices } j\right).$$

If  $\mathbf{u}$  satisfies the property above, we say that  $\mathbf{u}$  is  $\beta_0$ -orthogonal to almost all blocks of  $\mathbf{M}_{n-1}$ . Because the blocks of  $\mathbf{M}_{n-1}$  are iid,

$$\mathbb{P}_{\beta_0}(\mathbf{u}) \leq (\mathbb{P}_{X^{(1)}, \dots, X^{(n)}}(\sum_{1 \leq i \leq n} X^{(i)} \mathbf{u}^{(i)} \in B(0, \beta_0))^{n-O(1)} := \gamma_{\beta_0}(\mathbf{u})^{n-O(1)},$$

where  $X^{(i)}$  are iid copies of  $(x_{st})_{1 \leq s, t \leq d}$ .

If  $\gamma = \gamma_{\beta_0}(\mathbf{u})$  is small, say  $n^{-\Omega(1)}$ , then  $\mathbb{P}_{\beta_0}(\mathbf{u})$  is  $n^{-\Omega(n)}$ . Thus the contribution of these  $\mathbb{P}_{\beta_0}(\mathbf{u})$  in the total sum  $\sum_{\mathbf{u}} \mathbb{P}_{\beta_0}(\mathbf{u})$  is negligible as  $\mathcal{N} = n^{O(n)}$ .

For the case  $\gamma$  is comparably large,  $\gamma = n^{-O(1)}$ , then by Theorem 6.1, 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  $|\langle \mathbf{u}', \mathbf{r}_i(\mathbf{M}_{n-1}) \rangle|$ , where we recall that  $\mathbf{r}_i(\mathbf{M}_{n-1})$  is the  $i$ -th row of  $\mathbf{M}_{n-1}$ , is still of order  $O(\beta_0)$  and the components of  $\mathbf{u}'$  are now from the new GAPs (after a linear transformation). The number  $\mathcal{N}'$  of these  $\mathbf{u}'$  can be bounded by  $(\gamma^{-1}/n^\varepsilon)^n$ , while we recall that  $\mathbb{P}_{\beta_0}(\mathbf{u}')$  is of order  $\gamma^{-n}$ . Thus, summing over  $\mathbf{u}'$  we obtain the desired bound

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

To proceed further, we need the following elementary claim.

**Claim 6.2.** *Assume that  $\mathbf{C} = (c_{ij})$  is a  $d \times d$  matrix such that  $\delta \leq \sigma_d(\mathbf{C}) \leq \sigma_1(\mathbf{C}) \leq \delta^{-1}$ . Let  $\mathbf{u}' = (\mathbf{u}'^{(1)}, \dots, \mathbf{u}'^{(n)})$ , where  $\mathbf{u}'^{(i)} = \mathbf{C}\mathbf{u}^{(i)}$ . Then we have*

$$\gamma_\beta(\mathbf{u}) \leq \gamma_{\delta^{-1}\beta}(\mathbf{u}') \leq \gamma_{\delta^{-2}\beta}(\mathbf{u}).$$

By paying a factor of  $n^{O_{B,\varepsilon}(1)}$  in probability, we may assume that  $|\langle \mathbf{u}, \mathbf{r}_i(\mathbf{M}_n) \rangle| \leq \beta_0$  for the first  $d(n-1) - O_{B,\varepsilon}(1)$  rows of  $\mathbf{M}_n$ . Also, by paying another factor of  $n^{n^\varepsilon}$  in probability, we may assume that the first  $dn_0$  components of  $\mathbf{u}$  belong to a GAP  $Q$ , and the Euclidean norm of  $\mathbf{u}^{(n_0)}$  is comparable,  $\|\mathbf{u}^{(n_0)}\| = \Omega(1/n)$  (recall that  $\|\mathbf{u}\| \asymp 1$ ), where

$$n_0 := n - n^\varepsilon.$$

We refer to the remaining  $u_i$ 's as exceptional components. Note that these extra factors do not affect our final bound  $\exp(-\Omega(n))$ . Because  $\|\mathbf{u}^{(n_0)}\| = \Omega(1/n)$  and  $X^{(n_0)}$  is not degenerate with high probability, there exist positive constants  $c_1, c_2$  such that  $c_2 < 1$  and for any  $\beta \leq c_1/\sqrt{n}$  we have

$$\gamma_\beta(\mathbf{u}) \leq \sup_a \mathbb{P}_{X^{(n_0)}}(|X^{(n_0)} \mathbf{u}^{(n_0)} - a| \leq \beta) \leq 1 - c_2. \quad (6.2)$$

**6.3. Classification.** Next, let  $C$  be a sufficiently large constant depending on  $B$  and  $\varepsilon$  but not  $A$ . We classify  $\mathbf{u}$  into two classes  $\mathcal{B}$  and  $\mathcal{B}'$ , depending on whether  $\mathbb{P}_{\beta_0}(\mathbf{u}) \geq n^{-Cn}$  or not. Because of (6.1), for  $C$  large enough,

$$\sum_{\mathbf{u} \in \mathcal{B}'} \mathbb{P}_{\beta_0}(\mathbf{u}) \leq n^{O_{B,\varepsilon}(n)} / n^{Cn} \leq n^{-n/2}. \quad (6.3)$$

For the set  $\mathcal{B}$  of remaining vectors, we divide it into two subfamilies. Set  $n' := n^{1-\varepsilon}$ . We say that  $\mathbf{u} \in \mathcal{B}$  is compressible if for any  $n'$  components  $\mathbf{u}^{(i_1)}, \dots, \mathbf{u}^{(i_{n'})}$  among the  $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(n_0)}$ , we have

$$\sup_a \mathbb{P}_{X_{i_1}, \dots, X_{i_{n'}}} (|X_{i_1} \mathbf{u}^{(i_1)} + \dots + X_{i_{n'}} \mathbf{u}^{(i_{n'})} - a| \leq n^{-B-4}) \geq (n')^{-1/2+o(1)}. \quad (6.4)$$

Let  $\mathcal{B}_1$  and  $\mathcal{B}_2$  be the set of compressible and incompressible vectors respectively. We focus on  $\mathcal{B}_1$  first.

**6.4. Approximation for compressible vectors.** Set  $\beta := n^{-B-4}$ . It follows from Theorem 6.1 that, among any  $\mathbf{u}^{(i_1)}, \dots, \mathbf{u}^{(i_{n'})}$ , there are, say, at least  $n'/2+1$  vectors that belong to a ball of radius  $\beta$  in  $\mathbb{C}^d$  (because our GAP now has only one element after a linear transformation  $\mathbf{C} = (c_{ij})_{1 \leq i, j \leq d}$ ). A simple covering argument then implies that there is a ball of radius  $2\beta$  that contains all but  $n'-1$  vectors  $\mathbf{u}^{(i)}$ .

Thus there exists a vector  $\mathbf{u}' = (\mathbf{u}'^{(1)}, \dots, \mathbf{u}'^{(n)}) \in (2\beta) \cdot (\mathbb{Z} + \sqrt{-1}\mathbb{Z})^{nd}$  such that

- $|\mathbf{C}\mathbf{u}^{(i)} - \mathbf{u}'^{(i)}| \leq 4\beta$  for all  $i$ ;
- $\mathbf{u}'^{(i)}$  takes the same vector-value for at least  $n_0 - n'$  indices  $i$ .

Because of the approximation and Assumption 4.2, whenever  $\sum_{1 \leq i \leq n} X^{(i)} \mathbf{u}^{(i)} \in B(\mathbf{u}, \beta)$ , we also have

$$\left| \sum_{1 \leq i \leq n} X^{(i)} \mathbf{u}'^{(i)} - \mathbf{C} \sum_{1 \leq i \leq n} X^{(i)} \mathbf{u}^{(i)} \right| \leq n(n^{B+1} + n^\alpha)(4\beta) + \beta_0 := \beta'.$$

By definition,  $\beta' \leq c_1/\sqrt{n}$ , and thus by (6.2),  $\mathbb{P}_{\beta'}(\mathbf{u}') \leq (1 - c_2)^{(1-o(1))n}$ . Now we bound the number of  $\mathbf{u}'$  obtained from the approximation. First, there are  $O(n^{n-n_0+n'}) = O(n^{2n^{1-\varepsilon}})$  ways to choose those  $\mathbf{u}'^{(i)}$  that take the same vector  $\mathbf{w} \in \mathbb{C}^d$ , and there are just  $O(\beta^{-d})$  ways to choose  $\mathbf{w}$ . The remaining components belong to the set  $(2\beta)^{-d} \cdot (\mathbb{Z} + i\mathbb{Z})^d$ , and thus there are at most  $O((\beta^{-d})^{n-n_0+n'}) = O(n^{O_{A,B,\varepsilon}(n^{1-\varepsilon})})$  ways to choose them. Hence we obtain the total bound

$$\sum_{\mathbf{u} \in \mathcal{B}_1} \mathbb{P}_{\beta_0}(\mathbf{u}) \leq \sum_{\mathbf{u}'} \mathbb{P}_{\beta'}(\mathbf{u}') \leq O(n^{2n^{1-\varepsilon}}) O(n^{O_{A,B,\varepsilon}(n^{1-\varepsilon})}) (1 - c_2)^{(1-o(1))n} \leq (1 - c_2)^{(1-o(1))n}.$$

**6.5. Approximation for incompressible vectors.** The treatment here is similar, we will sketch the main steps, leaving the details for the reader as an exercise.

First, by exposing the rows of  $\mathbf{M}_{n-1}$  accordingly, and by paying an extra factor  $\binom{n_0}{n'} = O(n^{n^{1-\varepsilon}})$  in probability, we can assume that the components  $\mathbf{u}^{(n_0-n'+1)}, \dots, \mathbf{u}^{(n_0)}$  satisfy

$$\begin{aligned} \sup_{\mathbf{a} \in \mathbb{C}^d} \mathbb{P}_{X^{(n_0-n'+1)}, \dots, X^{(n_0)}} (|X^{(n_0-n'+1)} \mathbf{u}^{(n_0-n'+1)} + \dots + X^{(n_0)} \mathbf{u}^{(n_0)} - \mathbf{a}| \leq n^{-B-4}) &\leq (n')^{-1/2+o(1)} \\ &\leq n^{-1/2+\varepsilon/2+o(1)}. \end{aligned} \quad (6.5)$$



Next, define a radius sequence  $\beta_k, k \geq 0$  where  $\beta_0 = n^{-A/2+O_{B,\varepsilon}(1)}$  is the bound obtained from the conclusion of Theorem 4.5, and  $\beta_{k+1} := (n^{B+2} + n^{\alpha+1} + 1)^2 \beta_k$ . Also define

$$\gamma_{\beta_k}(\mathbf{u}) := \sup_{\mathbf{a} \in \mathbb{C}^d} \mathbb{P}_{X^{(n_0-n'+1)}, \dots, X^{(n_0)}}(|X^{(n_0-n'+1)} \mathbf{u}^{(n_0-n'+1)} + \dots + X^{(n_0)} \mathbf{u}^{(n_0)} - \mathbf{a}| \leq \beta).$$

Clearly  $\mathbb{P}_{\beta_k}(\mathbf{u}) \leq \gamma_{\beta_k}^{n-1}(\mathbf{u})$ . As  $\mathbf{C}$  is non-degenerate, with  $\mathbf{u}'$  from Theorem 6.1,

$$\gamma_{\beta_k}(\mathbf{u}) \leq \gamma_{(n^{B+2}+n^{\alpha+1}+1)\beta_k}(\mathbf{u}') \leq \gamma_{\beta_{k+1}}(\mathbf{u}). \quad (6.6)$$

Furthermore, we have freedom to choose  $k$  before applying Theorem 6.1 to obtain  $\mathbf{u}'$ . By the pigeon-hole principle, there exists  $k = k_0(\mathbf{u}) \leq C\varepsilon^{-1}$  such that

$$\pi_{\beta_{k_0+1}}(\mathbf{u}) \leq n^{\varepsilon n} \pi_{\beta_{k_0}}(\mathbf{u}). \quad (6.7)$$

Since  $A$  was chosen sufficiently large compared to  $O_{B,\varepsilon}(1)$  and  $C$ , we have  $\beta_{k_0+1} \leq n^{-B-4}$ . With this choice of  $k_0$ , we apply Theorem 6.1 to obtain an approximation  $\mathbf{u}'$  of  $\mathbf{C}\mathbf{u}$  with the following properties.

