RANDOM DOUBLY STOCHASTIC MATRICES: THE CIRCULAR LAW

HOI H. NGUYEN

Abstract. Let X be a matrix sampled uniformly from the set of doubly stochastic matrices of size $n\times n$. We show that the empirical spectral distribution of the normalized matrix $\sqrt{n}(X-\mathbf{E}X)$ converges almost surely to the circular law. This confirms a conjecture of Chatterjee, Diaconis and Sly.

1. Introduction

Let M be a matrix of size $n\times n$ and let $\lambda_1,\dots,\lambda_n$ be the eigenvalues of M. The empirical spectral distribution (ESD) μ_M of M is defined as

$$\mu_M := \frac{1}{n}\sum_{i\le n}\delta_{\lambda_i}.$$

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

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

Resolving a long standing conjecture in random matrix theory, Tao and Vu (appendix by Krishnapur) have proved that the ESD of random i.i.d. matrices obeys the circular law.

Theorem 1.1. [34] *Assume that the entries of M are i.i.d. copies of a complex random variable of mean zero and variance one, then the ESD of the matrix $\frac{1}{\sqrt{n}}M$ converges almost surely to the circular measure $\mu_{\mathbf{cir}}$.*

This result is built on earlier developments by Girko [14, 15], Bai [1], Götze-Tikhomirov [16], Pan-Zhou [26] and by 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 [6] who proved the following.

Theorem 1.2. [6, 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*

2000 *Mathematics Subject Classification.* 11P70, 15A52, 60G50.

The author is partly supported by research grant DMS-1200898.

$\sqrt{n}\bar{X}$, 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 [7]).

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. [30, 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 $\mu_{\mathbf{cir}}$.*

With a slightly different assumption of dependence, Vu and the current author showed in [25] the following.

Theorem 1.4. [25, Theorem 1.2] *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 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, the matrix model in Theorem 1.4 is a discrete version of the random Markov matrices considered in Theorem 1.2 where the entries are now restricted to $\pm 1/s$. However, it is probably more suitable to compare this 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 Theorem 1.4 serves as a local version of the law.

Although the entries of the matrices above are mildly correlated, the rows are still independent. This allows sufficient room so that we can adapt the existing approaches to bear with the problems. Our focus in this note is on a matrix model whose rows and columns are not independent.

Theorem 1.5 (Circular law for random doubly stochastic matrices)**.** *Let X be a matrix chosen uniformly from the set of doubly stochastic matrices. Then the ESD of the normalized matrix $\sqrt{n}(X - \mathbf{E}X)$ converges almost surely to $\mu_{\mathbf{cir}}$.*

Little is known about the properties of random doubly stochastic matrices as it falls outside the scope of techniques from the usual random matrix theory. However, there have been recent breakthrough by Barvinok and Hartigan (see for instance [3, 4, 5]). The Birkhoff polytope $\mathcal{M}_n$, which is the set of doubly stochastic matrices of size $n \times n$, is the basic object in operation research because of its appearance as the feasible set for the assignment problem. Doubly stochastic matrices also serve as a natural model for priors in statistical analysis of Markov chains. There is a close connection between the Birkhoff polytope and $MS(n,c)$, the set of matrices of size $n \times n$ with non-negative integer entries and all column sums and row sums equal c. These matrices are called magic squares, which are well known in enumerative combinatorics. We refer the reader to the work of Chatterjee, Diaconis and Sly [8] for further discussion.

There is a strong belief that random doubly stochastic matrices behave like i.i.d. random matrices. This intuition has been verified in [8] in many ways. Among other things, it has been shown that the normalized entry nx_{11} converges in total variation to an exponential random variable of mean one. More general, the authors of [8] showed that the normalized projection nX_k, where X_k is the submatrix generated by the first k rows and columns of X and where $k = O(\frac{\sqrt{n}}{\log n})$, converges in total variation to the matrix of independent exponential random variables.

Regarding spectral distribution of X, it has been shown by Chatterjee, Diaconis and Sly that the empirical distribution of the singular values of $\sqrt{n}(X - \mathbf{E}X)$ obey the quarter-circular law.

