Journal of Combinatorics
Volume 4, Number 1, 1–30, 2013

Circular law for random discrete matrices of given row sum

Hoi H. Nguyen* and Van H. Vu†

Let M_n be a random matrix of size $n \times n$ and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of M_n. The *empirical spectral distribution* μ_{M_n} of M_n is defined as

$$\mu_{M_n}(s,t) = \frac{1}{n}\#\{k \le n, \Re(\lambda_k) \le s; \Im(\lambda_k) \le t\}.$$

The circular law theorem in random matrix theory asserts that if the entries of M_n are i.i.d. copies of a random variable with mean zero and variance σ^2, then the empirical spectral distribution of the normalized matrix $\frac{1}{\sigma\sqrt{n}}M_n$ of M_n converges almost surely to the uniform distribution $\mu_{\mathbf{cir}}$ over the unit disk as n tends to infinity.

In this paper, we show that the empirical spectral distribution of the normalized matrix of M_n, a random matrix whose rows are independent random $(-1,1)$ vectors of given row-sum s with some fixed integer s satisfying $|s| \le (1-o(1))n$, also obeys the circular law. The key ingredient is a new polynomial estimate on the least singular value of M_n.

1. Introduction

Let M_n be a matrix of size $n \times n$ and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of M_n. Then the empirical spectral distribution (ESD) μ_{M_n} of M_n is defined as

$$\mu_{M_n}(s,t) = \frac{1}{n}\#\{k \le n, \Re(\lambda_k) \le s; \Im(\lambda_k) \le t\}.$$

We also define $\mu_{\mathbf{cir}}$ as the uniform distribution over the unit disk,

$$\mu_{\mathbf{cir}}(s,t) = \frac{1}{\pi} mes(|z| \le 1; \Re(z) \le s, \Im(z) \le t).$$

arXiv: 1203.5941
*The first author is supported by research grant DMS-1256802.
†The second author is supported by research grants DMS-0901216 and AFOSAR-FA-9550-12-1-0083.

Confirming a long standing conjecture in random matrix theory, a recent result of Tao and Vu (appendix by Krishnapur) proves a universal law for the ESD of random i.i.d. matrices.

Theorem 1.1. *[32] Assume that the entries of M_n are i.i.d. copies of a complex random variable of mean zero and finite non-zero variance σ^2, then the ESD of the matrix $\frac{1}{\sigma\sqrt{n}}M_n$ converges to $\mu_{\mathbf{cir}}$ almost surely as n tends to ∞.*

The proof of this result is built upon previous important developments of Girko [9, 10], Bai [1], Götze-Tikhomirov [11], Pan-Zhou [21], Tao-Vu [28] and many others.

In view of universality phenomenon, it is of importance to study the law for random matrices of non-independent entries. Probably one of the first results in this direction is due to Bordenave, Caputo and Chafai [3] who prove the law for random Markov matrices.

Theorem 1.2. *[3, Theorem 1.3] Let X be a random matrix of size $n \times n$ whose entries are i.i.d. copies of a non-negative continuous random variable with finite variance σ^2 and bounded density function. Then with probability one the ESD of the normalized matrix $\sqrt{n}\bar{X}$, where $\bar{X} = (\bar{x}_{ij})_{1\le i,j\le n}$ and $\bar{x}_{ij} := x_{ij}/(x_{i1} + \cdots + x_{in})$, converges weakly to the circular measure $\mu_{\mathbf{cir}}$.*

In particular, when x_{11} follows the exponential law of mean one, Theorem 1.2 establishes the circular law for the Dirichlet Markov ensemble (see also [4]). We remark that the assumptions of continuity and boundedness are crucial in the proof of Theorem 1.2.

Related results with "linear" assumption of independence include a result of Tao, who among other things proves the circular law for random zero-sum matrices.

Theorem 1.3. *[25, Theorem 1.13] Let X be a random matrix of size $n \times n$ whose entries are i.i.d. copies of a random variable of mean zero and variance one. Then the ESD of the normalized matrix $\frac{1}{\sqrt{n}}\bar{X}$, where $\bar{X} = (\bar{x}_{ij})_{1\le i,j\le n}$ and $\bar{x}_{ij} := x_{ij} - \frac{1}{n}(x_{i1} + \cdots + x_{in})$, converges almost surely to the circular measure $\mu_{\mathbf{cir}}$.*

The main goal of this note is to show that the circular law also holds for random discrete matrices of similar constraints.

Theorem 1.4 (Main result)**.** *Let $0 < \epsilon \le 1$ be a positive constant. Let M_n be a random $(-1, 1)$ matrix of size $n \times n$ whose rows are independent vectors of given row-sum s with some s satisfying $|s| \le (1 - \epsilon)n$. Then the ESD of*

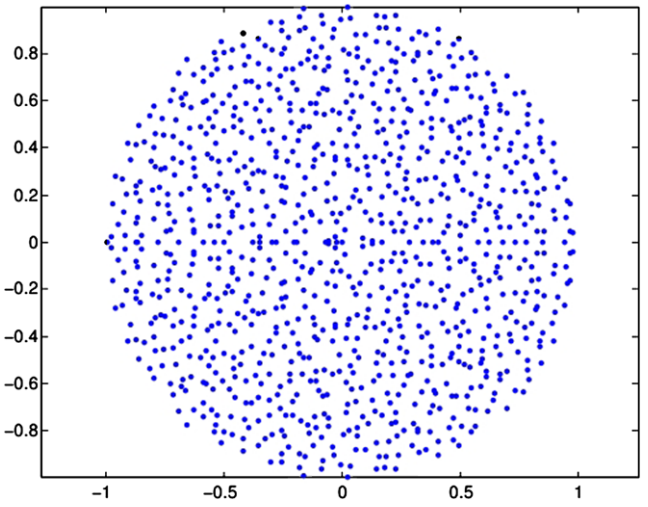

Figure 1: The ESD of a random matrix of size 1000 by 1000 whose rows are $(-1, 1)$ vectors of zero-sum, picture by Phillip Woods.

the normalized matrix $\frac{1}{\sigma\sqrt{n}}M_n$*, where* $\sigma^2 = 1-(\frac{s}{n})^2$*, converges almost surely to the distribution* $\mu_{\mathbf{cir}}$ *as* n *tends to* ∞*.*

To some extent, our matrix is a discrete version of the random Markov matrices considered in Theorem 1.2 where the entries are restricted to $\pm 1/s$. However, it is probably more suitable to compare our model with that of random Bernoulli matrices. By Theorem 1.1, the ESD of the normalized random Bernoulli matrices obeys the circular law, and hence our Theorem 1.4 serves as a local version of the law.

We remark that in a very recent result [19], the first author is able to prove a similar law for random doubly stochastic matrices, thus confirming the universality principle for another type of matrix of independent entries. Although the results are similar in spirit, the difficulties in each note are very different. The main obstacle of this note is to study the singularity of M_n and its perturbed variants. Inverse techniques developed in the literature to deal with this problem do not seem to suffice. This leads us to a new development to be discussed in Section 3. In what follows we present some reduction steps to simplify our problem.

Observe that, by letting X_{n-1} be the submatrix generated by the first $n-1$ rows and columns of M_n, the spectra of M_n is the union of s and the spectra of the pertubed matrix $X_{n-1} - F_{n-1}$ where all of the rows of F_{n-1}

are identical copies of $(m_{n1}, \dots, m_{n(n-1)})$, here by m_{ij} we mean the ij-th entry of M_n.

Indeed, consider the matrix $M := M_n - \lambda I_n$. We have

$$\det(M) = \det(M'),$$

where M' is obtained from M by adding its first $n-1$ columns to its last one.

On the other hand, we also have

$$\det(M') = (s - \lambda) \det(M''),$$

where

$$M'' := \begin{pmatrix} m_{11} - \lambda & \cdots & m_{1(n-1)} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ m_{(n-1)1} & \cdots & m_{(n-1)(n-1)} - \lambda & 1 \\ m_{n1} & \cdots & m_{n(n-1)} & 1 \end{pmatrix}.$$

It is clear that $\det(M'') = \det(M''')$, where $M''' := (X_{n-1} - F_{n-1}) - \lambda I_{n-1}$. Thus, the spectra of M_n is indeed the union of s and the spectra of the pertubed matrix $X_{n-1} - F_{n-1}$.

The observation above suggests a way to prove Theorem 1.4 by looking at the ESD of $X_{n-1} - F_{n-1}$. This alternative helps us avoid the outlier eigenvalue s of M_n which may cause certain technical difficulty for any direct study on M_n.

Notice that the rows of X_{n-1} above are independent vectors chosen uniformly from the set of all $(-1, 1)$ vectors of row-sum either $s-1$ or $s+1$. So for Theorem 1.4 it suffices to show the following.

Theorem 1.5 (Circular law for pertubed matrices)**.** *Let X_n be a random $(-1, 1)$ matrix whose rows are independent random vectors of row-sum either $s-1$ or $s+1$ with given s satisfying $|s| \le (1-\epsilon)n$. Let F_n be a deterministic matrix whose rows are identical copies of a given $(-1, 1)$ vector $\mathbf{f}$. Then the ESD of $\frac{1}{\sigma\sqrt{n}}(X_n + F_n)$, where $\sigma^2 = 1 - (\frac{s}{n})^2$, converges almost surely to the distribution of $\mu_{\mathbf{cir}}$ as n tends to ∞.*

For short, by $\mathcal{S}$ we denote the set of all $(-1, 1)$ vectors $\mathbf{x} = (x_1, \dots, x_n)$ of row-sum either $s-1$ or $s+1$. To establish Theorem 1.5, we will relate X_n to a random matrix X'_n whose entries are i.i.d. copies of a random Bernoulli variable x of the following form

$$\begin{cases} \mathbf{P}(x = -1) = \frac{1}{2} - \frac{s}{2n}, \\ \mathbf{P}(x = 1) = \frac{1}{2} + \frac{s}{2n}. \end{cases} \tag{1}$$

It is known that the ESD of $\frac{1}{\sigma\sqrt{n}}(X'_n + F_n)$ converges uniformly to $\mu_{\mathbf{cir}}$ (see for instance [32, Corollary 1.15]). As we desire to pass this result to $X_n + F_n$, we will make use of a so-called replacement principle below.

Theorem 1.6. *[32, Theorem 2.1] Suppose for each n that $A_n = (a_{ij}), B_n = (b_{ij})$ are random matrices of size $n \times n$. Assume that*

- *the sum*
$$\frac{1}{n^2}\sum_{ij}(|a_{ij}|^2 + |b_{ij}|^2)$$
is bounded almost surely;
- *for almost all complex numbers z*
$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI_n\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI_n\right)\right|$$
converges almost surely to zero.