- (i)  $|\mathbf{C}\mathbf{u}^{(i)} - \mathbf{u}'^{(i)}| \leq \beta_{k_0}$  for all  $i$ .
- (ii) The components of  $\mathbf{u}'^{(i)}$  belong to  $Q$  for all but  $n^{1-2\varepsilon}$  indices  $i$ , and the generators of  $Q$ , belong to the set  $\beta_{k_0} \cdot \{p/q + \sqrt{-1}p'/q', |p|, |q|, |p'|, |q'| \leq n^{A/2+O_{B,\varepsilon}(1)}\}$ .
- (iii)  $Q$  has rank  $O_{B,\varepsilon}(1)$  and size  $|Q| = O(\gamma_{\beta_{k_0}}^{-1}/n^{1/2-\varepsilon})$ .

Let  $\mathcal{B}'$  be the collection of such  $\mathbf{u}'$ . By definition,

$$\mathbb{P}_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}(\mathbf{u}') = \gamma_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}^{n-1}(\mathbf{u}') \leq \gamma_{\beta_{k_0+1}}^{n-1}(\mathbf{u}) \leq n^{\varepsilon n} \gamma_{\beta_{k_0}}^{n-1}(\mathbf{u}). \quad (6.8)$$

Arguing similarly to the treatment for  $\mathcal{N}$ , we can bound the cardinality  $\mathcal{N}'$  of  $\mathcal{B}'$  by

$$\mathcal{N}' \leq \gamma_{\beta_{k_0}}(\mathbf{u})^{-n} / n^{(1/2-\varepsilon-o(1))n}. \quad (6.9)$$

It follows from (6.8) and (6.9) that

$$\sum_{\mathbf{u}' \in \mathcal{B}'} \mathbb{P}_{(n^{B+2}+n^{\alpha+1}+1)\beta_{k_0}}(\mathbf{u}') \leq n^{-(1/2-4\varepsilon-o(1))n},$$

completing the treatment for incompressible vectors.

## 7. UNIVERSALITY OF RANDOM BLOCK MATRICES: PROOF OF THEOREM 2.7

This section is devoted to Theorem 2.7. We begin by introducing the following notation. Given a  $n \times n$  matrix  $\mathbf{M}$ , we let  $\mu_{\mathbf{M}}$  denote the empirical measure built from the eigenvalues of  $\mathbf{M}$  and  $\nu_{\mathbf{M}}$  denote the symmetric empirical measure built from the singular values of  $\mathbf{M}$ . That is,

$$\mu_{\mathbf{M}} := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(\mathbf{M})} \quad \text{and} \quad \nu_{\mathbf{M}} := \frac{1}{2n} \sum_{i=1}^n (\delta_{\sigma_i(\mathbf{M})} + \delta_{-\sigma_i(\mathbf{M})}),$$

where  $\lambda_1(\mathbf{M}), \dots, \lambda_n(\mathbf{M}) \in \mathbb{C}$  are the eigenvalues of  $\mathbf{M}$  and  $\sigma_1(\mathbf{M}) \geq \dots \geq \sigma_n(\mathbf{M})$  are the singular values of  $\mathbf{M}$ . Recall that  $F^{\mathbf{M}}$  is the ESD of  $\mathbf{M}$ . In particular, we have

$$F^{\mathbf{M}}(x, y) = \int_{-\infty}^x \int_{-\infty}^y \mu_{\mathbf{M}}(z) dt ds,$$

where  $z = s + \sqrt{-1}t$ .

Many of the techniques used to study Hermitian matrices fail to work for non-Hermitian matrices [3, Section 11.1]. Consider a  $n \times n$  non-Hermitian matrix  $\mathbf{M}$ . In [14, 15], Girko introduced a natural connection between  $\mu_{\mathbf{M}}$  and the collection of measures  $\{\nu_{\mathbf{M}-z\mathbf{I}}\}_{z \in \mathbb{C}}$ . Formally, we present this connection as Lemma 7.1 below.

Lemma 7.1 follows from [3, Lemma 11.2] and is based on Girko's original observation [14, 15]. The lemma has appeared in a number of different forms; for example, see [6, Lemma 4.3] and [16].

**Lemma 7.1** (Lemma 11.2 from [3]). *Let  $\mathbf{M}$  be a  $n \times n$  matrix. For any  $uv \neq 0$ , we have*

$$\begin{aligned} & \iint e^{\sqrt{-1}ux + \sqrt{-1}uy} F^{\mathbf{M}}(dx, dy) \\ &= \frac{u^2 + v^2}{4\sqrt{-1}u\pi} \iint \frac{\partial}{\partial s} \left[ \int_0^\infty \ln |x|^2 \nu_{\mathbf{M}-z\mathbf{I}}(dx) \right] e^{\sqrt{-1}us + \sqrt{-1}vt} dt ds, \end{aligned}$$

where  $z = s + \sqrt{-1}t$ .

We define the function

$$g_{\mathbf{M}}(s, t) := \frac{\partial}{\partial s} \int_0^\infty \log |x|^2 \nu_{\mathbf{M}-z\mathbf{I}}(dx), \quad (7.1)$$

where  $z = s + \sqrt{-1}t$ . We also define

$$g(s, t) := \begin{cases} \frac{2s}{s^2 + t^2}, & \text{if } s^2 + t^2 > 1 \\ 2s, & \text{otherwise} \end{cases}. \quad (7.2)$$

Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ . We define the  $2dn \times 2dn$  Hermitian matrix

$$\mathbf{H}_n = \mathbf{H}_n(z) := \begin{bmatrix} \mathbf{0} & \frac{1}{\sqrt{n}} \mathbf{X}_n - z\mathbf{I} \\ \frac{1}{\sqrt{n}} \mathbf{X}_n^* - \bar{z}\mathbf{I} & \mathbf{0} \end{bmatrix}$$

for  $z \in \mathbb{C}$ . It is straight-forward to verify that the eigenvalues of  $\mathbf{H}_n$  are given by

$$\pm \sigma_1 \left( \frac{1}{\sqrt{n}} \mathbf{X}_n - z\mathbf{I} \right), \pm \sigma_2 \left( \frac{1}{\sqrt{n}} \mathbf{X}_n - z\mathbf{I} \right), \dots, \pm \sigma_{dn} \left( \frac{1}{\sqrt{n}} \mathbf{X}_n - z\mathbf{I} \right).$$

In other words,  $\nu_{\frac{1}{\sqrt{n}} \mathbf{X}_n - z\mathbf{I}}$  is the empirical spectral measure of  $\mathbf{H}_n$ . By Lemma 7.1, the problem of studying  $\mu_{\frac{1}{\sqrt{n}} \mathbf{X}_n}$  reduces to studying the eigenvalue distribution of  $\mathbf{H}_n$ .

**7.2. Truncation.** In practice, it will be more convenient to work with a truncated version of  $\mathbf{H}_n$ . That is, we will work with a new matrix  $\hat{\mathbf{H}}_n$  whose entries are truncated versions of the entries of the original matrix  $\mathbf{H}_n$ . This subsection is devoted to the following standard truncation arguments.

Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume

$$m_{2+\eta} := \max_{1 \leq s, t \leq d} \mathbb{E} |\xi_{st}|^{2+\eta} < \infty, \quad (7.3)$$

for some  $\eta > 0$ . Let  $\delta > 0$ . For each  $s, t \in \{1, \dots, d\}$ , we define

$$\tilde{\xi}_{st}^{(n)} := \xi_{st} \mathbf{1}_{\{|\xi_{st}| \leq n^\delta\}} - \mathbb{E} [\xi_{st} \mathbf{1}_{\{|\xi_{st}| \leq n^\delta\}}] \quad \text{and} \quad \hat{\xi}_{st}^{(n)} := \frac{\tilde{\xi}_{st}^{(n)}}{\sqrt{d \operatorname{Var}(\tilde{\xi}_{st}^{(n)})}}.$$

Here  $\mathbf{1}_E$  denotes the indicator function of the event  $E$ . We present the following standard truncation lemma.

**Lemma 7.3.** *Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume (7.3) holds for some  $\eta > 0$ . For each  $\delta > 0$ , there exists  $n_0 > 0$  such that the following holds for all  $n > n_0$ .*

- (i) For each  $s, t \in \{1, \dots, d\}$ ,  $\hat{\xi}_{st}^{(n)}$  has mean zero and variance  $1/d$ .
- (ii) a.s.  $\max_{1 \leq s, t \leq d} \left| \hat{\xi}_{st}^{(n)} \right| \leq 4n^\delta$ .
- (iii) We have

$$\max_{1 \leq s, t \leq d} \left| 1/d - \operatorname{Var}(\tilde{\xi}_{st}^{(n)}) \right| \leq 2 \frac{m_{2+\eta}}{n^{\delta\eta}}.$$

- (iv) We have

$$\max_{(s,t) \neq (u,v)} \left| \mathbb{E} \left[ \hat{\xi}_{st}^{(n)} \overline{\hat{\xi}_{uv}^{(n)}} \right] \right| \leq 10 \frac{\sqrt{m_{2+\eta}}}{n^{\delta\eta/2}}.$$

*Proof.* (of Lemma 7.3) We first note that

$$\operatorname{Var}(\tilde{\xi}_{st}^{(n)}) \leq \mathbb{E} |\xi_{st}|^2 \mathbf{1}_{\{|\xi_{st}| \leq n^\delta\}} \leq \mathbb{E} |\xi_{st}|^2 = 1/d \quad (7.4)$$

for all  $s, t \in \{1, \dots, d\}$ . We also note that

$$\left| 1/d - \operatorname{Var}(\tilde{\xi}_{st}^{(n)}) \right| = \left| 1/d - \mathbb{E} \left| \tilde{\xi}_{st}^{(n)} \right|^2 \right| \leq 2 \mathbb{E} |\xi_{st}|^2 \mathbf{1}_{\{|\xi_{st}| > n^\delta\}} \leq 2 \frac{m_{2+\eta}}{n^{\delta\eta}}. \quad (7.5)$$

Since this holds for all  $s, t \in \{1, \dots, d\}$ , we obtain (iii). We now take  $n_0$  sufficiently large such that

$$8m_{2+\eta} \leq n_0^{\delta\eta} \quad (7.6)$$

and

$$\min_{1 \leq s, t \leq d} \operatorname{Var}(\tilde{\xi}_{st}^{(n)}) > \frac{1}{2d}$$

for all  $n > n_0$ ; let  $n > n_0$ . Then each  $\hat{\xi}_{st}^{(n)}$  has mean zero and variance  $1/d$  by construction. Moreover, we have a.s.