Theorem 1.6. [8, Theorem 3] *Let $0 \le \sigma_1, \dots, \sigma_n$ be the singular values of $\sqrt{n}(X - \mathbf{E}X)$, where X is a random doubly stochastic matrix. Then the empirical spectral measure $\frac{1}{n}\sum_{i\le n} \delta_{\sigma_i}$ converges in probability and in weak topology to the quarter-circle measure $\frac{1}{\pi}\sqrt{4-x^2}\mathbf{1}_{[0,2]}dx$.*

The key ingredients in the proof of Theorem 1.6 are a sharp concentration result coupled with two transference principles (Lemmas 2.2 and 2.3 below). These principles help translate results from i.i.d random matrices of independent random exponential variables to random doubly stochastic matrices.

It has been conjectured in [8] that the empirical spectral distribution of $\sqrt{n}(X - \mathbf{E}X)$ obeys the circular law, which we confirm now. For the rest of this section we sketch the general plan to attack Theorem 1.5.

For the entries of X are exchangeable, $\mathbf{E}X$ is the matrix J_n of all $1/n$. The matrix $X - \mathbf{E}X$ has a zero eigenvalue and we want to single this outlier out due to several technical reasons. One way to do this is passing to $\bar{X}$, a matrix of size $(n-1)\times(n-1)$ defined as

$$\bar{X} := \begin{pmatrix} x_{22}-x_{21} & \cdots & x_{2n}-x_{21} \\ x_{32}-x_{31} & \cdots & x_{3n}-x_{31} \\ \vdots & \vdots & \vdots \\ x_{n2}-x_{n1} & \cdots & x_{nn}-x_{n1} \end{pmatrix}.$$

It is not hard to show that the spectra of $\sqrt{n}(X - \mathbf{E}X)$ is the union of zero and the spectra of $\sqrt{n}\bar{X}$. Indeed, consider the matrix $\lambda I_n - \sqrt{n}(X - \mathbf{E}X)$. By adding all other rows to its first row, and then subtracting the first column from every other column, we arrive at a matrix whose determinant is $\lambda \det(\lambda I_{n-1} - \sqrt{n}\bar{X})$, thus confirming our observation. Hence, it is enough to prove the circular law for $\bar{X}$.

Theorem 1.7 (Main theorem)**.** *Let X be a matrix chosen uniformly from the set of doubly stochastic matrices. Then the ESD of the matrix $\sqrt{n}\bar{X}$ converges almost surely to $\mu_{\mathbf{cir}}$.*

One way to prove our main result above is to showing that the Stieltjes transform of $\mu_{\sqrt{n}\bar{X}}$ converges to that of the circular measure. However, it is slightly more convenient to work with the logarithmic potential. We will mainly rely on the following machinery from [34, Theorem 2.1].

Lemma 1.8. *Suppose that $M = (m_{ij})_{1\le i,j\le n}$ is a random matrix. Assume that*

- $\frac{1}{n}\|M\|_{HS}^2 = \frac{1}{n}\sum_{i,j} m_{ij}^2$ *is bounded almost surely;*
- *for almost all complex numbers z_0, the logarithmic potential $\frac{1}{n}\log|\det(M - z_0 I_n)|$ converges almost surely to $f(z_0) = \int_{\mathbf{C}} \log|w - z_0| d\mu_{\mathbf{cir}}(w)$.*

Then μ_M converges almost surely to $\mu_{\mathbf{cir}}$.

We will break the main task into two parts, one showing the boundedness and one proving the convergence.

Theorem 1.9. *Let X be a matrix chosen uniformly from the set of doubly stochastic matrices. Then the square sum $\sum_{2\le i,j\le n}(x_{ij} - x_{i1})^2$ is bounded almost surely.*

The proof of Theorem 1.9 will be presented at the end of Section 2. The heart of our paper is to establish the convergence of $\frac{1}{n}\log|\det(\sqrt{n}\bar{X} - z_0 I_{n-1})|$.

Theorem 1.10. *For almost all complex numbers z_0, $\frac{1}{n}\log|\det(\sqrt{n}\bar{X} - z_0 I_{n-1})|$ converges almost surely to $f(z_0)$.*

The main difficulty in establishing Theorem 1.10 is that the entries in each row and each column of $\bar{X}$ are not at all independent. To our best knowledge, the convergence for such model has not been studied before in the literature. We will present its proof in Section 6.