Then $\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n}$ converges almost surely to zero.

In application, $X_n + F_n$ plays the role of A_n and $X'_n + F_n$ plays that of B_n. It is clear that the first condition of Theorem 1.6 is satisfied. Thus, for Theorem 1.5 it suffices to justify the second condition.

Theorem 1.7. *For every fixed complex z we have*

$$\frac{1}{n}\log|\det((X_n + F_n) - z\sqrt{n}I_n)| - \frac{1}{n}\log|\det((X'_n + F_n) - z\sqrt{n}I_n)|$$

converges to zero almost surely.

We will outline a proof for Theorem 1.7 in the next section.

Notation. Here and later, asymptotic notations such as O, Ω, Θ, and so for, are used under the assumption that $n \to \infty$. A notation such as $O_C(.)$ emphasizes that the hidden constant in O depends on C.

For $1 \le s \le n$, we denote by $\mathbf{e}_s$ the unit vector $(0, \dots, 0, 1, 0, \dots, 0)$, where all but the s-th component are zero. For a real or complex vector $\mathbf{v} = (v_1, \dots, v_n)$, we use the shorthand $\|\mathbf{v}\|$ for its L_2-norm $(\sum_i |v_i|^2)^{1/2}$.

For a matrix M, we use the notation $\mathbf{r}_i(M)$ and $\mathbf{c}_j(M)$ to denote its i-th row and j-th column respectively. For an event A, we use the subscript $\mathbf{P}_{\mathbf{x}}(A)$ to emphasize that the probability under consideration is taking according to the random vector $\mathbf{x}$.

2. Proof of Theorem 1.7: Outline

Let $\mathbf{f}_1, \dots, \mathbf{f}_n$ denote the (deterministic) rows of $F_n + \sqrt{n} z I_n$, and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ as well as $\mathbf{x}'_1, \dots, \mathbf{x}'_n$ be the rows of X_n and X'_n respectively.

For each $i \geq 2$, let V_{i-1} be the space spanned by $\mathbf{x}_1 + \mathbf{f}_1, \dots, \mathbf{x}_{i-1} + \mathbf{f}_{i-1}$ and let $\operatorname{dist}(\mathbf{x}_i + \mathbf{f}_i, V_{i-1})$ be the distance from $\mathbf{x}_i + \mathbf{f}_i$ to V_{i-1}. Define similarly for V'_{i-1} and $\operatorname{dist}(\mathbf{x}'_i + \mathbf{f}_i, V'_{i-1})$. By the "base times height" formula we have

$$\begin{aligned}
&\log \Big| \det((X_n + F_n) - z\sqrt{n} I_n) \Big| = \sum_i \log \operatorname{dist}((\mathbf{x}_i + \mathbf{f}_i), V_{i-1}). \\
&= \sum_{i \leq m} \log \operatorname{dist}((\mathbf{x}_i + \mathbf{f}_i), V_{i-1}) + \sum_{m < i} \log \operatorname{dist}((\mathbf{x}_i + \mathbf{f}_i), V_{i-1}) \\
&:= \log S_1 + \log S_2;
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\log \Big| \det((X'_n + F_n) - z\sqrt{n} I_n) \Big| = \sum_i \log \operatorname{dist}((\mathbf{x}'_i + \mathbf{f}_i), V_{i-1}). \\
&= \sum_{i \leq m} \log \operatorname{dist}((\mathbf{x}'_i + \mathbf{f}_i), V'_{i-1}) + \sum_{m < i} \log \operatorname{dist}((\mathbf{x}'_i + \mathbf{f}_i), V'_{i-1}) \\
&:= \log S'_1 + \log S'_2.
\end{aligned}$$

where we set the threshold m to be $m := n - \log^8 n$.

In order to compare $\log |\det((X_n + F_n) - z\sqrt{n} I_n)|$ with $\log |\det((X'_n + F_n) - z\sqrt{n} I_n)|$, we will show the following.

Theorem 2.1. *With probability* $1 - \exp(-\log^{2-o(1)} n)$ *we have*

$$\frac{1}{n} |\log S_1 - \log S'_1| = O(\log^{-2} n).$$

Theorem 2.2. *With probability* $1 - O(n^{-100})$ *we have*

$$\frac{1}{n} (|\log S_2| + |\log S'_2|) = O(\log^9 n / n).$$

It is clear that Theorem 1.7 follows from Theorem 2.1 and Theorem 2.2. In what follows we outline the approach to prove these results.

2.3. Sketch of the proof of Theorem 2.1

One of the main ingredients is the following row replacement principle.

Lemma 2.4. *Let i be an integer between 1 and m. Let $\mathbf{x}_1, \dots, \mathbf{x}_i, \mathbf{x}'_i, \mathbf{x}'_{i+1}, \dots, \mathbf{x}'_m$ be $m+1$ independent vectors where the $\mathbf{x}_j$'s are random vectors of type $\mathcal{S}$ and $\mathbf{x}'_k$'s are random vectors whose components are i.i.d copies of x from (1). Assume that vol_i is the m-dimensional volume of the parallelepiped generated by $\mathbf{x}_1 + \mathbf{f}_1, \dots, \mathbf{x}_i + \mathbf{f}_i, \mathbf{x}'_{i+1} + \mathbf{f}_{i+1}, \dots, \mathbf{x}'_m + \mathbf{f}_m$ and vol_{i-1} is that of the parallelepiped generated by $\mathbf{x}_1 + \mathbf{f}_1, \dots, \mathbf{x}_{i-1} + \mathbf{f}_{i-1}, \mathbf{x}'_i + \mathbf{f}_i, \dots, \mathbf{x}'_m + \mathbf{f}_m$. Then we have*

$$\begin{aligned}\mathbf{P}_{\mathbf{x}_1,\dots,\mathbf{x}_i,\mathbf{x}'_i,\mathbf{x}'_{i+1},\dots,\mathbf{x}'_m}\Big(|\log \mathrm{vol}_i - \log \mathrm{vol}_{i-1}| = O(\log^{-2} n)\Big)\\ = 1 - \exp(-\log^{2-o(1)} n).\end{aligned}$$

Lemma 2.1 then follows by a repeatedly use of Lemma 2.4 and the triangle inequality using the fact that S_1 and S'_1 are volumes of the parallelepipeds generated by $\mathbf{x}_1 + \mathbf{f}_1, \dots, \mathbf{x}_m + \mathbf{f}_m$ and by $\mathbf{x}'_1 + \mathbf{f}_1, \dots, \mathbf{x}'_m + \mathbf{f}_m$ respectively.

We now justify Lemma 2.4. We express vol_i as $\mathrm{vol}_i = d \times \mathrm{vol}$, where d is the distance from $\mathbf{x}_i + \mathbf{f}_i$ to the space V spanned by $\mathbf{x}_1 + \mathbf{f}_1, \dots, \mathbf{x}_{i-1} + \mathbf{f}_{i-1}, \mathbf{x}'_{i+1} + \mathbf{f}'_{i+1}, \dots, \mathbf{x}'_m + \mathbf{f}_m$ and vol is the volume of the parallelepiped generated by these vectors. Similarly, we can express vol_{i-1} as $\mathrm{vol}_{i-1} = d' \times \mathrm{vol}$, where d' is the distance from $\mathbf{x}'_i + \mathbf{f}_i$ to V.

Thus, we have

$$|\log(\mathrm{vol}_i) - \log(\mathrm{vol}_{i-1})| = |\log d - \log d'|.$$

We will next see that d and d' are almost identical with very high probability.

Let $\mathbf{f}$ be a fixed vector (whose coordinates may depend on n). In what follows we denote the translation $\mathbf{f} + (s/n, \dots, s/n)$ of $\mathbf{f}$ by $\mathbf{f}'$.

Lemma 2.5. *Assume that $V \subset \mathbf{C}^n$ is a subspace of dimension $\dim(V) = k \le n - 10$. Let $\mathbf{x}' = (x'_1, \dots, x'_n)$ be a random vector where x'_i are i.i.d. copies of x from (1) and let d' be the distance from $\mathbf{x}' + \mathbf{f}$ to V. Then for any $t > 0$ we have*

$$\mathbf{P}_{\mathbf{x}'}(|d' - \sqrt{n - k + d^2_{\mathbf{f}'}}| \ge t + 3) \le \exp\left(-\frac{t^2}{4}\right),$$

where $d_{\mathbf{f}'}$ is the distance from $\mathbf{f}'$ to V.

Lemma 2.5 can be proved by using a well-known result of Talagrand; we defer its proof to Section 7.

As $\mathbf{E}(\sum_i x'_i) = s$ and $\mathbf{Var}(\sum_i x'_i) = \Theta(n)$, the probability that a random vector $\mathbf{x}'$ belongs to the set of $(-1,1)$ vectors of row-sum $s+1$ (or $s-1$) is $\Theta(1/\sqrt{n})$. Furthermore, condition on $\mathbf{x}' \in \mathcal{S}$, $\mathbf{x}'$ is uniformly distributed over these sets. We thus infer from Lemma 2.5 the following.

Corollary 2.6. *Let* $\mathbf{x}$ *be a vector uniformly sampled from* $\mathcal{S}$ *and let* d *be the distance from* $\mathbf{x}+\mathbf{f}$ *to* V*. Then for any* $t > 0$ *we have*

$$\mathbf{P}_{\mathbf{x}}(|d - \sqrt{n-k+d^2_{\mathbf{f}'}}| \geq t+3) = O\left(\sqrt{n}\exp\left(-\frac{t^2}{4}\right)\right).$$

One immediate consequence of Lemma 2.5 and Corollary 2.6 is that if $k \leq n - \log^4 n$, then by setting $t = \log n$, d is nonzero with probability at least $1 - O(\exp(-\log^{2-o(1)} n))$. By applying this fact m times, we conclude that all the vol_i are non-zero with probability at least $1 - O(\exp(-\log^{2-o(1)} n)$. So it is safe to assume that V has dimension exactly $m-1$ for any V spanned by $\mathbf{x}_1+\mathbf{f}_1, \dots, \mathbf{x}_{i-1}+\mathbf{f}_{i-1}, \mathbf{x}'_{i+1}+\mathbf{f}'_{i+1}, \dots, \mathbf{x}'_m+\mathbf{f}_m$. Next, by applying Lemma 2.5 and Corollary 2.6 once more, with probability $1 - O(\exp(-\log^{2-o(1)} n))$ with respect to $\mathbf{x}_i$ and $\mathbf{x}'_i$ we have

$$|d - \sqrt{n-m+1+d^2_{\mathbf{f}'_i}}| \leq \log n$$

and

$$|d' - \sqrt{n-m+1+d^2_{\mathbf{f}'_i}}| \leq \log n.$$

It then follows that

$$|\log d - \log d'| \leq \log\left(1 + \frac{2\log n}{\log^4 n - \log n}\right) = O(\log^{-2} n),$$

completing the proof of Lemma 2.4.