$$\left| \hat{\xi}_{st}^{(n)} \right| \leq \frac{2n^\delta}{\sqrt{d \operatorname{Var}(\tilde{\xi}_{st}^{(n)})}} \leq 4n^\delta$$

for all  $s, t \in \{1, \dots, d\}$ . We now verify (iv); fix  $s, t, u, v \in \{1, \dots, d\}$  with  $(s, t) \neq (u, v)$ . We have

$$\begin{aligned} \left| \mathbb{E} \left[ \hat{\xi}_{st}^{(n)} \overline{\hat{\xi}_{uv}^{(n)}} \right] - \mathbb{E} \left[ \tilde{\xi}_{st}^{(n)} \overline{\tilde{\xi}_{uv}^{(n)}} \right] \right| &\leq \mathbb{E} \left| \hat{\xi}_{st}^{(n)} \overline{\hat{\xi}_{uv}^{(n)}} \right| \left| 1 - d \sqrt{\operatorname{Var}(\tilde{\xi}_{st}^{(n)})} \sqrt{\operatorname{Var}(\tilde{\xi}_{uv}^{(n)})} \right| \\ &\leq \left| \frac{1}{d} - \sqrt{\operatorname{Var}(\tilde{\xi}_{st}^{(n)})} \sqrt{\operatorname{Var}(\tilde{\xi}_{uv}^{(n)})} \right| \\ &\leq d \left| \frac{1}{d^2} - \operatorname{Var}(\tilde{\xi}_{st}^{(n)}) \operatorname{Var}(\tilde{\xi}_{uv}^{(n)}) \right| \\ &\leq \left| \operatorname{Var}(\tilde{\xi}_{st}^{(n)}) - \frac{1}{d} \right| + \left| \operatorname{Var}(\tilde{\xi}_{uv}^{(n)}) - \frac{1}{d} \right| \\ &\leq 4 \frac{m_{2+\eta}}{n^{\delta\eta}} \end{aligned}$$

by (7.4) and the Cauchy-Schwarz inequality. Here the last inequality follows from (7.5). By another application of Cauchy-Schwarz and (7.4), we obtain

$$\begin{aligned} \left| \mathbb{E} \left[ \begin{array}{c} \tilde{\xi}_{st}^{(n)} \\ \tilde{\xi}_{uv}^{(n)} \end{array} \right] \right| &\leq \sqrt{\mathbb{E} |\xi_{st}|^2 \mathbf{1}_{\{|\xi_{st}| > n^\delta\}}} + \sqrt{\mathbb{E} |\xi_{uv}|^2 \mathbf{1}_{\{|\xi_{uv}| > n^\delta\}}} \\ &\quad + 2\mathbb{E} |\xi_{st}|^2 \mathbf{1}_{\{|\xi_{st}| > n^\delta\}} + 2\mathbb{E} |\xi_{uv}|^2 \mathbf{1}_{\{|\xi_{uv}| > n^\delta\}} \\ &\leq 2\frac{\sqrt{m_{2+\eta}}}{n^{\delta\eta/2}} + 4\frac{m_{2+\eta}}{n^{\delta\eta}}. \end{aligned}$$

Combining the bounds above with (7.6), we obtain

$$\left| \mathbb{E} \left[ \begin{array}{c} \hat{\xi}_{st}^{(n)} \\ \hat{\xi}_{uv}^{(n)} \end{array} \right] \right| \leq 10\frac{\sqrt{m_{2+\eta}}}{n^{\delta\eta/2}}. \quad (7.7)$$

Since (7.7) holds for any  $(s, t) \neq (u, v)$ , the proof of the lemma is complete.  $\square$

We will continue to use the notation introduced in Definition 2.6. That is, for any  $s, t \in \{1, \dots, d\}$  and  $1 \leq i, j \leq n$ , we let  $x_{st;ij}$  denote the  $(i, j)$ -entry of the matrix  $\mathbf{X}_{n,st}$ . For every  $s, t \in \{1, \dots, d\}$ ,  $n \geq 1$ , and  $1 \leq i, j \leq n$ , we define

$$\tilde{x}_{st;ij}^{(n)} := x_{st;ij} \mathbf{1}_{\{|x_{st;ij}| \leq n^\delta\}} - \mathbb{E} [x_{st;ij} \mathbf{1}_{\{|x_{st;ij}| \leq n^\delta\}}]$$

and

$$\hat{x}_{st;ij}^{(n)} := \frac{\tilde{x}_{st;ij}^{(n)}}{\sqrt{d \operatorname{Var}(\tilde{x}_{st;ij}^{(n)})}}.$$

Set  $\tilde{\mathbf{X}}_{n,st} := \left( \tilde{x}_{st;ij}^{(n)} \right)_{i,j=1}^n$  and  $\hat{\mathbf{X}}_{n,st} := \left( \hat{x}_{st;ij}^{(n)} \right)_{i,j=1}^n$  for every  $n \geq 1$  and  $s, t \in \{1, \dots, d\}$ . We also define the  $dn \times dn$  random block matrices

$$\tilde{\mathbf{X}}_n := \left( \tilde{\mathbf{X}}_{n,st} \right)_{s,t=1}^d, \quad \hat{\mathbf{X}}_n := \left( \hat{\mathbf{X}}_{n,st} \right)_{s,t=1}^d.$$

For  $z \in \mathbb{C}$ , we define the  $2dn \times 2dn$  matrices

$$\tilde{\mathbf{H}}_n = \tilde{\mathbf{H}}_n(z) := \begin{bmatrix} \mathbf{0} & \frac{1}{\sqrt{n}} \tilde{\mathbf{X}}_n - z\mathbf{I} \\ \frac{1}{\sqrt{n}} \tilde{\mathbf{X}}_n^* - \bar{z}\mathbf{I} & \mathbf{0} \end{bmatrix}$$

and

$$\hat{\mathbf{H}}_n = \hat{\mathbf{H}}_n(z) := \begin{bmatrix} \mathbf{0} & \frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n - z\mathbf{I} \\ \frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n^* - \bar{z}\mathbf{I} & \mathbf{0} \end{bmatrix}.$$

We will make use of the following corollary to the law of large numbers.

**Lemma 7.4** (Law of large numbers). *Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of random matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume (7.3) holds for some  $\eta > 0$ . Let  $\delta > 0$ . Then a.s.*

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \|\mathbf{X}_n\|_2^2 \leq d, \quad (7.8)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \|\hat{\mathbf{X}}_n\|_2^2 \leq 8d, \quad (7.9)$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n |x_{st;ij}|^{2+\eta} \mathbf{1}_{\{|x_{st;ij}| > n^\delta\}} = 0. \quad (7.10)$$

*Proof.* (of Lemma 7.4) We first prove (7.8). We begin by noting that

$$\frac{1}{n^2} \|\mathbf{X}_n\|_2^2 = \sum_{s,t=1}^d \frac{1}{n^2} \sum_{i,j=1}^n |x_{st;ij}|^2.$$

For any  $s, t \in \{1, \dots, d\}$ , we apply the law of large numbers and obtain a.s.

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i,j=1}^n |x_{st;ij}|^2 \leq \mathbb{E}|\xi_{st}|^2 = \frac{1}{d}.$$

Since  $d$  is fixed, independent of  $n$ , we conclude that a.s.

$$\limsup_{n \rightarrow \infty} \sum_{s,t=1}^d \frac{1}{n^2} \sum_{i,j=1}^n |x_{st;ij}|^2 \leq d,$$

and the proof of (7.8) is complete.

For (7.9), we apply the bounds in Lemma 7.3 to obtain

$$\sum_{s,t=1}^d \sum_{i,j=1}^n \left| \hat{x}_{st;ij}^{(n)} \right|^2 \leq 2 \sum_{s,t=1}^d \sum_{i,j=1}^n \left| \tilde{x}_{st;ij}^{(n)} \right|^2 \leq 4 \sum_{s,t=1}^d \sum_{i,j=1}^n (|x_{st;ij}|^2 + \mathbb{E}|x_{st;ij}|^2)$$

for  $n$  sufficiently large. Hence (7.9) follows from (7.8).

We now prove (7.10); fix  $s, t \in \{1, \dots, d\}$ . By the law of large numbers, for any  $M > 0$ , we have a.s.

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i,j=1}^n |x_{st;ij}|^{2+\eta} \mathbf{1}_{\{|x_{st;ij}| > n^\delta\}} \leq \mathbb{E}|\xi_{st}|^{2+\eta} \mathbf{1}_{\{|\xi_{st}| > M\}}.$$

By the dominated convergence theorem, it follows that

$$\lim_{M \rightarrow \infty} \mathbb{E}|\xi_{st}|^{2+\eta} \mathbf{1}_{\{|\xi_{st}| > M\}} = 0.$$

We conclude that a.s.

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i,j=1}^n |x_{st;ij}|^{2+\eta} \mathbf{1}_{\{|x_{st;ij}| > n^\delta\}} = 0.$$

Since  $d$  is fixed, independent of  $n$ , the claim follows.  $\square$

We let  $L(F, G)$  denote the Levy distance between two distribution functions  $F, G$ . That is,

$$L(F, G) := \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R}\}. \quad (7.11)$$

Convergence in Levy distance implies convergence in distribution [3, Remark A.40]. We will compare the ESD of  $\hat{\mathbf{H}}_n$  to the ESD of  $\mathbf{H}_n$  using the Levy metric.

**Lemma 7.5.** *Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of random matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume (7.3) holds for some  $\eta > 0$ . Let  $\delta > 0$ . Then a.s.*

$$\sup_{z \in \mathbb{C}} L(F^{\mathbf{H}_n}, F^{\hat{\mathbf{H}}_n}) = o(n^{-\delta\eta/3}).$$

*Proof.* (of Lemma 7.5) We will apply [3, Corollary A.41] to bound  $L(F^{\mathbf{H}_n}, F^{\tilde{\mathbf{H}}_n})$  and  $L(F^{\tilde{\mathbf{H}}_n}, F^{\hat{\mathbf{H}}_n})$  separately. Thus, we have

$$\begin{aligned} \sup_{z \in \mathbb{C}} n^{\delta\eta} L^3(F^{\mathbf{H}_n}, F^{\tilde{\mathbf{H}}_n}) &\leq \frac{n^{\delta\eta}}{n^2} \left\| \mathbf{X}_n - \tilde{\mathbf{X}}_n \right\|_2^2 \\ &\leq 2 \frac{n^{\delta\eta}}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n \left( |x_{st,ij}|^2 \mathbf{1}_{\{|x_{st,ij}| > n^\delta\}} + \mathbb{E}|x_{st,ij}|^2 \mathbf{1}_{\{|x_{st,ij}| > n^\delta\}} \right) \\ &\leq \frac{2}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n \left( |x_{st,ij}|^{2+\eta} \mathbf{1}_{\{|x_{st,ij}| > n^\delta\}} + \mathbb{E}|x_{st,ij}|^{2+\eta} \mathbf{1}_{\{|x_{st,ij}| > n^\delta\}} \right). \end{aligned}$$

We note that

$$\frac{1}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n \mathbb{E}|x_{st,ij}|^{2+\eta} \mathbf{1}_{\{|x_{st,ij}| > n^\delta\}} = \sum_{s,t=1}^n \mathbb{E}|\xi_{st}|^{2+\eta} \mathbf{1}_{\{|\xi_{st}| > n^\delta\}},$$

and thus, by the dominated convergence theorem, we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n \mathbb{E}|x_{st,ij}|^{2+\eta} \mathbf{1}_{\{|x_{st,ij}| > n^\delta\}} = 0.$$

Therefore, by Lemma 7.4, we conclude that a.s.

$$\lim_{n \rightarrow \infty} \sup_{z \in \mathbb{C}} n^{\delta\eta} L^3(F^{\mathbf{H}_n}, F^{\tilde{\mathbf{H}}_n}) = 0,$$

and hence a.s.

$$\sup_{z \in \mathbb{C}} L(F^{\mathbf{H}_n}, F^{\tilde{\mathbf{H}}_n}) = o(n^{-\delta\eta/3}). \quad (7.12)$$

Applying [3, Corollary A.41] again, we obtain

$$\begin{aligned} \sup_{z \in \mathbb{C}} n^{\delta\eta} L^3(F^{\tilde{\mathbf{H}}_n}, F^{\hat{\mathbf{H}}_n}) &\leq \frac{n^{\delta\eta}}{n^2} \left\| \tilde{\mathbf{X}}_n - \hat{\mathbf{X}}_n \right\|_2^2 \\ &\leq \frac{n^{\delta\eta}}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n \left| \hat{x}_{st,ij}^{(n)} \right|^2 \left| 1 - \sqrt{d \operatorname{Var}(\hat{x}_{st,ij}^{(n)})} \right|^2 \\ &\leq \frac{n^{\delta\eta}}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n \left| \hat{x}_{st,ij}^{(n)} \right|^2 \left| 1 - d \operatorname{Var}(\hat{x}_{st,ij}^{(n)}) \right|^2 \\ &\leq d^2 \frac{n^{\delta\eta}}{n^2} \sum_{s,t=1}^d \sum_{i,j=1}^n \left| \hat{x}_{st,ij}^{(n)} \right|^2 \left| \frac{1}{d} - \operatorname{Var}(\tilde{\xi}_{st}^{(n)}) \right|^2 \\ &\leq 4d^2 \frac{m_{2+\eta}^2}{n^{\delta\eta}} \frac{1}{n^2} \left\| \hat{\mathbf{X}}_n \right\|_2^2. \end{aligned}$$

Here the last inequality follows from Lemma 7.3. By Lemma 7.4, we have a.s.