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 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 $\mathbf{x}$.

For a real or complex vector $\mathbf{v} = (v_1, \dots, v_n)$, we will use the shorthand $\|\mathbf{v}\|$ for its L_2-norm $\big(\sum_i |v_i|^2\big)^{1/2}$.

2. Some properties of random doubly stochastic matrices

We will gather here some basic properties of random doubly stochastic matrices. The reader is invited to consult [8] for further insight and applications.

2.1. **Relation to random i.i.d matrix of exponentials.**

Let $\mathcal{M}_n$ be the Birkhoff polytope generated by the permutation matrices. Let Φ be the projection from $\mathbf{R}^{n^2}$ to $\mathbf{R}^{(n-1)^2}$ by mapping $(x_{ij})_{1\le i,j\le n}$ to $(x_{ij})_{2\le i,j\le n}$.

Let $\Gamma : \mathbf{R}^{(n-1)^2} \to \mathbf{R}^{n^2}$ denote the following function

$$\Gamma(X)=\Gamma(X)_{ij}:=\begin{cases} x_{ij} & 2\le i,j\le n;\\ 1-\sum_{k=2}^n x_{ik} & 2\le i\le n, j=1;\\ 1-\sum_{k=2}^n x_{kj} & 2\le j\le n, i=1;\\ 1-\sum_{l=2}^n(1-\sum_{k=2}^n x_{kl}) & i=j=1.\end{cases}$$

Thus Γ extends a matrix X of size $(n-1)\times(n-1)$ to a doubly stochastic matrix of size $n\times n$ whose bottom right corner is X. With the above notation, the doubly stochastic matrices correspond to $(n-1)\times(n-1)$-matrices of the set

$$S_n:=\left\{X=(x_{ij})_{2\le i,j\le n}\in[0,1]^{(n-1)^2}:0\le\Gamma(X)_{ij}\le 1\right\}.$$

The distribution of X as a random doubly stochastic matrix is then given by the uniform distribution on S_n. We next introduce an asymptotic formula by Canfield and Mckay [11] for the volume of S_n

$$\mathrm{Vol}(S_n)=\frac{1}{n^{n-1}}\frac{1}{(2\pi)^{n-1/2}n^{(n-1)^2}}\exp(\frac{1}{3}+n^2+o(1)). \tag{1}$$

This formula plays a crucial role in the transference principles to be introduced next.

Define

$$D_n:=\left\{Y=(y_{ij})_{1\le i,j\le n}:\Phi(\frac{1}{n}Y)\in S_n,\min\left\{\frac{1}{n}y_{ij}-\Gamma(\Phi(\frac{1}{n}Y))_{ij}\right\}\ge 0\right\},$$

where $\Phi:\mathbf{R}^{n^2}\to\mathbf{R}^{(n-1)^2}$ is the projection $X=(x_{ij})_{1\le i,j\le n}\mapsto(x_{ij})_{2\le i,j\le n}$.

Let $Y=(y_{ij})_{1\le i,j\le n}$ be a random matrix where y_{ij} are i.i.d. copies of a random exponential variable with mean one. As an application of (1), it is not hard to deduce the following transference principle between random doubly stochastic matrices X and random i.i.d matrices Y.

Lemma 2.2. [8, Lemma 2.1] *Condition on $Y\in D_n$, we have $(\frac{1}{n}y_{ij})_{2\le i,j\le n}$ is uniform on S_n. Furthermore, for large n we have*

$$\mathbf{P}(Y\in D_n)\ge n^{-4n}.$$

Lemma 2.2 is useful when we want to pass an extremely rare event from the model $\frac{1}{n}Y$ to the model X. In applications (in particular when working with concentration results), it is more useful to work with matrices of bounded entries. With this goal in mind we define

$$\tilde{S}_n:=\left\{\tilde{X}=(\tilde{x}_{ij})_{2\le i,j\le n}\in[0,1]^{(n-1)^2},0\le\Gamma(\tilde{X})_{ij}\le\frac{10\log n}{n}\right\},$$

and

$$\tilde{D}_n := \left\{ \tilde{Y} = (\tilde{y}_{ij})_{1\le i,j\le n} \in [0, 10\log n]^{n^2}, \frac{1}{n}\tilde{Y} \in \tilde{S}_n, 0 \le \frac{1}{n}\tilde{y}_{ij} - \Gamma(\Phi(\frac{1}{n}\tilde{Y}))_{ij} \le n^{-4} \right\}.$$