2.7. Sketch of the proof of Theorem 2.2

Our key lemma here is to show that the least singular value of $X_n + F_n + z\sqrt{n}I_n$, for any fixed complex number z, is at least $n^{-O(1)}$ with probability $1 - O(n^{-100})$.

Theorem 2.8. *Assume that F is a deterministic complex matrix of size $n \times n$ such that $|f_{ij}| \leq n^{\gamma}$ for some constant γ. Then for any $B > 0$ there exists $A > 0$ depending on B and γ such that*

$$\mathbf{P}\big(\sigma_n(X_n + F) < n^{-A}\big) \leq O(n^{-B}).$$

This theorem is an analog of the Bernoulli counterpart $X'_n + F$ whose proof can be found in either [33] or in other papers of the second author with Tao such as [29, 31, 32]. Unfortunately, these proofs do not seem to cover Theorem 3.1 in any trivial way. Henceforth, a large part of this note will be devoted to prove it, starting from Section 3.

We next invoke the following two linear algebra results.

Lemma 2.9 (Cauchy's interlacing law)**.** *[32, Lemma A.1] Let A be a matrix of size $n \times n$ and A' be the submatrix formed by the first $n - k$ rows of A. Let $\sigma_1(A) \geq \cdots \geq \sigma_n(A) \geq 0$ be the singular values of A, and similarly for A'. Then we have*

$$\sigma_i(A) \geq \sigma_i(A') \geq \sigma_{i+k}(A)$$

for every $1 \leq i \leq n - k$.

Lemma 2.10 (Negative second moment)**.** *[32, Lemma A.4] Let $1 \leq n' \leq n$, and let A' be a full rank matrix of size n' by n with singular values $\sigma_1(A') \geq \cdots \geq \sigma_n(A') \geq 0$ and rows $\mathbf{r}_1, \ldots, \mathbf{r}_{n'} \in \mathbf{C}^n$. For each $1 \leq i \leq n'$, let W_i be the subspace generated by the $n' - 1$ rows $\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, \mathbf{r}_{i+1}, \ldots, \mathbf{r}_{n'}$. Then we have*

$$\sum_{i=1}^{n'} \sigma_i^{-2}(A') = \sum_{i=1}^{n'} \operatorname{dist}^{-2}(\mathbf{r}_i, W_i).$$

We now prove Theorem 2.2. By Theorem 2.8, we can assume that $\mathbf{x}_1 + \mathbf{f}, \ldots, \mathbf{x}_n + \mathbf{f}$ spans the whole space $\mathbf{R}^n$ with probability at least $1 - O(n^{-100})$, and so in particular, all the V_i have full rank. Applying Lemma 2.10 for the matrix A' generated by the first k rows $\mathbf{x}_1 + \mathbf{f}, \ldots, \mathbf{x}_k + \mathbf{f}$ with any $k > m = n - \log^8 n$, we obtain the following with probability at least $1 - O(n^{-100})$

$$\operatorname{dist}^{-2}(\mathbf{x}_k + \mathbf{f}, V_{k-1}) < \sum_{i=1}^{k} \sigma_i^{-2}(A') = O(n^{O(1)}),$$

where in the RHS estimate we applied Lemma 2.9 and then Theorem 2.8.

Thus, for any $k > m$

$$O(n^{-O(1)}) = \operatorname{dist}(\mathbf{x}_k + \mathbf{f}, V_{k-1}) \leq \|\mathbf{x}_k + \mathbf{f}\| = O(\sqrt{n}). \tag{2}$$

Similarly, by applying the known variant of Theorem 2.8 for $(X'_n + F_n) - z\sqrt{n}I_n$ and by Lemmas 2.9 and 2.10 we also have

$$O(n^{-O(1)}) = \mathrm{dist}(\mathbf{x}'_k + \mathbf{f}, V'_{k-1}) = O(\sqrt{n}). \tag{3}$$

Owing to the estimates (2) and (3), we infer that

$$\mathbf{P}\left(\frac{1}{n}(|\log S_2| + |\log S'_2|) = O(\log^9 n/n)\right) = 1 - O(n^{-100}),$$

proving Lemma 2.2.

3. The least singular value bound

For the reader's convenience, we restate Theorem 2.8 below.

Theorem 3.1. *Assume that F is a deterministic complex matrix such that $|f_{ij}| \le n^\gamma$ for some constant γ. Then for any $B > 0$ there exists $A > 0$ depending on B and γ such that*

$$\mathbf{P}\big(\sigma_n(X_n + F) < n^{-A}\big) \le O(n^{-B}).$$

We refer the reader to [17] for a simple discrete version of Theorem 3.1. This section is devoted to provide an overview of our approach to prove Theorem 3.1; more details of the proofs will be discussed in subsequent sections.

We use the shorthand X for the matrix $X_n + F$. To prove Theorem 3.1, we assume that there exist vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbf{C}^n$ such that $\|\mathbf{a}\| = 1, \|\mathbf{b}\| < n^{-A}$ and

$$X\mathbf{a} = \mathbf{b}.$$

We next consider two cases.

Case 1. X is non-singular. Let $C(X) = (c_{ij}(X))$, $1 \le i, j \le n$, be the matrix of the cofactors of X. We then have

$$C(X)\mathbf{b} = \det(X) \cdot \mathbf{a}.$$

Thus,

$$\|C(X)\mathbf{b}\| = |\det(X)|.$$

By paying a factor of n in probability, without loss of generality we can assume that

$$|c_{11}(X)b_1 + \cdots + c_{1n}(X)b_n| \ge |\det(X)|/n^{1/2}.$$

Note that $\|\mathbf{b}\| \leq n^{-A}$, thus by Cauchy-Schwarz inequality

$$\sum_{i=1}^{n} |c_{1i}(X)|^2 \geq n^{2A-1} \det(X)^2. \tag{4}$$

We next express $\det(X)$ as a linear form of its first row $\mathbf{r}_1(X) = (x_1 + f_{11}, \dots, x_n + f_{1n})$

$$\det(Q) = x_1 c_{11}(X) + \cdots + x_n c_{1n}(X) + f_{11} c_{11}(X) + \cdots + f_{1n} c_{1n}(X).$$

Thus, with $c := \sqrt{\sum_i c_{1i}(X)^2}$ (which is $\neq 0$ as $(c_{11}, \dots, c_{1n}) \neq \mathbf{0}$), (4) can be rewritten as

$$\left| x_1 \frac{c_{11}(X)}{c} + \cdots + x_n \frac{c_{1n}(X)}{c} + \frac{1}{c}(f_{11} c_{11}(X) + \cdots + f_{1n} c_{1n}(Q)) \right| \leq n^{-A+1/2}.$$

Roughly speaking, our approach to prove Theorem 3.1 consists of two main steps.

- *Step 1.* Condition on X', the matrix of the last $n-1$ rows of X, if

$$\sup_v \mathbf{P}_{x_1,\dots,x_n} \left(\left| \sum_{i=1}^{n} x_i \frac{c_{1i}(X_n)}{c} - v \right| \leq n^{-A} \right) \geq n^{-B},$$

  then there is a strong structure among the cofactors c_{1i}.
- *Step 2.* The probability, with respect to X', that there is a strong additive structure among the c_{1i} is negligible.

We pause to discuss the structure mentioned in the inverse step. A set $Q \subset \mathbf{C}$ is a *GAP of rank* r if it can be expressed as in the form

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

for some $(g_0, \dots, g_r) \in \mathbf{C}^{r+1}$ and $(K_1, \dots, K_r), (K_1', \dots, K_r') \in \mathbf{Z}^r$.

It is convenient to think of Q as the image of an integer box $B := \{(k_1, \dots, k_r) \in \mathbf{Z}^r | K_i \leq k_i \leq K_i'\}$ under the linear map $\Phi : (k_1, \dots, k_r) \mapsto g_0 + k_1 g_1 + \cdots + k_r g_r$.

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

We are now ready to state our steps in details.

Theorem 3.2 (Step 1)**.** *Let $0 < \alpha < 1/2$ be a given constant. Assume that*

$$\rho^*_{n^{-A}}(\{v_1, \dots, v_n\}) := \sup_v \mathbf{P}_{x_1,\dots,x_n}\big(|\sum_{i=1}^n x_i v_i - v| \le n^{-A}\big) \ge n^{-B}$$

for some sufficiently large A, where $v_i = c_{1i}(X)/c$. Then there exists a vector $\mathbf{u} = (u_1, \dots, u_n)$ and a real number β of the form $n^{-A+k(5B+5+\gamma)}$ where $0 \le k \le A/(10B+10+2\gamma), k \in \mathbf{Z}$ such that the following holds.

- *$\|\mathbf{u}\| \asymp 1$ and $|\langle \mathbf{u}, \mathbf{r}_i(X)\rangle| \le \beta n^{5B+4+\gamma}$ for $n-1$ rows $\mathbf{r}_i$ of X.*
- *There exists a generalized arithmetic progression Q^* of rank $O_{\alpha,B}(1)$ and size $|Q^*| = \max(1, O_{\alpha,B}((\rho^*_{\beta n^{5B+4+\gamma}}(\{u_1, \dots, u_n\}))^{-1}/n^{\alpha/2}))$ which contains at least $n - n^{1/2+\alpha}$ complex numbers u_i.*
- *All the components of u_i and of the generators of Q^* are rational numbers of the form p/q, where $|p|, |q| \le n^{A+1}$.*

Roughly speaking, the quantity $(\rho^*_{\beta n^{5B+4+\gamma}}(\{u_1, \dots, u_n\}))^{-1}$ appearing in the bound of $|Q^*|$ guarantees that the containment is economical.

In the second step of the approach, we show that the probability for Q' having the above properties is negligible.

Theorem 3.3 (Step 2)**.** *With respect to X', the probability that there exists a vector $\mathbf{u}$ and a number β as in Theorem 3.2 is $\exp(-\Omega(n))$.*

We remark here that the choice of α being near $1/2$ would optimize the probability bound in Theorem 3.3. However, we prefer to keep α abstract to demonstrate the flexibility of our approach.

We now study the remaining case.

Case 2. X is singular. We show that the probability of this event is bounded by $O(n^{-B})$ for any $B > 0$, where the implied constant depends on B. The approach is identical (if not easier) to that of **Case 1**.