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2+\delta\eta}} \left\| \hat{\mathbf{X}}_n \right\|_2^2 = 0,$$

and we conclude that a.s.

$$\sup_{z \in \mathbb{C}} L(F^{\tilde{\mathbf{H}}_n}, F^{\hat{\mathbf{H}}_n}) = o(n^{-\delta\eta/3}). \quad (7.13)$$

The claim now follows from (7.12), (7.13), and the triangle inequality for Levy distance.  $\square$

**7.6. Cubic Relation.** We now consider the distribution of eigenvalues of  $\mathbf{H}_n$ . In fact, by Lemma 7.5, it will suffice to consider the eigenvalues of  $\hat{\mathbf{H}}_n$ . To this end, we will study the resolvent of  $\hat{\mathbf{H}}_n$  in Theorem 7.8 below. Indeed, for  $w \in \mathbb{C}$  with  $\text{Im}(w) > 0$ ,

$$\hat{m}_n(z, w) := \frac{1}{2dn} \text{tr} \left( \hat{\mathbf{H}}_n(z) - w\mathbf{I} \right)^{-1} = \int_{\mathbb{R}} \frac{1}{x - w} \nu_{\frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n - z\mathbf{I}}(dx)$$

is the Stieltjes transform of the measure  $\nu_{\frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n - z\mathbf{I}}$ . It follows from standard Stieltjes transform techniques (e.g. [3, Theorem B.9]) that computing the limiting ESD of  $\hat{\mathbf{H}}_n(z)$  is equivalent to computing the limit of  $\hat{m}_n(z, w)$  for all  $w \in \mathbb{C}$  with  $\text{Im}(w) > 0$ .

As is standard in random matrix theory, we will not compute  $\hat{m}_n$  explicitly. Instead we will derive a fixed point equation. Indeed, we will show

$$\hat{m}_n(z, w) + \frac{\hat{m}_n(z, w) + w}{(\hat{m}_n(z, w) + w)^2 - |z|^2} = o(1) \quad (7.14)$$

for  $z, w \in \mathbb{C}$  with  $\text{Im}(w) > 0$ . We will then conclude that  $\hat{m}_n(z, w)$  converges to a limit, which we denote by  $m(z, w)$ . It follows that  $m(z, w)$  satisfies the equation

$$m(z, w) + \frac{m(z, w) + w}{(m(z, w) + w)^2 - |z|^2} = 0. \quad (7.15)$$

From (7.15) we will deduce the limiting ESD of  $\hat{\mathbf{H}}_n$ . Equation (7.15) has appeared previously in [6, 17] and in a slightly different form in [3, Chapter 11]. We refer to equation (7.15) as a cubic relation since it can be rewritten as the cubic polynomial equation

$$m(z, w)^3 + 2wm(z, w)^2 + (w^2 - |z|^2 + 1)m(z, w) + w = 0.$$

In this subsection we will show  $\hat{m}_n$  satisfies (7.14). We begin with the following concentration result for bilinear forms from [29].

**Lemma 7.7** (Lemma 3.10 of [29]). *Let  $(x, y)$  be a random vector in  $\mathbb{C}^2$  where  $x, y$  both have mean zero, unit variance, and satisfy*

- $\max\{|x|, |y|\} \leq L$  a.s.,
- $\mathbb{E}[\bar{x}y] = \rho$ .

*Let  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  be iid copies of  $(x, y)$ , and set  $X = (x_1, x_2, \dots, x_n)^T$  and  $Y = (y_1, y_2, \dots, y_n)^T$ . Let  $\mathbf{B}$  be a  $n \times n$  random matrix, independent of  $X$  and  $Y$ , which satisfies  $\|\mathbf{B}\| \leq n^{1/4}$  a.s. Then, for any  $p \geq 2$ ,*

$$\mathbb{P} \left( \left| \frac{1}{n} X^* \mathbf{B} Y - \frac{\rho}{n} \text{tr} \mathbf{B} \right| > n^{-1/8} \right) = O_p \left( \frac{L^{2p}}{n^{p/8}} \right).$$

We formally establish (7.14) in the following theorem.

**Theorem 7.8.** *Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of random matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume (7.3) holds for some  $\eta > 0$ . Let  $0 < \delta < 1/100$ . Consider the truncated random matrices  $\{\hat{\mathbf{X}}_n\}_{n \geq 1}$  and  $\{\hat{\mathbf{H}}_n(z)\}_{n \geq 1}$ . For  $z, w \in \mathbb{C}$  with  $\text{Im}(w) > 0$ , define*

$$\hat{\mathbf{G}}_n(z, w) := \left( \hat{\mathbf{H}}_n(z) - w\mathbf{I} \right)^{-1} \quad \text{and} \quad \hat{m}_n(z, w) := \frac{1}{2dn} \text{tr} \hat{\mathbf{G}}_n(z, w).$$

*Let  $M, \beta > 0$ . Then, for  $v_n := \max\{n^{-\eta\delta/100}, n^{-1/100}\}$ , a.s.*

$$\sup_{|z| \leq M} \sup_{|w| \leq \beta, \text{Im}(w) \geq v_n} \left| \hat{m}_n(z, w) + \frac{\hat{m}_n(z, w) + w}{(\hat{m}_n(z, w) + w)^2 - |z|^2} \right| = O_{M, \beta, d, m_{2+\eta}}(v_n^5).$$

In order to prove Theorem 7.8, we will need the following deterministic lemmas.

**Lemma 7.9.** *Let  $\mathbf{R}$  be the  $2n \times 2n$  block matrix given by*

$$\mathbf{R} = \begin{bmatrix} -w\mathbf{I} & \mathbf{B} \\ \mathbf{B}^* & -w\mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{R}_3 & \mathbf{R}_4 \end{bmatrix},$$

where  $\mathbf{B}, \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \mathbf{R}_4$  are  $n \times n$  matrices. Then  $\text{tr } \mathbf{R}_1 = \text{tr } \mathbf{R}_4$  for any  $w \in \mathbb{C}$  with  $\text{Im}(w) > 0$ .

*Proof.* (of Lemma 7.9) We first note that

$$\begin{bmatrix} -w\mathbf{I} & \mathbf{B} \\ \mathbf{B}^* & -w\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{0} \end{bmatrix} - w\mathbf{I}$$

is invertible for any  $w \in \mathbb{C}$  with  $\text{Im}(w) > 0$ . Let  $\sigma_1, \sigma_2, \dots, \sigma_n \geq 0$  denote the singular values of  $\mathbf{B}$ . Then  $-w\mathbf{I} + w^{-1}\mathbf{B}\mathbf{B}^*$  has eigenvalues  $-w + w^{-1}\sigma_i^2$  for  $i = 1, 2, \dots, n$ . In particular

$$\text{Im} \left( -w + \frac{\sigma_i^2}{w} \right) = -\text{Im}(w) - \text{Im}(w) \frac{\sigma_i^2}{|w|^2} < 0$$

for  $\text{Im}(w) > 0$ . Thus  $-w\mathbf{I} + w^{-1}\mathbf{B}\mathbf{B}^*$  is invertible. Similarly,  $-w\mathbf{I} + w^{-1}\mathbf{B}^*\mathbf{B}$  has the same eigenvalues and is also invertible. By the Schur complement [20, Section 0.7.3],

$$\begin{aligned} \mathbf{R}_1 &= (-w\mathbf{I} + w^{-1}\mathbf{B}\mathbf{B}^*)^{-1}, \\ \mathbf{R}_4 &= (-w\mathbf{I} + w^{-1}\mathbf{B}^*\mathbf{B})^{-1}. \end{aligned}$$

Since  $\mathbf{R}_1$  and  $\mathbf{R}_4$  have the same eigenvalues,  $\text{tr } \mathbf{R}_1 = \text{tr } \mathbf{R}_4$ . □

We introduce  $\varepsilon$ -nets as a convenient way to discretize a compact set. Let  $\varepsilon > 0$ . A set  $X$  is an  $\varepsilon$ -net of a set  $Y$  if for any  $y \in Y$ , there exists  $x \in X$  such that  $\|x - y\| \leq \varepsilon$ . In order to prove Theorem 7.8, we will need the following well-known estimate for the maximum size of an  $\varepsilon$ -net.

**Lemma 7.10** (Lemma 3.11 of [29]). *The set  $\{w \in \mathbb{C} : |w| \leq \beta, \text{Im}(w) \geq \alpha\}$  admits an  $\varepsilon$ -net of size at most*

$$\left( 1 + \frac{2\beta}{\varepsilon} \right)^2.$$

We will also take advantage of the following facts, which can be found in [19, 20]. Let  $\mathbf{B}$  be a  $n \times n$  matrix with singular values  $\sigma_1(\mathbf{B}) \geq \dots \geq \sigma_n(\mathbf{B}) \geq 0$ . Then the  $2n \times 2n$  matrix

$$\begin{bmatrix} -w\mathbf{I} & \mathbf{B} \\ \mathbf{B}^* & -w\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{0} \end{bmatrix} - w\mathbf{I} \tag{7.16}$$

has an orthonormal basis of eigenvectors with eigenvalues  $\pm\sigma_1(\mathbf{B}) - w, \dots, \pm\sigma_n(\mathbf{B}) - w$ . Thus, if  $\text{Im}(w) > 0$ , the matrix (7.16) is invertible. Let

$$\mathbf{G} := \begin{bmatrix} -w\mathbf{I} & \mathbf{B} \\ \mathbf{B}^* & -w\mathbf{I} \end{bmatrix}^{-1}.$$

Then, for  $\text{Im}(w) > 0$ , we have

$$\|\mathbf{G}\| = \max_{1 \leq i \leq n} \frac{1}{|\pm\sigma_i(\mathbf{B}) - w|} \leq \frac{1}{\text{Im}(w)}. \tag{7.17}$$

We will make use of the following identity: for any invertible  $n \times n$  matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}. \tag{7.18}$$



A special case of (7.18) is the resolvent identity (also known as Hilbert's identity): for any Hermitian  $n \times n$  matrix  $\mathbf{A}$ ,

$$(\mathbf{A} - w\mathbf{I})^{-1} - (\mathbf{A} - w'\mathbf{I})^{-1} = (w - w')(\mathbf{A} - w\mathbf{I})^{-1}(\mathbf{A} - w'\mathbf{I})^{-1} \quad (7.19)$$

for all  $w, w' \in \mathbb{C}$  with  $\text{Im}(w), \text{Im}(w') > 0$ .

We are now ready to prove Theorem 7.8.

*Proof.* (of Theorem 7.8) Fix  $M, \beta > 0$ . For the remainder of the proof, the implicit constants in our asymptotic notation (such as  $O, o, \Omega, \ll$ ) depend only on the constants  $M, \beta, d$ , and  $m_{2+\eta}$ ; for simplicity, we no longer include these subscripts in our notation.