Observe that $\tilde{S_n}$ corresponds to doubly stochastic matrices $\tilde{X}$ of entries bounded by $10\log n/n$.

Let $\tilde{Y} = (\tilde{y}_{ij})_{1\le i,j\le n}$ where $\tilde{y}_{ij}$ are i.i.d. copies of a truncated exponetial $\tilde{y}$ of the following density function

$$\rho_{\tilde{y}}(x) = \begin{cases} \exp(-x)/(1-n^{-10}) & \text{if } x \in [0, 10\log n], \\ \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It is clear that $\mathbf{E}(\tilde{y}^2) = \Theta(1)$ and $\mathbf{E}(\tilde{y}^4) = \Theta(1)$. We now introduce another transference principle which is an analogue of Lemma 2.2.

Lemma 2.3. [8, Lemma 4.1] *Condition on $\tilde{Y} \in \tilde{D}_n$, we have that $(\frac{1}{n}\tilde{y}_{ij})_{2\le i,j\le n}$ is uniform on $\tilde{S}_n$. Furthermore, for large n we have*

$$\mathbf{P}(\tilde{Y} \in \tilde{D}_n) \ge n^{-10n}.$$

Notice that in the corresponding definition of $\tilde{D}_n$ in [8, Section 4] the bound $10\log n$ was replaced by $6\log n$, but one can easily check that this modification does not affect the validity of Lemma 2.3.

2.4. **Relation to random stochastic matrices.** Let $\mathcal{R} = \mathcal{R}_{r,n}$ denote the $r(n-1)$-dimensional polytope of nonnegative matrices of size $r \times n$ whose rows sum to 1. let μ_r denote the uniform probability measure on $\mathcal{R}$ and let ν_r denote the measure on $\mathcal{R}$ induced by the first r rows of a random doubly stochastic matrix X. As another application of (1) (to be more precise, we need a more general form for volume of polytopes generated by rectangular matrices of constant row and column sums), one can show that these two measures are comparable as long as r is small.

Lemma 2.5. [8, Lemma 3.3] *For a fixed integer $r \ge 1$ and $n > r$ the Radon-Nikodym derivative of the measures μ_r and ν_r satisfies*

$$\frac{d\nu_r}{d\mu_r} \le (1 + o(1))\exp(r/2)$$

as $n \to \infty$.

It then follows that, in terms of order, there is not much difference between the models X and $\tilde{X}$.

Theorem 2.6. *Assume that $B > 4$ is a constant, then*

$$\mathbf{P}_X(n^{-B} \le nx_{11} \le B\log n) \ge 1 - O(n^{-B/2}).$$

In particular, since the entries of X are exchangeable, Theorem 2.6 yields the following.

Corollary 2.7. *Assume that X is a random doubly stochastic matrix, then*

$$\mathbf{P}(X \in \tilde{S}_n) = \mathbf{P}(|x_{ij}| \le 10\log n/n \text{ for all } 1 \le i,j \le n) \ge 1 - O(n^{-3}).$$

Proof. (of Theorem 2.6) It follows from Lemma 2.5 (for $r = 1$) that

$$\mathbf{P}(n^{-B} \le nx_{11} \le B\log n) \le \big(1 + o(1)\big)\exp(1/2)\mathbf{P}(n^{-B} \le nx_1 \le B\log n),$$

where x_1 has distribution $B(1, n-1)$.