First of all, by paying a factor of n in probability and without loss of generality, it suffices to consider the event that $\mathbf{x}_1 + \mathbf{f}_1$ belongs to the subspace generated by $\mathbf{x}_2 + \mathbf{f}_2, \dots, \mathbf{x}_n + \mathbf{f}_n$. We show

Theorem 3.4. *Assume that X_n is a random matrix whose rows $\mathbf{x}_1, \dots, \mathbf{x}_n$ are independent random vectors sampled uniformly from $\mathcal{S}$. Then for any $B > 0$*

$$\mathbf{P}(\mathbf{x}_1 + \mathbf{f}_1 \text{ belongs to the subspace } H \text{ generated by } \mathbf{x}_2 + \mathbf{f}_2, \dots, \mathbf{x}_n + \mathbf{f}_n) = O(n^{-B}),$$

where the implied constant depends on B.

Condition on $\mathbf{x}_2, \dots, \mathbf{x}_n$, let $\mathbf{v} = (v_1, \dots, v_n)$ be a unit vector which is orthogonal to H. Then the probability that $\mathbf{x}_1 + \mathbf{f}_1 = (x_1 + f_{11}, \dots, x_n + f_{1n})$ belongs to H is bounded by $\mathbf{P}_{x_1,\dots,x_n}(x_1v_1 + \dots + x_nv_n + (f_{11}v_1 + \dots + f_{1n}v_n) = 0)$, and so crudely by

$$\mathbf{P}(\mathbf{x}_1 + \mathbf{f}_1 \in H) \leq \sup_v \mathbf{P}_{x_1,\dots,x_n}(|x_1v_1 + \dots + x_nv_n - v| \leq n^{-A}).$$

We again apply Theorem 3.2 to obtain a structural vector $\mathbf{u}$ and then use Theorem 3.3 to conclude that the probability for the existence of such $\mathbf{u}$ is negligible, completing the proof of Theorem 3.4.

The rest of the paper is organized as follows. In Section 4, we introduce our key lemmas. Theorems 3.2 and 3.3 will be proven in Sections 5 and 6 respectively.

4. The main tools for proving Theorem 3.2

We need to study the concentration of $\sum_i x_i v_i$ in a small ball, where $\mathbf{x} = (x_1, \dots, x_n)$ is sampled uniformly from the set $\mathcal{S}$ of all $(-1, 1)$ vectors of row-sum either $s-1$ or $s+1$. As customary, we first study a similar problem for $\mathbf{x}'$, a random vector whose components are i.i.d. copy of the Bernoulli variable x defined in (1).

Let $V = \{v_1, \dots, v_n\}$ be a multiset in $\mathbf{R}^d$, where d is a fixed integer. For $\beta > 0$, we define the *small ball probability* as

$$\rho_\beta(V) := \sup_{v \in \mathbf{R}^d} \mathbf{P}_{\mathbf{x}'}\big(v_1x'_1 + \dots + v_nx'_n \in B(v, \beta)\big),$$

where by $B(v, \beta)$ we denote the closed disk of radius β centered at v in $\mathbf{R}^d$.

A well-known result of Erdős [6] and Littlewood-Offord [16] asserts that if v_i are real numbers of magnitude $|v_i| \geq \beta$, then

$$\rho_\beta(V) = O(n^{-1/2}).$$

This remarkable inequality has generated an impressive way of research. We refer the reader to [12, 15, 20, 29] and the references therein for further discussion regarding these developments.

In the reverse direction, we would like to find the underlying reason as to why the small ball probability is large (say, polynomial in n).

Typical examples of V, where ρ_β is large, involve generalized arithmetic progressions introduced in the previous section.

Example 4.1. Let $Q = \{\sum_{i=1}^r k_i g_i | -K_i \le k_i \le K_i\}$ be a proper symmetric GAP of rank $r = O(1)$ and size $N = n^{O(1)}$ in $\mathbf{R}^d$. Assume that for each v_i there exists $q_i \in Q$ such that $\|v_i - q\| \le \delta$. Then, because the random sum $\sum_i q_i x'_i$ takes value in the GAP $nQ := \{\sum_{i=1}^r k_i g_i | - nK_i \le k_i \le nK_i\}$, and because $|nQ| \le n^r N = n^{O(1)}$, 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

$$\rho_{n\delta}(V) = n^{-O(1)}. \tag{5}$$

The above example shows that if v_i are *close* to a *GAP* of rank $O(1)$ and size $n^{O(1)}$ in $\mathbf{R}^d$, then V has large small ball probability. It was shown by Tao and the second author in [29, 31–33] and by the current authors in [20] that these are essentially the only examples of large small ball probability. We present here a somewhat optimal version.

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

Theorem 4.2 (Continuous Inverse Littlewood-Offord theorem for Bernoulli distribution)**.** *[20, Theorem 2.9] Let $0 < \alpha < 1/2; 0 < C$ be constants. Let $\beta > 0$ be a parameter that may depend on n. Suppose that $V = \{v_1, \dots, v_n\}$ is a multi-subset of $\mathbf{R}^d$ such that $\sum_{i=1}^n \|v_i\|^2 = 1$ and that V has large small ball probability*

$$\rho := \rho_\beta(V) \ge n^{-C},$$

where in the definition of ρ_β we assume $x'_1, \dots, x'_n$ to be i.i.d. copies of the Bernoulli random variable x defined in (1). Then for any number $n^\alpha \le n' \le n$, there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r k_i g_i : |k_i| \le K_i\}$ such that the following holds.

- *(Full dimension) There exists $\sqrt{\frac{n'}{\log n}} \ll k \ll \sqrt{n'}$ such that the dilate $P := (\beta/k)^{-1} \cdot Q$ contains the discrete hypercube $\{0,1\}^d$. Furthermore, P is an integral set, $P \subset \mathbf{Z}^d$.*
- *(Approximation) At least $n - n'$ elements of V (counting multiplicity) are $O(\frac{\beta}{k})$-close to Q.*
- *(Small rank and cardinality) Q has constant rank $d \le r = O(1)$, and small cardinality*

$$|Q| = \max\Big(1, O_{\alpha,d,C}(\rho^{-1} n'^{(-r+d)/2})\Big).$$

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

We note that [20, Theorem 2.9] was originally stated for more general distribution of the x'_i. Another slight difference is that we require P to be a subset of $\mathbf{Z}^d$ here. However, this additional fact is not new as it has been explicitly verified in the proof of Theorem 2.9 (see the last part of [20, Section 6]).

Remark 4.3. As noticed in [20, Corollary 2.10], the above theorem implies that if we use a coarser structure (which $O(\beta)$-approximates the v_i rather than $O(\beta/k)$-approximates as stated in Theorem 4.2), then we can obtain a bound of at most $\max(O(\rho^{-1}/\sqrt{n'}), 1)$ in the size of Q. As it turned out, the saving factor $1/\sqrt{n'}$ here plays a crucial role in any applications of Theorem 4.2 in the literature.

From now on we will be mainly working with $\mathbf{R}^2$ (equivalently, $\mathbf{C}$). Our method naturally extends to $\mathbf{R}^d$ for any fixed d, but we do not attempt to do so here. To prove Theorem 3.2, we need to modify our notion of concentration probability as follows. Let $V = \{v_1, \dots, v_n\}$ be a multiset in $\mathbf{R}^2$. For any $\beta > 0$, we define

$$\rho^*_\beta(V) := \sup_{v \in \mathbf{R}^2} \mathbf{P}_{\mathbf{x}}\big(v_1 x_1 + \cdots + v_n x_n \in B(v, \beta)\big),$$

where the probability is taken uniformly over all $(-1, 1)$ vectors $\mathbf{x} = (x_1, \dots, x_n)$ of given entry sum $\bar{s}$, where $|\bar{s}| \le (1 - \epsilon)n$. (In later application, we will set $\bar{s}$ to be either $s - 1$ or $s + 1$.)

By definition, ρ^* is invariant under translation. One observes that for any β and V we have

$$\rho_\beta(V) = \Omega(\rho^*_\beta(V)/\sqrt{n}). \tag{6}$$

This relation suggests that if $\rho^* := \rho^*_\beta(V)$ is large, then Theorem 4.2 (more precisely, Remark 4.3) implies that all the v_i can be approximated by a *GAP* Q of size $O((\rho^*)^{-1}\sqrt{n}/\sqrt{n'})$. This bound, unfortunately, falls short for any application as the saving factor $\sqrt{n}/\sqrt{n'}$ here is greater than 1 (we refer the reader to Remark 6.5 of Section 6 for more explanation).

The above discussion shows that a sole application of (6) is not enough to obtain a useful inverse result regarding ρ^*. In the following result, by using the extra translation invariance property of ρ^*, we provide a more economical inverse result.

Theorem 4.4 (Inverse Littlewood-Offord result with respect to ρ^*)**.** *Suppose that $V = \{v_1, \dots, v_n\}$ is a multi-subset of $\mathbf{R}^2$ such that $\sum_{i=1}^n \|v_i\|^2 = 1$ and that*

$$\rho^* := \rho^*_\beta(V) \ge n^{-C}$$

for some $\beta = O(n^{-21C-12})$. Then for any number $n^\alpha \le n' \le n$ there exists a proper GAP $Q^* = \{g_0 + \sum_{i=1}^r k_i g_i : |k_i| \le K_i\}$ such that

- *At least $n - n'$ elements of V are βn^{5C+3}-close to Q^*.*
- *Q^* has small rank $r = O(1)$, and small cardinality*

$$|Q| = \max\Big(1, O_{\alpha,C}((\rho^*)^{-1}\sqrt{n}/n')\Big).$$

- *There is a non-zero integer $p = O(\sqrt{n'})$ such that all steps $g_i = (g_{i1}, g_{i2}), 0 \le i \le r$ of Q^* have the form $g_{ij} = \beta \cdot \frac{p_{ij}}{p}$ with $p_{ij} \in \mathbf{Z}$ and $p_{ij} = O(\beta^{-1}\sqrt{n'})$.*

Note that the approximation in this case is not as fine as in Theorem 4.2(or as in Remark 4.3) and the structure Q^* is not necessarily symmetric. On the other hand, the size of Q^* is bounded by $O((\rho^*)^{-1}\sqrt{n}/n')$, which is considerably smaller than $O((\rho^*)^{-1}\sqrt{n}/\sqrt{n'})$ obtained by (6).

Before proving Theorem 4.4, let us provide a useful fact whose proof is simple and hence omitted.