For notational convenience, we will write  $\mathbf{X}_n$  instead of  $\hat{\mathbf{X}}_n$ . That is, we let  $\{\mathbf{X}_n\}_{n \geq 1}$  denote the sequence of truncated matrices. Similarly, we write  $\mathbf{X}_{n,st}$  instead of  $\hat{\mathbf{X}}_{n,st}$  for  $s, t \in \{1, \dots, d\}$ . We define the  $2dn \times 2dn$  matrix

$$\mathbf{G}_n(z, w) := \begin{bmatrix} -w\mathbf{I} & \frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I} \\ \frac{1}{\sqrt{n}}\mathbf{X}_n^* - \bar{z}\mathbf{I} & -w\mathbf{I} \end{bmatrix}^{-1}.$$

We will often drop the dependence on  $z, w$  and simply write  $\mathbf{G}_n$  instead of  $\mathbf{G}_n(z, w)$ . We write  $\mathbf{G}_n = (\mathbf{G}_{n,st})_{s,t=1}^{2d}$  where each  $\mathbf{G}_{n,st}$  is a  $n \times n$  matrix. Then  $\mathbf{G}_{n,st}(i, j)$  denotes the  $(i, j)$ -entry of  $\mathbf{G}_{n,st}$ . We define  $m_{n,st}(z, w) := \frac{1}{n} \text{tr} \mathbf{G}_{n,st}$  for  $s, t \in \{1, \dots, 2d\}$  and  $m_n(z, w) := \frac{1}{2dn} \text{tr} \mathbf{G}_n$ . We will often drop the dependence on  $z, w$  and simply write  $m_n$  and  $m_{n,st}$  instead of  $m_n(z, w)$  and  $m_{n,st}(z, w)$ .

Let  $1 \leq k \leq n$ . We let  $\mathbf{r}_k(\mathbf{X}_{n,st})$  denote the  $k$ -th row of  $\mathbf{X}_{n,st}$  with the  $k$ -th entry removed. Similarly, we let  $\mathbf{c}_k(\mathbf{X}_{n,st})$  denote the  $k$ -th column of  $\mathbf{X}_{n,st}$  with the  $k$ -th entry removed. We let  $\mathbf{X}_{n,st}^{(k)}$  be the  $(n-1) \times (n-1)$  matrix constructed from  $\mathbf{X}_{n,st}$  by removing the  $k$ -th column and  $k$ -th row. We let  $\mathbf{X}_n^{(k)}$  be the  $d(n-1) \times d(n-1)$  block matrix given by  $\mathbf{X}_n^{(k)} := \left( \mathbf{X}_{n,st}^{(k)} \right)_{s,t=1}^d$ . Define the  $2d(n-1) \times 2d(n-1)$  matrix

$$\mathbf{G}_n^{(k)}(z, w) := \begin{bmatrix} -w\mathbf{I} & \frac{1}{\sqrt{n}}\mathbf{X}_n^{(k)} - z\mathbf{I} \\ \frac{1}{\sqrt{n}}\mathbf{X}_n^{(k)*} - \bar{z}\mathbf{I} & -w\mathbf{I} \end{bmatrix}^{-1}.$$

We will often drop the dependence on  $z, w$  and simply write  $\mathbf{G}_n^{(k)}$ . We again write  $\mathbf{G}_n^{(k)} = \left( \mathbf{G}_{n,st}^{(k)} \right)_{s,t=1}^{2d}$  where each  $\mathbf{G}_{n,st}^{(k)}$  is a  $(n-1) \times (n-1)$  matrix. We let  $m_{n,st}^{(k)}(z, w) := \frac{1}{n} \text{tr} \mathbf{G}_{n,st}^{(k)}$  for  $s, t \in \{1, \dots, 2d\}$  and  $m_n^{(k)}(z, w) := \frac{1}{2d(n-1)} \text{tr} \mathbf{G}_n^{(k)}$ . We will often drop the dependence on  $z, w$  and write  $m_n^{(k)}$  and  $m_{n,st}^{(k)}$  instead of  $m_n^{(k)}(z, w)$  and  $m_{n,st}^{(k)}(z, w)$ .

From (7.17), for  $w \in \mathbb{C}$  with  $\text{Im}(w) \geq v_n$ , we have the deterministic bounds  $\|\mathbf{G}_n(z, w)\| \leq v_n^{-1}$ ,  $|m_n(z, w)| \leq v_n^{-1}$ , and  $|m_{n,st}(z, w)| \leq v_n^{-1}$ . By Cauchy's interlacing theorem [20, Theorem 4.3.8] (or alternatively [3, (A.1.12)]), we have the deterministic bound

$$\sup_{1 \leq k \leq n} \sup_{|z| \leq M} \sup_{|w| \leq \beta, \text{Im}(w) \geq v_n} \left| m_n^{(k)}(z, w) - m_n(z, w) \right| = O\left(\frac{1}{v_n n}\right). \quad (7.20)$$

By Lemma 7.9, we have

$$\sum_{s=1}^d m_{n,ss}(z, w) = \sum_{s=d+1}^{2d} m_{n,ss}(z, w)$$

and

$$\sum_{s=1}^d m_{n,ss}^{(k)}(z, w) = \sum_{s=d+1}^{2d} m_{n,ss}^{(k)}(z, w)$$

for any  $1 \leq k \leq n$ . Thus, from (7.20), we find

$$\sup_{1 \leq k \leq n} \sup_{|z| \leq M} \sup_{|w| \leq \beta, \text{Im}(w) \geq v_n} \left| \frac{1}{d} \sum_{s=1}^d m_{n,ss}^{(k)}(z, w) - m_n(z, w) \right| = O\left(\frac{1}{nv_n}\right) \quad (7.21)$$

and

$$\sup_{1 \leq k \leq n} \sup_{|z| \leq M} \sup_{|w| \leq \beta, \text{Im}(w) \geq v_n} \left| \frac{1}{d} \sum_{s=d+1}^{2d} m_{n,ss}^{(k)}(z, w) - m_n(z, w) \right| = O\left(\frac{1}{nv_n}\right). \quad (7.22)$$

Fix  $1 \leq k \leq n$  and  $z \in \mathbb{C}$  with  $|z| \leq M$ . Fix  $w \in \mathbb{C}$  with  $|w| \leq \beta$  and  $\text{Im}(w) \geq v_n$ . Let  $\mathbf{Q}_k$  be the  $2d \times 2d$  matrix given by  $\mathbf{Q}_k := (\mathbf{G}_{n,st}(k, k))_{s,t=1}^{2d}$ . By the Schur complement [20, Section 0.7.3],

$$\mathbf{Q}_k^{-1} = \begin{bmatrix} -w\mathbf{I} & \left(\frac{1}{\sqrt{n}}x_{st;kk} - z\delta_{s,t}\right)_{s,t=1}^d \\ \left(\frac{1}{\sqrt{n}}\bar{x}_{st;kk} - \bar{z}\delta_{s,t}\right)_{s,t=1}^d & -w\mathbf{I} \end{bmatrix} - \frac{1}{n} \mathbf{R}_k \mathbf{G}_n^{(k)} \mathbf{R}_k^*,$$

where  $\delta_{s,t}$  is the Kronecker delta and

$$\mathbf{R}_k := \begin{bmatrix} \mathbf{0} & (\mathbf{r}_k(\mathbf{X}_{n,st}))_{s,t=1}^d \\ (\mathbf{c}_k(\mathbf{X}_{n,ts}))_{s,t=1}^d & \mathbf{0} \end{bmatrix}.$$

By the truncation assumption and Lemma 7.3, we have a.s.

$$\max_{1 \leq k \leq n} \max_{1 \leq s, t \leq d} \frac{|x_{st;kk}|}{\sqrt{n}} \ll \frac{n^\delta}{\sqrt{n}} \leq \frac{1}{n^{2/5}}$$

for  $n$  sufficiently large.

We observe that  $\mathbf{R}_k$  and  $\mathbf{G}_n^{(k)}$  are independent random matrices. By expanding the product, we note that the entries of  $\mathbf{R}_k \mathbf{G}_n^{(k)} \mathbf{R}_k^*$  are linear combinations of bilinear forms. We will apply Lemma 7.7 to control each bilinear form. Applying the bound  $\|\mathbf{G}_n(z, w)\| \leq v_n^{-1}$  and Lemma 7.7, we obtain

$$\left\| \frac{1}{n} \mathbf{R}_k \mathbf{G}_n^{(k)} \mathbf{R}_k^* - \begin{bmatrix} \left(\frac{1}{d} \sum_{s=d+1}^{2d} m_{n,ss}^{(k)}\right) \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \left(\frac{1}{d} \sum_{s=1}^d m_{n,ss}^{(k)}\right) \mathbf{I}_d \end{bmatrix} \right\| \ll n^{-1/8} + \frac{1}{n^{\delta\eta/2} v_n} \quad (7.23)$$

with probability  $1 - O(n^{-100})$ . Here we obtain the bound on the spectral norm by bounding each entry individually and noting that

$$\|\mathbf{B}\| \leq \|\mathbf{B}\|_2 \leq 2d \max_{i,j} |\mathbf{B}_{ij}|$$

for any  $2d \times 2d$  matrix  $\mathbf{B}$  (recall that the matrices above are  $2d \times 2d$ ). The bound (7.23) holds with probability  $1 - O(n^{-100})$  by taking  $p$  sufficiently large in Lemma 7.7. The factor  $1/d$  appears because the entries of  $\mathbf{R}_k$  have variance  $1/d$ . We also used (iv) from Lemma 7.3 and the deterministic bound  $\sup_{s,t \in \{1, \dots, 2d\}} |m_{n,st}^{(k)}| \leq \|\mathbf{G}_n^{(k)}\| \leq v_n^{-1}$ .

By (7.21), (7.22), and the union bound over  $1 \leq k \leq n$ , we obtain

$$\sup_{1 \leq k \leq n} \|\mathbf{Q}_k^{-1} - \mathbf{M}_n\| \ll n^{-1/8} + \frac{1}{n^{\delta\eta/2} v_n} + \frac{1}{nv_n} \quad (7.24)$$

with probability  $1 - O(n^{-99})$ , where

$$\mathbf{M}_n := \begin{bmatrix} -(m_n(z, w) + w)\mathbf{I}_d & -z\mathbf{I}_d \\ -\bar{z}\mathbf{I}_d & -(m_n(z, w) + w)\mathbf{I}_d \end{bmatrix}.$$

We note that

$$|m_n(z, w) + w| \geq \operatorname{Im}(m_n(z, w) + w) \geq \operatorname{Im}(w) \geq v_n > 0. \quad (7.25)$$

It follows that  $M_n$  is invertible and the inverse is given (in block form) by

$$\mathbf{M}_n^{-1} = \begin{bmatrix} \mathbf{M}_{n,1} & \mathbf{M}_{n,2} \\ \mathbf{M}_{n,3} & \mathbf{M}_{n,4} \end{bmatrix},$$

where  $\mathbf{M}_{n,1}, \mathbf{M}_{n,2}, \mathbf{M}_{n,3}, \mathbf{M}_{n,4}$  are  $d \times d$  matrices with

$$\begin{aligned} \mathbf{M}_{n,1} &= \mathbf{M}_{n,4} := -\frac{m_n(z, w) + w}{(m_n(z, w) + w)^2 - |z|^2} \mathbf{I}, \\ \mathbf{M}_{n,2} &:= -\frac{z}{m_n(z, w) + w} \mathbf{M}_{n,1}, \\ \mathbf{M}_{n,3} &:= -\frac{\bar{z}}{m_n(z, w) + w} \mathbf{M}_{n,1}. \end{aligned}$$

Using (7.25), we obtain

$$\inf_{\operatorname{Im}(w) \geq v_n} |(m_n(z, w) + w)^2 - |z|^2| \geq \inf_{\operatorname{Im}(w) \geq v_n} |m_n(z, w) + w - |z|| |m_n(z, w) + w + |z|| \geq v_n^2 \quad (7.26)$$

and hence

$$\sup_{|z| \leq M} \sup_{|w| \leq \beta, \operatorname{Im}(w) \geq v_n} \|\mathbf{M}_n^{-1}\| = O(v_n^{-4}).$$

Since  $\sup_{\operatorname{Im}(w) \geq v_n} \|\mathbf{G}_n(z, w)\| \leq v_n^{-1}$ , we obtain

$$\sup_{1 \leq k \leq n} \sup_{|w| \leq \beta, \operatorname{Im}(w) \geq v_n} \|\mathbf{Q}_k\| = O(v_n^{-1}).$$

Therefore, by (7.24), we have

$$\begin{aligned} \sup_{1 \leq k \leq n} \|\mathbf{Q}_k - \mathbf{M}_n^{-1}\| &\leq \sup_{1 \leq k \leq n} \|\mathbf{Q}_k(\mathbf{Q}_k^{-1} - \mathbf{M}_n)\mathbf{M}_n^{-1}\| \\ &\leq \sup_{1 \leq k \leq n} \|\mathbf{Q}_k\| \|\mathbf{Q}_k^{-1} - \mathbf{M}_n\| \|\mathbf{M}_n^{-1}\| \\ &\ll \frac{1}{n^{1/8} v_n^6} + \frac{1}{n^{\delta\eta/2} v_n^6} \end{aligned}$$

with probability at least  $1 - O(n^{-99})$ . Since  $m_n(z, w)$  is the normalized sum of the diagonal elements of  $\mathbf{G}_n$ , we now consider the diagonal elements of  $\mathbf{Q}_k$  and  $\mathbf{M}_n$ ; from the above estimate, we conclude that

$$\left| m_n(z, w) + \frac{m_n(z, w) + w}{(m_n(z, w) + w)^2 - |z|^2} \right| \ll v_n^5 \quad (7.27)$$

with probability  $1 - O(n^{-99})$ . Here we used the fact that

$$\frac{1}{n^{1/8} v_n^6} + \frac{1}{n^{\delta\eta/2} v_n^6} \leq 2v_n^5$$

by definition of  $v_n$ .

We now use an  $\varepsilon$ -net argument to extend (7.27) to all  $|z| \leq M$  and  $|w| \leq \beta$  with  $\operatorname{Im}(w) \geq v_n$ . Since  $\sup_{\operatorname{Im}(w) \geq v_n} \|\mathbf{G}_n(z, w)\| \leq v_n^{-1}$ , we apply (7.18) and the resolvent identity (7.19) to obtain the deterministic bounds

$$|m_n(z, w) - m_n(z, w')| \leq \frac{|w - w'|}{v_n^2}$$

and

$$|m_n(z, w) - m_n(z', w)| \leq \frac{|z - z'|}{v_n^2}$$

for all  $z, z' \in \mathbb{C}$  and  $w, w' \in \mathbb{C}$  with  $\text{Im}(w), \text{Im}(w') \geq v_n$ . Applying (7.26) and the triangle inequality, we obtain

$$\begin{aligned} & \left| \frac{m_n(z, w) + w}{(m_n(z, w) + w)^2 - |z|^2} - \frac{m_n(z, w') + w'}{(m_n(z, w') + w')^2 - |z|^2} \right| \\ & \ll \frac{1}{v_n^2} |m_n(z, w) + w - m_n(z, w') - w'| \\ & \quad + \frac{1}{v_n} \left| \frac{1}{(m_n(z, w) + w)^2 - |z|^2} - \frac{1}{(m_n(z, w') + w')^2 - |z|^2} \right| \\ & \ll \frac{|w - w'|}{v_n^4} + \frac{1}{v_n^5} |(m_n(z, w) + w)^2 - (m_n(z, w') + w')^2| \\ & \ll \frac{|w - w'|}{v_n^8} \end{aligned}$$

for all  $|z| \leq M$  and  $|w|, |w'| \leq \beta$  with  $\text{Im}(w), \text{Im}(w') \geq v_n$ . Similarly,

$$\left| \frac{m_n(z, w) + w}{(m_n(z, w) + w)^2 - |z|^2} - \frac{m_n(z', w) + w}{(m_n(z', w) + w)^2 - |z'|^2} \right| \ll \frac{|z - z'|}{v_n^8}$$

for all  $|z|, |z'| \leq M$  and  $|w| \leq \beta$  with  $\text{Im}(w) \geq v_n$ .

We now apply an  $\varepsilon$ -net argument with  $\varepsilon = v_n^{13}$  to the sets  $\{z \in \mathbb{C} : |z| \leq M\}$  and  $\{w \in \mathbb{C} : |w| \leq \beta, \text{Im}(w) \geq v_n\}$ . Let  $\mathcal{N}_1$  and  $\mathcal{N}_2$  denote the respective  $\varepsilon$ -nets of the two sets. By Lemma 7.10,

$$|\mathcal{N}_1| + |\mathcal{N}_2| \ll v_n^{-26} \leq n^{1/2}.$$

Therefore, by a standard  $\varepsilon$ -net argument and the union bound, we conclude that

$$\sup_{|z| \leq M} \sup_{|w| \leq \beta, \text{Im}(w) \geq v_n} \left| m_n(z, w) + \frac{m_n(z, w) + w}{(m_n(z, w) + w)^2 - |z|^2} \right| = O(v_n^5)$$

with probability (say)  $1 - O(n^{-2})$ . The claim now follows from an application of the Borel-Cantelli lemma.  $\square$

**7.11. Proof of Theorem 2.7.** This subsection is devoted to the proof of Theorem 2.7. With Lemma 7.5 and Theorem 7.8 in hand, the proof of Theorem 2.7 will follow from a standard (and somewhat technical) argument; see [3, Chapter 11], [36], and references therein. We detail the argument below.

Recall the definition of the functions  $g_{\mathbf{M}}$  and  $g$  given in (7.1) and (7.2). By [3, Chapter 11] (see also [6] and [17, Section 3]), for each  $z \in \mathbb{C}$ , there exists a probability measure  $\nu_z$  on the real line such that

$$g(s, t) = \frac{\partial}{\partial s} \int_0^\infty \log |x|^2 d\nu_z(dx),$$

where  $z = s + \sqrt{-1}t$ .

Assume  $\{\mathbf{X}_n\}_{n \geq 1}$  and  $\{\mathbf{N}_n\}_{n \geq 1}$  satisfy the assumptions of Theorem 2.7. By Lemma 7.1 and [3, Lemma 11.5], in order to prove Theorem 2.7, it suffices to show that a.s.

$$\iint \left[ g_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n)}(s, t) - g(s, t) \right] e^{\sqrt{-1}us + \sqrt{-1}vt} dt ds \rightarrow 0$$

as  $n \rightarrow \infty$ .

By the triangle inequality and Lemma 7.4, we have that a.s.

$$\frac{1}{n} \left\| \frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) \right\|_2^2 = O_d(1). \quad (7.28)$$

Let  $A > 0$ . Define

$$T := \{(s, t) : |s| \leq A, |t| \leq A^3\}.$$

By [3, Lemma 11.7] and (7.28), in order to prove Theorem 2.7 it suffices to show that for each fixed  $A > 0$  a.s.

$$\iint_T \left[ g_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n)}(s, t) - g(s, t) \right] e^{\sqrt{-1}us + \sqrt{-1}vt} dt ds \longrightarrow 0$$

as  $n \rightarrow \infty$ .

Let  $\varepsilon_n := n^{-B}$  for some sufficiently large  $B > 0$  (independent of  $n$ ) to be chosen later. Following the integration by parts argument from [3, Section 11.7], it suffices to show that a.s.

$$\limsup_{n \rightarrow \infty} \iint_T \left| \int_{\varepsilon_n}^{\infty} \log |x|^2 \left( \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) - \nu_z(dx) \right) \right| dt ds = 0 \quad (7.29)$$

and

$$\limsup_{n \rightarrow \infty} \iint_T \left| \int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) \right| dt ds = 0, \quad (7.30)$$

and similarly with the two-dimensional integral on  $T$  replaced by one-dimensional integrals on the boundary of  $T$ . We shall only estimate the two-dimensional integrals, as the treatment of the one-dimensional integrals are similar.

We prove (7.29) first. By (7.28), it follows that  $\nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}$  is supported on  $[-n^{50}, n^{50}]$  a.s. Thus, it suffices to show that a.s.

$$\limsup_{n \rightarrow \infty} \iint_T \left| \int_{\varepsilon_n}^{n^{50}} \log |x|^2 \left( \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) - \nu_z(dx) \right) \right| dt ds = 0.$$

By definition of  $\varepsilon_n$ , it suffices to show that a.s.

$$\limsup_{n \rightarrow \infty} (\log n) \sup_{z \in T} \sup_{x \in \mathbb{R}} \left| \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}((-\infty, x)) - \nu_z((-\infty, x)) \right| = 0. \quad (7.31)$$

(7.31) will follow from Lemma 7.12 below.

We now prove (7.30). By Theorem 2.9 (and the Borel-Cantelli lemma), for some sufficiently large  $B > 0$ , we have the following:

$$\text{for a.e. } z \in T, \text{ a.s. } \lim_{n \rightarrow \infty} \int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) = 0. \quad (7.32)$$

We now observe that it is possible to switch the quantifiers ‘‘a.e.’’ on  $z$  and ‘‘a.s.’’ on  $\omega$  in (7.32) using the arguments from [6, Section 4] and Fubini’s theorem, where  $\omega$  denotes an element of the sample space. Thus, we have

$$\text{a.s., for a.e. } z \in T, \lim_{n \rightarrow \infty} \int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) = 0. \quad (7.33)$$

Using the  $L^2$ -norm argument in [36, Section 12], it follows that a.s.

$$\left( \iint_T \left| \int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) \right|^2 dt ds \right)^{1/2} \quad (7.34)$$

is bounded uniformly in  $n$ , and hence the sequence of functions  $\int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx)$  is a.s. uniformly integrable on  $T$ . Let  $L > 1$  be a large parameter and define  $T_{L,n}$  to be the set of all

$z \in T$  such that  $\left| \int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) \right| \leq L$ . By (7.33) and the dominated convergence theorem, we have a.s.

$$\lim_{n \rightarrow \infty} \iint_{T_{L,n}} \left| \int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) \right| dt ds = 0.$$

On the other hand, from the uniform boundedness of (7.34), we obtain a.s.

$$\limsup_{n \rightarrow \infty} \iint_{T \setminus T_{L,n}} \left| \int_0^{\varepsilon_n} \log |x|^2 \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}}(dx) \right| dt ds \ll \frac{1}{L}.$$

Combining the bounds above and taking  $L \rightarrow \infty$  yields (7.30).

It remains to establish the following lemma.

**Lemma 7.12.** *Let  $\{\mathbf{X}_n\}_{n \geq 1}$  be a sequence of random matrices that satisfies condition **C0** with parameter  $d \geq 2$  and atom variables  $(\xi_{st})_{s,t=1}^d$ , and assume (7.3) holds for some  $\eta > 0$ . For each  $n \geq 1$ , let  $\mathbf{N}_n$  is a  $dn \times dn$  matrix such that  $\text{rank}(\mathbf{N}_n) = O(n^{1-\varepsilon})$  for some  $\varepsilon > 0$ . Then there exists  $\alpha > 0$  such that a.s.*

$$\sup_{|z| \leq M} \left\| \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}} - \nu_z \right\| = O_{M, m_2 + \eta, d}(n^{-\alpha}),$$

where  $\|\nu - \mu\| := \sup_{x \in \mathbb{R}} |\nu((-\infty, x)) - \mu((-\infty, x))|$  for any two probability measures  $\nu, \mu$  on the real line.

*Proof.* (of Lemma 7.12) The proof of Lemma 7.12 is based on the arguments from [39, Lemma 64]. By [3, Theorem A.43],

$$\sup_{z \in \mathbb{C}} \left\| \nu_{\frac{1}{\sqrt{n}}(\mathbf{X}_n + \mathbf{N}_n) - z\mathbf{I}} - \nu_{\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}} \right\| = O(n^{-\varepsilon}).$$

Thus, by the triangle inequality, it suffices to show that a.s.

$$\sup_{|z| \leq M} \left\| \nu_{\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}} - \nu_z \right\| = O_{M, m_2 + \eta}(n^{-\alpha})$$

for some  $\alpha > 0$ .