Fact 4.5. *Assume that $P = \{k_1 g_1 + \cdots + k_r g_r | - K_i \le k_i \le K_i\}$ is a proper symmetric GAP which contains $w_1, \ldots, w_r$, where each w_i can be written as $k_{i1}g_1 + \cdots + k_{ir}g_r, k_{ij} \in \mathbf{Z}, |k_{ij}| \le K_i$.*

(i) Assume that the vectors $\mathbf{k}_i = (k_{i1}, \ldots, k_{ir}), 1 \le i \le r$, have full rank in $\mathbf{R}^r$. Then we can express each generator g_i as $g_i = y_{i1}w_1 + \cdots + y_{ir}w_r$, where y_{ij} are rational numbers of the form p/q with $|p|, |q| = O_r(|P|^r)$.

(ii) Assume that $\mathbf{k}_r$ belongs to the space spanned by $\mathbf{k}_1, \ldots, \mathbf{k}_{r-1}$, then we can write $\mathbf{k}_r$ as $\mathbf{k}_r = y_1\mathbf{k}_1 + \cdots + y_{r-1}\mathbf{k}_{r-1}$, where y_i are rational numbers of the form p/q with $|p|, |q| = O_r(|P|^r)$.

We now proceed to justify the main result of this section.

Proof of Theorem 4.4. Define a new set $U \subset \mathbf{R}^3$ as

$$U = \{u_1, \ldots, u_n\} := \left\{\frac{1}{2} \cdot \left(v_1, \frac{1}{\sqrt{n}}\right), \ldots, \frac{1}{2} \cdot \left(v_n, \frac{1}{\sqrt{n}}\right)\right\}.$$

By definition, we have $\sum_i \|u_i\|^2 = 1$ and $\rho^*_\beta(V) = \rho^*_{\beta/2}(U)$. Thus, by (6)

$$\rho_{\beta/2}(U) = \Omega(\rho^*_{\beta/2}(U)/\sqrt{n}) = \Omega(\rho^*_\beta(V)/\sqrt{n}) = \Omega(n^{-C-1/2}).$$

We apply Theorem 4.2 to U to obtain two GAPs Q and $P = (\beta/2)^{-1}k \cdot Q$ respectively. First, observe that if the rank r of Q (and P) is at least 5, then

$$|Q| = O((\rho^*)^{-1}\sqrt{n}/n'^{(r-3)/2}) = O((\rho^*)^{-1}\sqrt{n}/n'),$$

and so we are done by letting Q^* be the GAP generated by the first two coordinates of the generators of Q. Note that $g_0 = 0$ because Q is homogeneous. Also, we obtained a very good approximation (of order $O(\beta/k)$) in this case.

Next, we observe that r cannot be 3. Assume otherwise that $P = \{\sum_{i=1}^{3} k_i g_i : |k_i| \leq K_i\}$, where $g_i = (g_{i1}, g_{i2}, g_{i3}) \in \mathbf{Z}^3$ are the generators of P. Because $P \subset \mathbf{Z}^3$ and it contains $(1,0,0), (0,1,0)$ and $(0,0,1)$, by Fact 4.5 (i) the generators g_i must have the form (g_{i1}, g_{i2}, g_{i3}) where $|g_{ij}|$ are bounded by $O(|P|^3)$. But P has size $O(\rho_{\beta/2}^{-1}(U)) = O(n^{C+1/2})$, thus $|g_{ij}| = O(n^{3C+3/2})$. As a consequence, all of the elements of P must have norm at most $O(n^{4C+2})$. However, this is impossible because as one of the elements of P is $O(1)$-close to an element of $(\beta/2k)^{-1} \cdot U$, its second coordinate must be of order at least $\frac{\beta^{-1}k}{\sqrt{n}}$, which is greater than n^{4C+2} by the assumption of β of being sufficiently small.

We now consider the case $r = 4$, $P = \{\sum_{i=1}^{4} k_i g_i : |k_i| \leq K_i\}$, where $g_i = (g_{i1}, g_{i2}, g_{i3}) \in \mathbf{Z}^3$. Let $(w_1, l), \ldots, (w_{n-n'}, l)$ be the elements of P which are $O(1)$-close to $n - n'$ elements of the dilated set $(\beta/2k)^{-1} \cdot U$. Apparently, $l = \Theta(\beta^{-1}k/\sqrt{n})$. We next consider two cases.

Case 1. If all $\|w_i\|$ are smaller that n^{4C+2}, then we would be done because in this case the order of all $\|u_i\|$ is at most $O((\beta/2k)n^{4C+2})$, which is bounded by βn^{4C+2}.

Case 2. Assume otherwise that, say $\|w_1\| \geq n^{4C+2}$. Consider the following elements of P, $\mathbf{b}_1 := (1,0,0), \mathbf{b}_2 := (0,1,0), \mathbf{b}_3 := (0,0,1)$ and $\mathbf{b}_4 := (w_1, l)$. Because $\|w_1\|$ is greater than n^{4C+2}, one checks that the condition of Fact 4.5 (ii) does not hold for $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ and $\mathbf{b}_4$. We thus apply Fact 4.5 (i) to conclude that each g_i can be expressed as in the form $c_{i1}\mathbf{b}_1 + c_{i2}\mathbf{b}_2 + c_{i3}\mathbf{b}_3 + c_{i4}\mathbf{b}_4$, where $c_{ij} = p/q$ and $|p|, |q| = O(n^{4C+2})$.

Next, consider any $\mathbf{b} = (w_{i_0}, l)$ from the set $\{(w_1, l), \ldots, (w_{n-n'}, l)\}$. There exist $k_1, k_2, k_3, k_4 \in \mathbf{Z}, |k_i| \leq K_i$, such that $\mathbf{b} = k_1 g_1 + k_2 g_2 + k_3 g_3 + k_4 g_4$, and so

$$\begin{aligned}\mathbf{b} = {} & (k_1 c_{11} + k_2 c_{21} + k_3 c_{31} + k_4 c_{41})\mathbf{b}_1 + (k_1 c_{12} + x_2 k_{22} + x_3 c_{32} + k_4 c_{42})\mathbf{b}_2 \\ & + (k_1 c_{13} + k_2 c_{23} + k_3 c_{33} + k_4 c_{43})\mathbf{b}_3 + (k_1 c_{14} + k_2 c_{24} + k_3 c_{34} + k_4 c_{44})\mathbf{b}_4.\end{aligned}$$

Notice that $l = \Theta(\beta^{-1}k/\sqrt{n}) \geq n^{21C+11}$, meanwhile $|k_1c_{13} + k_2c_{23} + k_3c_{33} + k_4c_{43}| = O(n^{5C+5/2})$ and $|k_1c_{14} + k_2c_{24} + k_3c_{34} + k_4c_{44}| = \Theta(n^{-16C-8})$ as c_{ij} are rational numbers whose denominators are bounded by $O(n^{4C+4})$ and $k_1c_{14} + k_2c_{24} + k_3c_{34} + k_4c_{44}$ cannot be zero. We conclude that the coefficients of $\mathbf{b}_3$ and $\mathbf{b}_4$ must be 0 and 1 respectively,

It thus follows that, by considering the first two coordinates of $\mathbf{b}_1$ and $\mathbf{b}_2$,

$$\begin{aligned}\|w_{i_0} - w_1\|^2 &= \big((k_1c_{11} + k_2c_{21} + k_3c_{31} + k_4c_{41})^2 \\ &\quad + (k_1c_{12} + k_2c_{22} + k_3c_{32} + k_4c_{42})^2\big)^{1/2} \\ &= O(n^{5C+5/2}) < n^{5C+3}.\end{aligned}$$

Combining Cases 1 and 2, we infer that if $r = 4$, then all but n' elements of V are βn^{5C+3}-close to a common point. To complete the proof, we just simply set $g_0 = \beta n^{5C+3} \cdot p$ be this approximated point where p is a complex number of integral coordinates and $|p| \leq \beta^{-1}n^{-5C-3}$. We set other generators to be zero. □

We now deduce an important corollary of Theorem 4.4 which, similarly to the result of Erdős and Littlewood-Offord, states that as long as the multi-set V is not too degenerated (for a given β), its concentration probability ρ^* must be small.

Corollary 4.6. *Let $0 < \alpha < 1/2$ be a positive constant and let n' be a number satisfying $n^{1/2+\alpha} < n' < n$. Assume that $\beta \leq n^{-24}$ and V is a multi-set in $\mathbf{R}^2$ so that any of its $n - n'$ elements cannot be βn^6-close to a common point. Then we have*

$$\rho^*_\beta(V) = O(\sqrt{n}/n').$$

Proof of Corollary 4.6. Assume otherwise that $\rho^*_\beta(V) \geq C\sqrt{n}/n'$ for some large constant C to be chosen. So

$$\rho^*(V) \geq Cn^{-1/2}.$$

We next apply Theorem 4.4 to V to obtain a GAP Q^* which is $\beta n^{11/2}$ to all but $n - n'$ elements of V. Notice that because there are no more than $n - n' - 1$ elements of V that are βn^6-close to one common point, Q^* must have size at least 2. On the other hand, from the conclusion of

Theorem 4.4, assuming that C is sufficiently large depending on α, the size of Q^* is bounded by

$$|Q^*| = \max(1, O_\alpha((\rho^*)^{-1}\sqrt{n}/n') = \max\left(1, O_\alpha\left(\frac{1}{C}\right)\right) = 1.$$

This contradiction completes the proof of our corollary. □

5. Proof of Theorem 3.2

We will invoke Theorem 4.4. Define a radius sequence $(\beta_k)_0^\infty$ where $\beta_0 := n^{-A}$ and

$$\beta_{i+1} = n^{5B+5+\gamma}\beta_i.$$

Let V be the multi-set of $v_1, \ldots, v_n$. Then the assumption of Theorem 3.2 becomes

$$\rho_{\beta_0}{}^*(V) \geq n^{-B}$$

with either $\bar{s} = s - 1$ or $\bar{s} = s + 1$.

Next, because the increasing sequence $\rho^*_{\beta_i}(V)$ is bounded from above by 1, by pigeonhole principle there exists $0 \leq k_0 \leq 2B/\alpha$ such that

$$\rho^*_{\beta_{k_0+1}}(V) \leq n^{\alpha/2}\rho^*_{\beta_{k_0}}(V).$$

As A was chosen to be sufficiently large, one has $\beta_{k_0} \leq n^{-A/2}$. We next apply Theorem 4.4 to V with $n' = n^{1/2+\alpha}$ and $\beta = \beta_{k_0}$ to obtain a GAP $Q^* = \{g_0 + \sum_{i=1}^r k_i g_i, |k_i| \leq K_i\}$ for which the following holds.