From [17, Remark 3.1], it follows that for each  $z \in \mathbb{C}$ ,  $\nu_z$  has density  $\rho_z$  with

$$\sup_{z \in \mathbb{C}} \sup_{x \in \mathbb{R}} |\rho_z(x)| \leq 1.$$

By [3, Lemma B.18], it suffices to show that a.s.

$$\sup_{|z| \leq M} L\left(F^{\mathbf{H}_n(z)}, F_z\right) = O_{M, m_2 + \eta}(n^{-\alpha}),$$

where  $F_z$  is the cumulative distribution function of  $\nu_z$ . We remind the reader that  $L(F, G)$  denotes the Levy distance, defined in (7.11), between the distribution functions  $F$  and  $G$ .

By Lemma 7.5, it suffices to show that a.s.

$$\sup_{|z| \leq M} \left\| \nu_{\frac{1}{\sqrt{n}}\hat{\mathbf{X}}_n - z\mathbf{I}} - \nu_z \right\| = O_{M, m_2 + \eta}(n^{-\alpha}), \quad (7.35)$$

where  $\{\hat{\mathbf{X}}_n\}_{n \geq 1}$  is the sequence of truncated matrices from Lemma 7.5 for some  $0 < \delta < 1/100$ . Let  $m(z, w)$  denote the Stieltjes transform of  $\nu_z$ . That is,

$$m(z, w) := \int \frac{1}{x - w} d\nu_z(dx) = \int \frac{\rho_z(x) dx}{x - w}.$$

From [17, Section 3], it follows that  $m(z, w)$  is a solution of

$$m(z, w) + \frac{m(z, w) + w}{(m(z, w) + w)^2 - |z|^2} = 0$$

analytic in the upper-half plane  $\{w \in \mathbb{C} : \text{Im}(w) > 0\}$ .

By [17, Remark 3.1], we choose  $\beta > 100$  sufficiently large (depending only on  $M$ ) such that  $\rho_z$  is supported inside the interval  $[-\beta/2, \beta/2]$  for all  $|z| \leq M$ . By Theorem 7.8 and [17, Lemma 2.4], it follows that a.s.

$$\sup_{|z| \leq M} \sup_{|w| \leq 4\beta, \text{Im}(w) \geq v_n} |\hat{m}_n(z, w) - m(z, w)| = O_{M, m_{2+\eta}}(v_n^4), \quad (7.36)$$

where  $v_n$  is defined in Theorem 7.8.

For the remainder of the proof, we fix a realization in which (7.36) holds. The implicit constants in our asymptotic notation (such as  $O, o, \Omega, \ll$ ) depend only on the constants  $M, m_{2+\eta}, d$ ; for simplicity, we no longer include these subscripts in our notation. By [17, (3.2)] and (7.36), it follows that

$$\sup_{|z| \leq M} \sup_{|w| \leq 4\beta, \text{Im}(w) \geq v_n} \text{Im}(\hat{m}_n(z, w)) \ll 1.$$

Write  $w = u + \sqrt{-1}v$ . For any interval  $I \subset \mathbb{R}$ , we define

$$N_I(z) := \#\left\{1 \leq i \leq 2dn : \lambda_i(\hat{\mathbf{H}}_n(z)) \in I\right\},$$

where  $\lambda_1(\hat{\mathbf{H}}_n(z)), \lambda_2(\hat{\mathbf{H}}_n(z)), \dots, \lambda_{2dn}(\hat{\mathbf{H}}_n(z))$  are the eigenvalues of  $\hat{\mathbf{H}}_n(z)$ . We remind the reader that the eigenvalues of  $\hat{\mathbf{H}}_n(z)$  are given by

$$\pm\sigma_1 \left( \frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n - z\mathbf{I} \right), \pm\sigma_2 \left( \frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n - z\mathbf{I} \right), \dots, \pm\sigma_{dn} \left( \frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n - z\mathbf{I} \right).$$

For an interval  $I \subset [-\beta, \beta]$  of length  $|I| = v \geq v_n$  centered at  $u$ , we have

$$\sup_{|z| \leq M} \frac{N_I(z)}{8dnv} \leq \sup_{|z| \leq M} \frac{1}{2dn} \sum_{i=1}^{2dn} \frac{v}{v^2 + |u - \lambda_i(\hat{\mathbf{H}}_n(z))|^2} = \sup_{|z| \leq M} \text{Im}(\hat{m}_n(z, u + \sqrt{-1}v)) \ll 1.$$

We conclude that for any interval  $I \subset [-\beta, \beta]$  of length  $|I| \geq v_n$ ,

$$\sup_{|z| \leq M} N_I(z) \ll |I|n. \quad (7.37)$$

Fix an interval  $I \subset [-\beta, \beta]$  with  $|I| \geq 10v_n$ . Define

$$f(y) := \frac{1}{\pi} \int_I \frac{v_n}{v_n^2 + (x - y)^2} dx.$$

We have

$$\frac{1}{2dn} \sum_{i=1}^{2dn} f(\lambda_i(\hat{\mathbf{H}}_n(z))) = \frac{1}{\pi} \text{Im} \int_I m_n(z, u + \sqrt{-1}v_n) du$$

and

$$\int_{\mathbb{R}} f(y) \rho_z(y) dy = \frac{1}{\pi} \text{Im} \int_I m(z, u + \sqrt{-1}v_n) du$$

for any  $z \in \mathbb{C}$ . Therefore, by (7.36), we obtain

$$\sup_{|z| \leq M} \left| \frac{1}{2dn} \sum_{i=1}^{2dn} f(\lambda_i(\hat{\mathbf{H}}_n(z))) - \int_I f(y) \rho_z(y) dy \right| = O(|I|v_n^4).$$

We note the following pointwise bounds:

$$f(y) \ll \frac{v_n |I|}{\text{dist}^2(y, I)}$$

when  $y \notin I$  and  $\text{dist}(y, I) \geq |I|$ , and

$$f(y) \ll \frac{1}{1 + \text{dist}(y, I)/v_n}$$

when  $y \notin I$  and  $\text{dist}(y, I) < |I|$ . In the case where  $y \in I$ , we have

$$f(y) = 1 + O\left(\frac{1}{1 + \text{dist}(y, I^c)/v_n}\right)$$

as  $\frac{1}{\pi} \frac{v_n}{v_n^2 + (x-y)^2}$  has total integral 1. Using these bounds, we find

$$\sup_{|z| \leq M} \left| \int_{\mathbb{R}} f(y) \rho_z(y) dy - \int_I \rho_z(y) dy \right| = O\left(v_n \log \frac{|I|}{v_n}\right).$$

Similarly, by (7.37), Riemann integration, and the trivial bound  $N_J \leq 2dn$  for any interval  $J$  outside of  $[-\beta, \beta]$ , we have

$$\sup_{|z| \leq M} \left| \frac{1}{2dn} \sum_{i=1}^{2dn} f(\lambda_i(\hat{\mathbf{H}}_n(z))) - \frac{1}{2dn} N_I(z) \right| = O\left(v_n \log \frac{|I|}{v_n}\right).$$

Combining the bounds above, we conclude that for any interval  $I \subset [-\beta, \beta]$  with  $|I| \geq 10v_n$ , we have

$$\sup_{|z| \leq M} \left| \frac{1}{2dn} N_I(z) - \int_I \rho_z(y) dy \right| = O(v_n^4 |I|) + O\left(v_n \log \frac{|I|}{v_n}\right).$$

In particular, since  $\rho_z$  is supported inside  $[-\beta/2, \beta/2]$ , we obtain

$$\sup_{|z| \leq M} \frac{1}{2dn} N_{[-\beta/2, \beta/2]^c}(z) = O(v_n \log v_n^{-1}),$$

where  $[-\beta/2, \beta/2]^c$  is the complement of the interval  $[-\beta/2, \beta/2]$ . Thus, we have

$$\begin{aligned} \sup_{|z| \leq M} \left\| \nu \frac{1}{\sqrt{n}} \hat{\mathbf{X}}_n - z \mathbf{I} - \nu_z \right\| &\ll v_n \log v_n^{-1} + \sup_{|z| \leq M} \sup_{x \in [-\beta/2, \beta/2]} \left| \frac{1}{2dn} N_{[-\beta, x]}(z) - \int_{-\beta}^x \rho_z(y) dy \right| \\ &\ll v_n \log v_n^{-1}. \end{aligned}$$

Since this bound holds for each fixed realization in which (7.36) holds, we obtain (7.35) a.s. The proof of the lemma is complete.  $\square$

## APPENDIX A. PROOF OF THEOREM 5.2 AND THEOREM 6.1

We will mainly focus on Theorem 6.1 as the proof for Theorem 5.2 is similar. Assume that  $z_{11}, \dots, z_{dd}$  are random variables satisfying (4.1). Then as  $\mathbb{E}|z_i|^{2+\eta}$  is bounded and as the covariance matrix  $(\mathbb{E}z_{ij}\bar{z}_{i'j'})_{ij, i'j'}$  is the identity matrix (and thus is non-degenerate), we have the following fact (whose proof is left as an exercise).

**Claim A.1.** *There exists a sufficiently small constant  $\delta > 0$  such that the following holds.*

- (1) *There exist measurable sets  $R_1, \dots, R_{d^2} \subset B(0, \delta^{-1}) \subset \mathbb{C}^{d^2}$  such that  $\mathbb{P}((z_{11}, \dots, z_{dd}) \in R_i) \geq \delta$  for all  $1 \leq i \leq d^2$  and for any collection of  $d^2$  vectors  $\mathbf{v}_1 \in R_1, \dots, \mathbf{v}_{d^2} \in R_{d^2}$ , the least singular value of the matrix formed from  $\mathbf{v}_1, \dots, \mathbf{v}_{d^2}$  is at least  $\delta$ .*
- (2) *There exist measurable sets  $R_1, \dots, R_d \subset B(0, \delta^{-1}) \subset \mathbb{C}^d$  such that  $\mathbb{P}((z_{11}, \dots, z_{1d}) \in R_1, \dots, (z_{d1}, \dots, z_{dd}) \in R_d) \geq \delta$ , and for any collection of  $d$  vectors  $\mathbf{v}_1 \in R_1, \dots, \mathbf{v}_d \in R_d$ , the least singular value of the matrix formed from  $\mathbf{v}_1, \dots, \mathbf{v}_d$  is at least  $\delta$ .*



We remark that (1) and (2) are useful for proving Theorem 5.2 and Theorem 6.1 respectively. Now we sketch the main steps to prove Theorem 6.1 following [26].

First, we have

$$\begin{aligned} \mathbb{P}\left(\sum_{1 \leq i \leq n} X^{(i)} \mathbf{u}^{(i)} \in B(\mathbf{u}, \beta)\right) &= \mathbb{P}\left(\exp(-\pi \|\sum_{i=1}^n X^{(i)} \mathbf{u}^{(i)} - \mathbf{u}\|^2) \geq \exp(-\pi \beta^2)\right) \\ &\leq \exp(\pi \beta^2) \mathbb{E} \exp(-\pi \|\sum_{i=1}^n X^{(i)} \mathbf{u}^{(i)} - \mathbf{u}\|^2) \\ &\leq \exp(\pi \beta^2) \int_{\mathbb{C}^d} \mathbb{E} e(\langle \sum_{1 \leq i \leq n} X^{(i)} \mathbf{u}^{(i)}, \mathbf{t} \rangle) e(-\langle \mathbf{u}, \mathbf{t} \rangle) \exp(-\pi \|\mathbf{t}\|^2) d\mathbf{t}. \end{aligned}$$

By using the independence of  $X^{(i)}$  and other elementary estimates such as  $|x| \leq |x|^2/2 + 1/2$  and  $|\cos(\pi x)| \leq \exp(-2\|x\|_{\mathbb{R}/\mathbb{Z}}^2)$ , we obtain

$$|\mathbb{E} e(\langle X^{(i)} \mathbf{u}^{(i)}, \mathbf{t} \rangle)| \leq \exp\left(-\mathbb{E}_Z \|\operatorname{Re}([z_{11}t_1 + \cdots + z_{d1}t_d] \mathbf{u}_1^{(i)} + \cdots + [z_{1d}t_1 + \cdots + z_{dd}t_d] \mathbf{u}_d^{(i)})\|_{\mathbb{R}/\mathbb{Z}}^2\right),$$

where  $X'$  is an iid copy of  $X = (x_{ij})_{1 \leq i, j \leq d}$  and  $Z = X - X'$ . By scaling the  $\mathbf{u}^{(i)}$  by a factor of  $\beta^{-1}$ , it is enough to assume  $\beta = 1$ . Set  $M := 2A \log n$  where  $A$  is large enough. From the fact that  $\gamma \geq n^{-O(1)}$  we easily obtain

$$\begin{aligned} \frac{\gamma}{2} &\leq \int_{\|\mathbf{t}\| \leq M} \exp\left(-\sum_{i=1}^n \mathbb{E}_Z \|\operatorname{Re}([z_{11}t_1 + \cdots + z_{d1}t_d] \mathbf{u}_1^{(i)} + \cdots + [z_{1d}t_1 + \cdots + z_{dd}t_d] \mathbf{u}_d^{(i)})\|_{\mathbb{R}/\mathbb{Z}}^2 - \pi \|\mathbf{t}\|^2\right) d\mathbf{t}. \end{aligned} \quad (\text{A.1})$$

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

$$S_m := \left\{ \mathbf{t} \in \mathbb{C}^d : \sum_{i=1}^n \mathbb{E}_Z \|\operatorname{Re}([z_{11}t_1 + \cdots + z_{d1}t_d] \mathbf{u}_1^{(i)} + \cdots + [z_{1d}t_1 + \cdots + z_{dd}t_d] \mathbf{u}_d^{(i)})\|_{\mathbb{R}/\mathbb{Z}}^2 + \|\mathbf{t}\|^2 \leq m \right\}.$$

Then it follows from (A.1) that there exists  $m \leq M$  such that  $\mu(S_m) \geq \gamma \exp(\frac{m}{4} - 2)$  and  $\mu(T) \geq c\gamma \exp(\frac{m}{4} - 2)m^{-2d}$ , where

$$T := \left\{ \mathbf{t} \in B(0, 1), \sum_{i=1}^n \mathbb{E}_Z \|\operatorname{Re}([z_{11}t_1 + \cdots + z_{d1}t_d] \mathbf{u}_1^{(i)} + \cdots + [z_{1d}t_1 + \cdots + z_{dd}t_d] \mathbf{u}_d^{(i)})\|_{\mathbb{R}/\mathbb{Z}}^2 \leq 4m \right\}.$$

By a proper discretization, with  $N$  a sufficiently large prime, we obtain the following discrete analog: there exists a subset  $S$  of size at least  $cN^{2d}\gamma \exp(\frac{m}{4} - 2)m^{-2d}$  of  $B_1 = \{k_1/N + \sqrt{-1}k_2/N : k_1, k_2 \in \mathbb{Z}, -2N \leq k_1, k_2 \leq 2N\}$  such that the following holds for any  $\mathbf{s} \in S$

$$\sum_{i=1}^n \mathbb{E}_Z \|\operatorname{Re}([z_{11}s_1 + \cdots + z_{d1}s_d] \mathbf{u}_1^{(i)} + \cdots + [z_{1d}s_1 + \cdots + z_{dd}s_d] \mathbf{u}_d^{(i)})\|_{\mathbb{R}/\mathbb{Z}}^2 \leq 16m.$$

Next, by the definition of  $S$ ,

$$\mathbb{E}_{\mathbb{Z}} \sum_{\mathbf{s} \in S} \sum_{i=1}^n \|\operatorname{Re}([z_{11}s_1 + \cdots + z_{d1}s_d]\mathbf{u}_1^{(i)} + \cdots + [z_{1d}s_1 + \cdots + z_{dd}s_d]\mathbf{u}_d^{(i)})\|_{\mathbb{R}/\mathbb{Z}}^2 \leq 16m|S|. \quad (\text{A.2})$$

It then follows that, by (2) of Claim A.1, there exists a matrix  $\mathbf{C} = (c_{11}, \dots, c_{dd})$  such that  $\delta \leq \sigma_d(\mathbf{C}) \leq \sigma_1 \leq \delta^{-1}$  and the following holds for some sufficiently large constant  $C$

$$\sum_{\mathbf{s} \in S} \sum_{i=1}^n \|\operatorname{Re}([c_{11}s_1 + \cdots + c_{d1}s_d]\mathbf{u}_1^{(i)} + \cdots + [c_{1d}s_1 + \cdots + c_{dd}s_d]\mathbf{u}_d^{(i)})\|_{\mathbb{R}/\mathbb{Z}}^2 \leq Cm|S|$$

$$\sum_{\mathbf{s} \in S} \sum_{i=1}^n \|\langle \mathbf{v}^{(i)}, \mathbf{s} \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \leq Cm|S|,$$

where  $\mathbf{v}^{(i)} := (c_{11}\mathbf{u}_1^{(i)} + \cdots + c_{1d}\mathbf{u}_d^{(i)}, \dots, c_{d1}\mathbf{u}_1^{(i)} + \cdots + c_{dd}\mathbf{u}_d^{(i)})$ .

Let  $n'$  be any number between  $n^\varepsilon$  and  $n$ . We say that an index  $1 \leq i \leq n$  is *bad* if  $\sum_{\mathbf{s} \in S} \|\langle \mathbf{s}, \mathbf{v}^{(i)} \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \geq \frac{Cm|S|}{n'}$ . Clearly the number of bad indices is at most  $n'$ . Let  $I$  be the set of good indices, and  $V$  be the set of vectors  $\mathbf{v}^{(i)}, i \in I$ . Recall that for an arbitrary vector  $\mathbf{v} \in V$

$$\sum_{\mathbf{s} \in S} \|\langle \mathbf{s}, \mathbf{v} \rangle\|_{\mathbb{R}/\mathbb{Z}}^2 \leq Cdm|S|/n'.$$

Set  $k := c\sqrt{\frac{n'}{m}}$  for some sufficiently small constant  $c$ , and let  $V_k := k(V \cup \{0\})$ . By the Cauchy-Schwarz inequality, for any  $\mathbf{v} \in V_k$ , we have

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

The last estimate implies that the size of  $V_k$  does not grow fast in terms of  $k$ . The treatment from here is identical to [26, Section 6] by using a Freiman-type inverse result [26, Theorem 3.2].

## REFERENCES

- [1] Z. D. Bai, J. Silverstein, Y. Q. Yin, *A note on the largest eigenvalue of large-dimensional sample covariance matrix*, J. Multivariate Anal. **26** (1988), 166–168.
- [2] Z. D. Bai, J. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab. Volume 26, Number 1 (1998), 316–345.
- [3] Z. D. Bai, J. Silverstein, *Spectral analysis of large dimensional random matrices*, Mathematics Monograph Series **2**, Science Press, Beijing 2006.
- [4] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), 494–529.
- [5] F. Benaych-Georges, F. Chapon, *Random right eigenvalues of Gaussian quaternionic matrices*, Random Matrices Theory Appl. **1** (2012).
- [6] C. Bordenave, D. Chafai, *Around the circular law*, Probability Surveys **9** (2012) 1–89.
- [7] D. L. Burkholder, *Distribution function inequalities for martingales*, Ann. Probab. **1** 19–42 (1973).
- [8] K. Costello, T. Tao, V. Vu, *Random symmetric matrices are almost surely non-singular*, Duke Math. J. **135** (2006), 395–413.
- [9] S. J. Dilworth, *Some probabilistic inequalities with applications to functional analysis*, Banach Spaces. Contemp. Math. **144** 53–67 (1993).
- [10] 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).
- [11] P. Erdős, *Extremal problems in number theory*, Proc. Sympos. pure Math., Vol. VIII, pp. 181–189 AMS (1965).
- [12] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902.
- [13] J. Ginibre, *Statistical ensembles of complex, quaternion and real matrices*, J. Math. Phys. **6** (1965), 440–449.
- [14] V. L. Girko, *Circular law*, Theory Probab. Appl. (1984), 694–706.
- [15] V. L. Girko, *The strong circular law, twenty years later*, II. Random Oper. Stochastic Equations **12** (2004), no. 3, 255–312.
- [16] I. Goldscheid, B. A. Khoruzhenko, *The Thouless formula for random non-Hermitian Jacobi matrices*, Israel J. Math., **148** (2005), 331–346.

- [17] F. Götze, T. Tikhomirov, *The circular law for random matrices*, Ann. Probab. Volume 38, Number 4, 1444–1491 (2010).
- [18] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. **8** (1977), no. 3–4, 197–211.
- [19] R. A. Horn, C. R. Johnson, *Topics in Matrix Analysis*, Cambridge Univ. Press (1991).
- [20] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press (1991).
- [21] J. E. Littlewood, 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.
- [22] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Acad. Press. (1967).
- [23] M. L. Mehta, *Random Matrices*, third edition. Elsevier/Academic Press, Amsterdam, (2004).
- [24] H. Nguyen, S. O’Rourke. *The elliptic law*, available at [arXiv:1208.5883](https://arxiv.org/abs/1208.5883) [math.PR].
- [25] H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Mathematics Journal Vol. 161, 4 (2012), 545–586.
- [26] H. Nguyen, V. Vu, *Optimal Littlewood-Offord theorems*, Advances in Math., Vol. 226 6 (2011), 5298–5319.
- [27] H. Nguyen, V. Vu, *Random matrices: law of the determinant*, Annals of Probability (2014), Vol. 42, No. 1, 146–167.
- [28] H. Nguyen, V. Vu, *Small probability, inverse theorems, and applications*, Bolyai Society Mathematical Studies, Vol. 25 (2013).
- [29] S. O’Rourke, D. Renfrew, *Low rank perturbations of large elliptic random matrices*, available at [arXiv:1309.5326](https://arxiv.org/abs/1309.5326) [math.PR].
- [30] G. Pan, W. Zhou, *Circular law, extreme singular values and potential theory*, Journal of Multivariate Analysis, **101** (2010), 645–656.
- [31] M. Rudelson, R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics **218** (2008), 600–633.
- [32] A. Soshnikov, *A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices*, J. Stat. Phys. **108** (2002), 1033–1056.
- [33] T. Tao, 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.
- [34] T. Tao, V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2) **169** (2009), no 2, 595–632.
- [35] T. Tao, V. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), 603–628.
- [36] T. Tao, V. Vu, *Random matrices: The Circular Law*, Communication in Contemporary Mathematics **10** (2008), 261–307.
- [37] T. Tao, V. Vu, *Smooth analysis of the condition number and the least singular value*, Mathematics of Computation **79** (2010), 2333–2352.
- [38] T. Tao, V. Vu, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab. Volume 38, Number 5 (2010), 2023–2065.
- [39] T. Tao, V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math **206** (2011), 127–204.
- [40] R. Vershynin, *Invertibility of symmetric random matrices*, Random Structures and Algorithms 44 (2014), 135–182.
- [41] E. P. Wigner, *On the distributions of the roots of certain symmetric matrices*, Ann. Math. **67**, 325–327 (1958).
- [42] F. Zhang, *Quaternions and Matrices of Quaternions*, Linear Algebra and its Applications **251** (1997), 21–57.

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, COLUMBUS, OH 43210, USA

*E-mail address:* [nguyen.1261@math.osu.edu](mailto:nguyen.1261@math.osu.edu)

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN , CT 06520, USA

*E-mail address:* [sean.orourke@yale.edu](mailto:sean.orourke@yale.edu)