- Q^* has small rank $r = O(1)$, and small cardinality

$$|Q^*| = \max\left(1, O_{\alpha,B}\big((\rho^*_{\beta_{k_0}}(V))^{-1}/n^\alpha\big)\right).$$

- There are $n_0 := n - n^{1/2+\alpha}$ elements $v_{i_1}, \ldots, v_{i_{n_0}}$ of V which are $O(\beta_{k_0} n^{5B+3})$-close to $n - n^{1/2+\alpha}$ elements $u_1, \ldots, u_{n_0}$ of Q^*.
- There is a non-zero integer $p = O(\sqrt{n^{1/2+\alpha}})$ such that all steps $g_i = (g_{i1}, g_{i2}), 0 \leq i \leq r$ of Q^* have the form $g_{ij} = \beta_{k_0}^{-1} p_{ij}/p$ with $p_{ij} \in \mathbf{Z}$ and $p_{ij} = O(\beta_{k_0}^{-1}\sqrt{n^{1/2+\alpha}})$. In particular, all the components of the elements of Q^* have the form p/q where $|p|, |q| \leq n^{A+1}$.

Next, for each v of the remaining $n^{1/2+\alpha}$ exceptional elements of V (which are not close to any element of Q^*), we trivially approximate it by a complex number v whose components are rational numbers of the form p/q with $|q| \leq n^{A+1}$ such that $|u - v| \leq \beta_{k_0} n^{5B+3}$.

By the approximation, we infer that

$$\|\mathbf{u}-\mathbf{v}\| = \left(\sum_i |u_i - v_i|^2\right)^{1/2} \le \beta_{k_0} n^{5B+7/2}.$$

Taking into account that $|f_{ij}| \le n^\gamma$, we thus have

$$\begin{aligned}\rho^*_{\beta_{k_0}}(V) &\le \rho^*_{\beta_{k_0}+\beta_{k_0}n^{5B+7/2+\gamma}}(U) \le \rho^*_{\beta_{k_0}n^{5B+4+\gamma}}(U)\\ &\le \rho^*_{\beta_{k_0}+\beta_{k_0}n^{5B+4+\gamma}}(V) \le \rho^*_{\beta_{k_0}n^{5B+5+\gamma}}(V) = \rho^*_{\beta_{k_0+1}}(V),\end{aligned}$$

where U is the multi-set $\{u_1,\dots,u_n\}$.

From the estimate above, as $\rho^*_{\beta_{k_0+1}}(V) \le n^{\alpha/2}\rho^*_{\beta_{k_0}}(V)$, it is implied that

$$\rho^*_{\beta_{k_0}n^{5B+4+\gamma}}(U) \le n^{\alpha/2}\rho^*_{\beta_{k_0}}(V).$$

So the size of Q^* is bounded by

$$|Q^*| = \max\left(1, O\big((\rho^*_{\beta_{k_0}n^{5B+4+\gamma}}(U))^{-1}/n^{\alpha/2}\big)\right).$$

In summary, we have obtained a vector $\mathbf{u} = (u_1,\dots,u_n)$ which satisfies the following properties.

- $\|\mathbf{u}\| \asymp 1$, and because $\langle \mathbf{v}, \mathbf{r}_i(X)\rangle = 0$ for any row $\mathbf{r}_i$ of X of index $i \ge 2$, we also have $|\langle \mathbf{u}, \mathbf{r}_i(X)\rangle| \le \beta_{k_0} n^{5B+4+\gamma}$.
- There exists a generalized arithmetic progression Q^* of rank $O_{B,\alpha}(1)$ and size $|Q^*| = \max(1, O((\rho^*_{\beta_{k_0}n^{5B+4+\gamma}}(U))^{-1}/n^{\alpha/2}))$ that contains at least $n - n^{1/2+\alpha}$ complex numbers u_i.
- All the components of u_i and of the generators of Q^* are rational numbers of the form p/q, where $|p|, |q| \le n^{A+1}$.

This completes the proof of Theorem 3.2.

6. Proof of Theorem 3.3

By applying Theorem 3.2, we obtain a structural vector $\mathbf{u}$ which satisfies all the described properties. Because the number of β is bounded by a constant, it is enough to verify Theorem 3.3 for one such β. By paying a factor of n in probability, we assume that $|\langle \mathbf{u}, \mathbf{r}_i(X)\rangle| \le \beta n^{5B+4+\gamma}$ for the last $n-1$ rows of X.

Set $\beta' := \beta n^{5B+4+\gamma}$. We will consider two cases depending on the structure of $\mathbf{u}$.

6.1. Degenerate u

We first consider the probability $\mathbf{P}_{\mathbf{major}}$ of the event $|\langle \mathbf{r}_i, \mathbf{u}\rangle| \le \beta', 2 \le i \le n$, for which there are $n_0 := n - n^{1/2+\alpha}$ complex numbers u_i which can be $\beta' n^4$-approximated by a common point $u_0' \in \beta' n^4 \cdot \mathbf{Z}^2$.

By paying a factor $\binom{n}{n_0}$ in probability, we may assume that this point approximates the first n_0 complex numbers $u_1, \dots, u_{n_0}$. Thus, by approximating the remaining u_i by $u_i' \in \beta' n^4 \cdot \mathbf{Z}^2$ such that $|u_i - u_i'| \le \beta' n^4$, the events $|\langle \mathbf{r}_i, \mathbf{u}\rangle| \le \beta'$ belongs to the event $|\langle \mathbf{r}_i, \mathbf{u}'\rangle| \le \beta' n^5$, where $\mathbf{u}' = (u_1', \dots, u_1', u_{n_0+1}', \dots, u_n')$ and $\|\mathbf{u}'\| \asymp 1$.

Let $X_{(n-1)\times n}$ be the matrix generated by the last $n-1$ rows of X, and let X' be the $n-1$ by $n-n_0$ matrix obtained from $X_{(n-1)\times n}$ by joining its first n_0 columns,

$$X' = \big[\mathbf{c}_1(X_{(n-1)\times n}) + \cdots + \mathbf{c}_{n_0}(X_{(n-1)\times n}), \\ \mathbf{c}_{n_0+1}(X_{(n-1)\times n}), \dots, \mathbf{c}_n(X_{(n-1)\times n})\big].$$

By definition, the row vectors of X' satisfy $|\langle \mathbf{r}_i(X'), \mathbf{u}_{\mathrm{tr}}'\rangle| \le \beta' n^5$ where $\mathbf{u}_{\mathrm{tr}}' := (u_1', u_{n_0+1}', \dots, u_n')$. It also follows from definition that the i-th row of X' has the form $\mathbf{r}_i(X') = \mathbf{x}' + \mathbf{f}'$, where $\mathbf{f}' = (f_{i1} + \cdots + f_{in_0}, f_{i(n_0+1)}, \dots, f_{in})$ and $\mathbf{x}' = (x_1 + \cdots + x_{n_0}, x_{n_0+1}, \dots, x_n) := (x_1', \dots, x_{n-n_0}')$.

As $\mathbf{x}$ is sampled uniformly from $\mathcal{S}$, the set of all $(-1,1)$ vectors of entry-sum either $s-1$ or $s+1$, $\mathbf{x}'$ is a random vector chosen from *type 1* or *type 2* defined below.

Type 1 (row-sum $s+1$)**.**

$$\mathbf{P}(x_1' = k) = \frac{\binom{n_0}{(n_0+k)/2}\binom{n-n_0}{(n-n_0+s+1-k)/2}}{\binom{n}{n/2+(s-1)/2} + \binom{n}{n/2+(s+1)/2}}$$

for all k such that $k + n_0$ is even; and $(x_2', \dots, x_{n-n_0}')$ are chosen uniformly from all $(-1,1)$ vectors of row-sum $s + 1 - x_1'$.

Type 2 (row-sum $s-1$)**.**

$$\mathbf{P}(x_1' = k) = \frac{\binom{n_0}{(n_0+k)/2}\binom{n-n_0}{(n-n_0+s-1-k)/2}}{\binom{n}{n/2+(s-1)/2} + \binom{n}{n/2+(s+1)/2}}$$

for all k such that $k + n_0$ is even; and $(x_2', \dots, x_{n-n_0}')$ are chosen uniformly from all $(-1,1)$ vectors of row-sum $s - 1 - x_1'$.

It is clear that

$$\mathbf{P}(\mathbf{x}' \in \text{type 1}) = \frac{\binom{n}{n/2+(s+1)/2}}{\binom{n}{n/2+(s-1)/2} + \binom{n}{n/2+(s+1)/2}}$$

and

$$\mathbf{P}(\mathbf{x}' \in \text{type 2}) = \frac{\binom{n}{n/2+(s-1)/2}}{\binom{n}{n/2+(s-1)/2} + \binom{n}{n/2+(s+1)/2}}.$$

Observe that as $|s| \le (1-\epsilon)n$, these two probabilities are comparable, each of which can be bounded crudely from below by $(1-\epsilon)/4$.

We next apply the following result.

Claim 6.2. *Let $\epsilon < 1/4$ be a fixed constant. Let $\mathbf{u}'_{\text{tr}} = (u'_1, u'_{n_0+1}, \dots, u'_n)$ be a vector in which the components of each complex u'_i is of the form $\beta' n^4 \cdot \mathbf{Z}$ and such that $n_0|u'_1|^2 + |u'_{n_0+1}|^2 + \dots + |u'_n|^2 \asymp 1$. Then, as n is sufficiently large and $\mathbf{f}$ is a fixed vector, one has*

$$\mathbf{P}_{\mathbf{x}'}(|\langle \mathbf{x}' + \mathbf{f}', \mathbf{u}'_{\text{tr}} \rangle| \le \beta' n^5) \le 1 - (1-\epsilon)/8.$$

Proof of Claim 6.2. We will consider two main cases below.

(i) We first assume that there exists $1 < i_0 < j_0$ such that $|u'_{i_0} - u'_{j_0}| \ge \beta' n^5$. Without loss of generality, assume that $i_0 = n-1$ and $j_0 = n$. It follows from the distribution of $\mathbf{x}'$ that the event of having exactly one -1 among the last two components of $\mathbf{x}'$ happens with probability at least $(1-\epsilon)/4$ asymptotically. Within this event, observe that for any tuple $(x'_1, \dots, x'_{n-n_0-2})$, either $\mathbf{x} = (x'_1, \dots, x'_{n-n_0-2}, -1, 1)$ or $\mathbf{x} = (x'_1, \dots, x'_{n-n_0-2}, 1, -1)$ does not satisfy $|\langle \mathbf{x}', \mathbf{u}'_{\text{tr}} \rangle + \langle \mathbf{f}', \mathbf{u}'_{\text{tr}} \rangle| \le \beta' n^5$. Thus, we have

$$\mathbf{P}_{\mathbf{x}'}(|\langle \mathbf{x}', \mathbf{u}'_{\text{tr}} \rangle + \langle \mathbf{f}', \mathbf{u}'_{\text{tr}} \rangle| \le \beta' n^5) \le 1 - (1-\epsilon)/8.$$

(ii) Assume otherwise that there exists u' such that all $|u' - u'_{n_0+1}|, \dots, |u' - u'_n|$ are bounded by $\beta' n^5$. In this case, the inequality $|\langle \mathbf{x}', \mathbf{u}'_{\text{tr}} \rangle + \langle \mathbf{f}', \mathbf{u}'_{\text{tr}} \rangle| \le \beta' n^5$ implies that

$$|x'_1(u'_1 - u') + u'(x'_1 + x'_2 + \dots + x'_{n-n_0}) + \langle \mathbf{f}', \mathbf{u}'_{\text{tr}} \rangle| \le \beta' n^6. \tag{7}$$

We next consider the subcase $|u'_1 - u'| \ge \beta' n^8$. If $x'_1 + x'_2 + \dots + x'_{n-n_0} = s+1$, then (7) implies that x'_1 belongs to the interval $\big[(-u'(s+1) - \langle \mathbf{f}', \mathbf{u}'_{\text{tr}} \rangle)(\beta' n^8)^{-1} - 1/n^2, (-u'(s+1) - \langle \mathbf{f}', \mathbf{u}'_{\text{tr}} \rangle)(\beta' n^8)^{-1} + 1/n^2\big]$. However, because this interval has length $2/n^2$, and so this probability is clearly bounded

by $\sup_k \mathbf{P}(x'_1 = k)$, which is clearly smaller than $1 - (1-\epsilon)/4$. We argue similarly for the case $x'_1 + x'_2 + \cdots + x'_{n-n_0} = s - 1$.

For the remaining subcase $|u'_1 - u'| \le \beta' n^8$, as A was chosen to be large enough, we have $|u'_1 - u'| \le n^{-2}$. Next, because $\|\mathbf{u}'_{\mathrm{tr}}\|^2 = n_0|u'_0|^2 + (n - n_0)|u'|^2 \asymp 1$, we infer that $|u'| \asymp 1/\sqrt{n}$. It then follows that

$$|u'(x'_1 + \cdots + x'_{n-n_0}) + \langle \mathbf{f}', \mathbf{u}'_{\mathrm{tr}} \rangle| \le \beta' n^9. \tag{8}$$

However, as $x'_1 + \cdots + x'_{n-n_0}$ takes value $s+1$ and $s-1$ each with probability at least $(1-\epsilon)/4$, the equation (8) above holds with probability at most $1 - (1-\epsilon)/4$. □

Now we estimate $\mathbf{P}_{\mathbf{major}}$. As the event $|\langle \mathbf{r}_i(X), \mathbf{u}' \rangle| \le \beta' n^5$ is controlled by $|\langle \mathbf{r}_i(X'), \mathbf{u}'_{\mathrm{tr}} \rangle| \le \beta' n^5$, and by Claim 6.2 the later holds with probability $(7+\epsilon)/8$, it follows that the probability that $|\langle \mathbf{r}_i(X), \mathbf{u}' \rangle| \le \beta' n^5$ for all $2 \le i \le n$ is bounded by $((7+\epsilon)/8)^{n-1}$.

Additionally, an elementary computation implies that the number of structural vectors $\mathbf{u}' \in (\beta' n^4 \cdot \mathbf{Z}^2)^{n-n_0+1}$ satisfying $\|\mathbf{u}'\| \asymp 1$ is bounded by

$$((\beta' n^4)^{-1})^{n-n_0+1} = O((n^A)^{n^{1/2+\epsilon}+1}) = O(n^{O_A(n^{1/2+\epsilon})}).$$

Putting together, we obtain the following bound for $\mathbf{P}_{\mathbf{major}}$

$$\mathbf{P}_{\mathbf{major}} = O(n^{O_A(n^{1/2+\epsilon})}) \binom{n}{n_0} \binom{n-1}{n-n_0-1} \left(\frac{7+\epsilon}{8}\right)^{n-1} = \left(\frac{7+\epsilon}{8}\right)^{(1-o(1))n}.$$

Remark 6.3. In the treatment above the fact that $\mathbf{x}'$ takes either type 1 or type 2 with comparable probability is crucial. The assumption of just one type would not be enough to estimate $\mathbf{P}_{\mathbf{major}}$ unless we had an additional assumption on $\mathbf{u}'$, say $u'_1 + \cdots + u'_n$ is nearly zero.

6.4. Non-degenerate u

We consider the probability $\mathbf{P}_{\mathbf{minor}}$ of the event that there exists a vector $\mathbf{u}$ for which $|\langle \mathbf{r}_i(X), \mathbf{u} \rangle| \le \beta', 2 \le i$ and the following holds

- $\|\mathbf{u}\| \asymp 1$ and there does not exist any u which is $\beta' n^4$-close to all but $n^{1/2+\alpha}$ complex numbers u_i. Thus, it follows from Corollary 4.6 that

$$\rho^*_{\beta'}(U) = O(n^{-\alpha}).$$

- There exists a generalized arithmetic progression Q^* of rank $O_{B,\alpha}(1)$ and size $|Q^*| = \max\left(1, O(\rho^*_{\beta'}(U)^{-1}/n^{\alpha/2})\right) = O(\rho^*_{\beta'}(U)^{-1}/n^{\alpha/2})$ that

contains at least $n - n^{1/2+\alpha}$ complex numbers u_i. (Here we used the estimate $\rho^*_{\beta'}(U)^{-1} = \Omega(n^\alpha)$ to eliminate the trivial constant 1 in the size estimate of Q^*.)

- All the components of u_i and of the generators of the generalized arithmetic progression are rational numbers of the form p/q, where $|p|, |q| \le n^{A+1}$.

Let $0 < \delta$ to be chosen (any $\delta < \alpha/3$ will suffice). We divide the interval $[n^{-B}, O_\alpha(n^{-\alpha/2})]$ into sub-intervals $[n^{-(k+1)\delta}, n^{-k\delta}]$, where $\alpha/2\delta \le k \le B/\delta$. For each k, let $\mathbf{G}_k$ be the collection of $\mathbf{u}$'s such that $\rho^*_{\beta'}(U) \in [n^{-(k+1)\delta}, n^{-k\delta}]$, and let $\mathbf{P}_k$ be the probability that $|\langle \mathbf{r}_i(X'), \mathbf{u}\rangle| \le \beta'$ for all i and for one of $\mathbf{u}$ from $\mathbf{G}_k$.

We now bound the size of $\mathbf{G}_k$. To do this, we first count the number of GAPs which may contain most of the u_i of vectors $\mathbf{u}$ from $\mathbf{G}_k$, and then count the number of $\mathbf{u}$'s whose u_i are chosen from the determined structure. Recall that all components of the GAP generators are of the form p/q, where $|p|, |q| \le n^{A+1}$. Because each GAP has rank $O_{B,\alpha}(1)$ and size $O((\rho^*)^{-1}/n^{\alpha/2}) = O(n^{\delta(k+1)}/n^{\alpha/2})$, the number of such GAPs is bounded by

$$(n^{4A+4})^{O_{B,\alpha}(1)}(n^{\delta(k+1)}/n^{\alpha/2})^{O_{B,\alpha}(1)} = O(n^{O_{B,\alpha,\delta}(1)}).$$

After choosing a Q^* of size $O(n^{\delta(k+1)}/n^{\alpha/2})$, the number of ways to choose $n - n^{1/2+\alpha}$ complex numbers u_i as Q^*'s elements is

$$\binom{n}{n^{1/2+\alpha}}\binom{O(n^{\delta(k+1)}/n^{\alpha/2})}{n - n^{1/2+\alpha}} = O\big(n^{n^{1/2+\alpha}}(n^{\delta(k+1)}/n^{\alpha/2})^{n-n^{1/2+\alpha}}\big).$$

For the remaining $n^{1/2+\alpha}$ exceptional elements, there are $(n^{4A+4})^{n^{1/2+\alpha}} = O(n^{O_A(n^{1/2+\alpha})})$ ways to choose them. Putting these bounds together, we obtain the following bound for the number of $\mathbf{u}$ of $\mathbf{G}_k$

$$|\mathbf{G}_k| = O\big(n^{O_{A,B,\alpha,\delta}(n^{1/2+\alpha})}(n^{\delta(k+1)}/n^{\alpha/2})^{n-n^{1/2+\alpha}}\big).$$

Now, for a given $\mathbf{u} \in \mathbf{G}_k$ the probability that $|\langle \mathbf{r}_i(X), \mathbf{u}\rangle| \le \beta'$ for all $2 \le i \le n$ is bounded by $(\rho^*_{\beta'}(\mathbf{u}))^{n-1} \le (n^{-\delta k})^{n-1}$. Thus, we can estimate $\mathbf{P}_k$ as

$$\begin{aligned}\mathbf{P}_k &\le |\mathbf{G}_k|(n^{-\delta k})^{n-1} \\ &= O\left(n^{O_{A,B,\alpha,\delta}(n^{1/2+\alpha})}(n^\delta)^n/(n^{\alpha/2})^{n-n^{1/2+\alpha}}\right) = o(n^{-\alpha n/6}),\end{aligned}$$

provided that δ was chosen to be smaller than $\alpha/3$.

Summing over k, we thus obtain

$$\mathbf{P}_{\mathbf{minor}} = \sum_{k \le B/\delta} \mathbf{P}_k = o(n^{-\alpha n/6}).$$

Remark 6.5. One observes that the saving factor $1/n^{\alpha/2}$ in the size of Q^* plays a key role in our analysis here. This explains the necessity of Theorem 4.4.

7. Concentration of distance

We now give a proof of Lemma 2.5 basing on [26]. Let $P = (p_{ij})$ be the n by n orthogonal projection matrix from $\mathbf{C}^n$ to $V^\perp$. Thus, P is Hermitian and $P^2 = P$. We first normalize x'_i by setting $y'_i := x'_i - s/n$ and $f'_i := f_i + s/n$ for $1 \le i \le n$. We then have $\mathbf{E}y'_i = 0, \mathbf{Var}(y'_i) = 1 - (s/n)^2$ and

$$\begin{aligned} d'^2 = \|P(\mathbf{f} + \mathbf{x}')\|^2 = \|P(\mathbf{f}' + \mathbf{y}')\|^2 &= \sum_{ij} p_{ij}(y'_i + f'_i)\overline{(y'_j + f'_j)} \\ &= \sum_{ij} p_{ij} y'_i y'_j + \sum_{ij} y'_i(p_{ij}\overline{f'_j} + p_{ji}f'_j) + \sum_{ij} p_{ij} f'_i \overline{f'_j} \\ &= \mathbf{Tr}(P) + \sum_{i \ne j} p_{ij} y'_i y'_j + \sum_{ij} y'_i(p_{ij}\overline{f'_j} + p_{ji}f'_j) \\ &\quad + \sum_{ij} p_{ji} y'_i f'_j + d^2_{\mathbf{f}'} \\ &:= (n-k) + d^2_{\mathbf{f}'} + Y. \end{aligned}$$

It is clear that $\mathbf{E}Y = 0$, thus

$$\mathbf{E}(d^2) = (n-k) + d^2_{\mathbf{f}'}.$$

Note that

$$\begin{aligned} \mathbf{E}|Y|^2 &= \mathbf{E}|\sum_{i \ne j} p_{ij} y'_i y'_j + \sum_{ij} y'_i(p_{ij}\overline{f'_j} + p_{ji}f'_j)|^2 \\ &= \mathbf{E}|\sum_{i \ne j} p_{ij} y'_i y'_j|^2 + \mathbf{E}|\sum_{ij} y'_i(p_{ij}\overline{f'_j} + p_{ji}f'_j)|^2 \\ &= (1 - (s/n)^2)\Big[\sum_{i \ne j} |p_{ij}|^2 + \sum_i |\sum_j p_{ij}\overline{f'_j} + \sum_j p_{ji} f'_j|^2\Big] \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i\neq j} |p_{ij}|^2 + 4\sum_i (\Re(\sum_j p_{ji} f'_j))^2 \\
&\leq \sum_{i\neq j} |p_{ij}|^2 + 4\sum_i |\sum_j p_{ji} f'_j|^2 \\
&= \sum_{i\neq j} |p_{ij}|^2 + 4\sum_{j_1 j_2}\sum_i p_{j_1 i}\overline{p_{j_2 i}} f'_{j_1}\overline{f'_{j_2}} \\
&= \sum_{i\neq j} |p_{ij}|^2 + 4\sum_{j_1 j_2} p_{j_1 j_2} f'_{j_1}\overline{f'_{j_2}} = \sum_{i\neq j} p_{ij}^2 + 4d_{\mathbf{f}'}^2.
\end{aligned}$$

Next, because $\sum_i p_{ii} = (n-k)$, by Cauchy-Schwarz inequality

$$\sum_i p_{ii}^2 \geq (n-k)^2/n.$$

Thus,

$$\sum_{i\neq j} |p_{ij}|^2 = \sum_{i,j} |p_{ij}|^2 - \sum_i p_{ii}^2 \leq (n-k) - (n-k)^2/n \leq \min(k, n-k).$$

It is implied that

$$\mathbf{E}Y^2 \leq \min(k, n-k) + 4d_{\mathbf{f}'}^2.$$

Consider the event $d \geq \sqrt{n-k+d_{\mathbf{f}'}^2} + 3$. The probability of this event is bounded from above by

$$\begin{aligned}
&\mathbf{P}\big(d'^2 \geq n-k+d_{\mathbf{f}'}^2 + 6\sqrt{n-k+d_{\mathbf{f}'}^2}\big) \\
&= P\big(Y \geq 6\sqrt{n-k+d_{\mathbf{f}'}^2}\big) \\
&\leq \mathbf{P}\big(Y^2 \geq 36(n-k+d_{\mathbf{f}'}^2)\big) \\
&\leq \frac{\mathbf{E}Y^2}{36(n-k+d_{\mathbf{f}'}^2)} \leq \frac{1}{9}.
\end{aligned}$$

Similarly, consider the event $d' \leq \sqrt{n-k+d_{\mathbf{f}'}^2} - 3$. The probability of this event is bounded from above by

$$\mathbf{P}\big(d'^2 \leq n-k+d_{\mathbf{f}'}^2 - 6\sqrt{n-k+d_{\mathbf{f}'}^2} + 9\big)$$

$$= P\big(Y \le -6\sqrt{n-k+d_{\mathbf{f}'}^2} + 9\big)$$
$$\le \mathbf{P}\big(Y^2 \ge 36(n-k+d_{\mathbf{f}'}^2) - 108\sqrt{n-k+d_{\mathbf{f}'}^2} + 81\big)$$
$$\le \frac{\mathbf{E}Y^2}{36(n-k+d_{\mathbf{f}'}^2) - 108\sqrt{n-k+d_{\mathbf{f}'}^2} + 81} \le \frac{1}{4}$$

provided that $k \le n - 10$.

Thus, the median M of d' satisfies $|M - \sqrt{n-k+d_{\mathbf{f}'}^2}| \le 3$.

Since the distance function is convex on $\{-1,1\}^n$ with Lipschitz constant 1. Talagrand's concentration inequality [24] implies that for any t

$$\mathbf{P}(|d' - M| \ge t) \le 4\exp(-t^2/16).$$

Since $|M - \sqrt{n-k+d_{\mathbf{f}'}^2}| \le 3$, Lemma 2.5 follows.

References

[1] Z. D. Bai (1997). Circular law. *Ann. Probab.* **25**, 494–529. MR1428519

[2] Z. D. Bai and J. Silverstein (2006). *Spectral Analysis of Large Dimensional Random Matrices*, Mathematics Monograph Series 2, Science Press, Beijing.

[3] C. Bordenave, P. Caputo and D. Chafai (2012). Circular law theorem for random Markov matrices. *Probability Theory and Related Fields* **152**(3–4), 751–779. MR2892961

[4] D. Chafai (2010). The Dirichlet Markov ensemble. *Journal of Multivariate Analysis* **101**, 555–567. MR2575404

[5] K. Costello, T. Tao and V. Vu (2006). Random symmetric matrices are almost surely non-singular. *Duke Mathematics Journal* **135**, 395–413. MR2267289

[6] P. Erdős (1945). On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51**, 898–902. MR0014608

[7] P. Erdős and L. Moser (1947). Elementary problems and solutions: Solutions: E736. *Amer. Math. Monthly* **54**(4), 229–230. MR1526680

[8] C. G. Esséen (1966). On the Kolmogorov-Rogozin inequality for the concentration function. *Z. Wahrsch. Verw. Gebiete* **5**, 210–216. MR0205297

[9] V. L. Girko (1984). Circular law, *Theory Probab. Appl.*, 694–706. MR0773436

[10] V. L. Girko (2004). The strong circular law, Twenty years later, II. *Random Oper. Stochastic Equations* **12**(3), 255–312. MR2085255

[11] F. Götze and A. N. Tikhomirov (2010). The circular law for random matrices. *Annals of Probability* **38**(4), 1444–1491. MR2663633

[12] G. Halász (1977). Estimates for the concentration function of combinatorial number theory and probability. *Period. Math. Hungar.* **8**(3–4), 197–211. MR0494478

[13] J. Kahn, J. Komlós and E. Szemerédi (1995). On the probability that a random ±1 matrix is singular. *J. Amer. Math. Soc.* **8**, 223–240. MR1260107

[14] G. Katona (1966). On a conjecture of Erdős and a stronger form of Sperner's theorem. *Studia Sci. Math. Hungar.* **1**, 59–63. MR0205864

[15] D. Kleitman (1970). On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors. *Advances in Math.* **5**, 155–157. MR0265923

[16] J. E. Littlewood and A. C. Offord (1943). On the number of real roots of a random algebraic equation. *III. Rec. Math. Mat. Sbornik N.S.* **12**, 277–286. MR0009656

[17] H. Nguyen (2013). On the singularity of random combinatorial matrices. *SIAM J. Discrete Mathematics* **27**(1), 447–458. MR3032929

[18] H. Nguyen (2012). Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. *Duke Mathematics Journal* **161**(4), 545–586. MR2891529

[19] H. Nguyen, Random doubly stochastic matrices: The circular law, submitted.

[20] H. Nguyen and V. Vu (2011). Optimal Littlewood-Offord theorems. *Advances in Mathematics* **226**(6), 5298–5319. MR2775902

[21] G. Pan and W. Zhou (2010). Circular law, extreme singular values and potential theory. *Journal of Multivariate Analysis* **101**, 645–656. MR2575411

[22] M. Rudelson and R. Vershynin (2008). The Littlewood-Offord problem and invertibility of random matrices. *Advances in Mathematics* **218**, 600–633. MR2407948

[23] A. Sárközy and E. Szemerédi (1965). Über ein Problem von Erdős und Moser. *Acta Arithmetica* **11**, 205–208. MR0182619

[24] M. Talagrand (1996). A new look at independence. *Annals of Probability* **24**(1), 1–34. MR1387624

[25] T. Tao (2013). Outliers in the spectrum of i.i.d. matrices with bounded rank perturbations. *Probability Theory and Related Fields* **155**, 231–263. MR3010398

[26] T. Tao and V. Vu (2006). On random ±1 matrices: Singularity and determinant. *Random Structures Algorithms* **28**, 1–23. MR2187480

[27] T. Tao and V. Vu (2007). On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.* **20**, 603–628. MR2291914

[28] T. Tao and V. Vu (2008). Random matrices: The circular law. *Communications in Contemporary Mathematics* **10**, 261–307. MR2409368

[29] T. Tao and V. Vu (2009). Inverse Littlewood-Offord theorems and the condition number of random matrices. *Annals of Mathematics (2)* **169**(2), 595–632. MR2480613

[30] T. Tao and V. Vu (2010). A sharp inverse Littlewood-Offord theorem. *Random Structures and Algorithms* **37**(4), 525–539. MR2760363

[31] T. Tao and V. Vu (2009). From the Littlewood-Offord problem to the circular law: Universality of the spectral distribution of random matrices. *Bull. Amer. Math. Soc. (N.S.)* **46**(3), 377–396. MR2507275

[32] T. Tao, V. Vu and appendix by M. Krishnapur (2010). Random matrices: universality of ESDs and the circular law. *Annals of Probability* **38**(5), 2023–2065. MR2722794

[33] T. Tao and V. Vu (2010). Smooth analysis of the condition number and the least singular value. *Mathematics of Computation* **79**, 2333–2352. MR2684367

[34] R. Vershynin. Invertibility of symmetric random matrices, to appear in *Random Structures and Algorithms*. arxiv.org/abs/1102.0300.

[35] V. Vu. *Discrete random matrices*. arxiv.org/abs/math/0611321.

Hoi H. Nguyen
Department of Mathematics
Yale University
New Haven, CT 06520
USA
E-mail address: hoi.nguyen@yale.edu

Van H. Vu
Department of Mathematics
Yale University
New Haven, CT 06520
USA
E-mail address: van.vu@yale.edu

Received April 13, 2013