The claim then follows because

$$\begin{aligned}\mathbf{P}(n^{-B} \le nx_1 \le B\log n) &= (n-1)\int_{n^{-B}}^{B\log n}(1-x)^{n-2}dx \\ &= 1 - (n-1)\Big(\int_0^{n^{-B}}(1-x)^{n-2}dx + \int_{B\log n}^1 (1-x)^{n-2}dx\Big) \\ &\ge 1 - O(n^{-B/2}).\end{aligned}$$

□

We end this section by giving a proof for the boundedness of Lemma 1.8.

2.8. **A proof for Theorem 1.9.** We first focus on the random vector $\mathbf{x} = (x_1, \dots, x_n)$ chosen uniformly from the simplex $S = \Big\{\mathbf{x} = (x_1, \dots, x_n), 0 \le x_i \le 1, \sum_i x_i = 1\Big\}$. Because each x_i has distribution $B(1, n-1)$, we have

$$\mathbf{E}_{\mathbf{x}}\|\mathbf{x}\|^2 = \frac{2}{n+1}. \tag{3}$$

Also, it can be shown that (for instance from [22, equation (19)])

$$\mathbf{E}_{\mathbf{x}}x_1x_2 = \frac{1}{n(n+1)}. \tag{4}$$

It thus follows from (3) that $\|\mathbf{x}\| = O(1/\sqrt{n})$ with high probability. It turns out that this probability is extremely close to one.

Lemma 2.9. *Assume that $\mathbf{x}$ is sampled uniformly from S and assume that $\epsilon > 0$ is a sufficiently small constant. Then there exists a positive constant $C > 0$ such that*

$$\mathbf{P}(\|\mathbf{x}\| \geq C/\sqrt{n}) \leq \exp(-\epsilon\sqrt{n}).$$

We assume Lemma 2.9 for the moment.

Proof. (of Theorem 1.9) First, it follows from Lemma 2.5 (for $r = 1$) that

$$\begin{aligned}\mathbf{P}(x_{21}^2 + \cdots + x_{n1}^2 \geq C/n) &\leq \big(1 + o(1)\big)\exp(1/2)\mathbf{P}(x_2^2 + \cdots + x_n^2 \geq C/n)\\ &= O(1)\mathbf{P}(x_1^2 + x_2^2 + \cdots + x_n^2 \geq C/n)\end{aligned}$$

where $(x_1, x_2, \dots, x_n)$ are sampled uniformly from the simplex S. But Lemma 2.9 indicates that the RHS is bounded by $\exp(-\epsilon\sqrt{n})$. Thus

$$\mathbf{P}(x_{21}^2 + \cdots + x_{n1}^2 \geq C/n) = O(\exp(-\epsilon\sqrt{n})). \tag{5}$$

And so, as x_{ij} are exchangeable, for any j we also have

$$\mathbf{P}(x_{2j}^2 + \cdots + x_{nj}^2 \geq C/n) = O(\exp(-\epsilon\sqrt{n})). \tag{6}$$

The claim of Theorem 1.9 then follows because $\sum_{2\leq i,j\leq n}(x_{ij} - x_{i1})^2 \geq C$ would imply $\sum_{i=2}^n x_{ij}^2 \geq C/4n$ for some j.

□

It remains to prove for Lemma 2.9. We apply the following concentration result by Paouris.

Theorem 2.10. [27, Theorem 1.1] *There exists an absolute constant $c > 0$ such that if K is an isotropic convex body in $\mathbf{R}^n$, then*

$$\mathbf{P}(\mathbf{x} \in K, \|\mathbf{x}\| \geq c\sqrt{n}L_K t) \leq \exp(-\sqrt{n}t)$$

for every $t \geq 1$, where L_K is the isotropic constant of K.

Observe that, by the triangle inequality, for Lemma 2.9 it is enough to give a similar probability bound for the event $\|\mathbf{x} - (1/n, \dots, 1/n)\| \geq C/\sqrt{n}$.

We first shift S to the hyperplane $H := \{\mathbf{x}' = (x'_1, \dots, x'_n), x'_1 + \cdots + x'_n = 0\}$ by the translation $\mathbf{x} = (x_1, \dots, x_n) \mapsto (x_1 - 1/n, \dots, x_n - 1/n)$. We then scale the obtained body by a factor $\alpha = \Theta(n)$ to obtain a regular simplex S' of volume one. Elementary computations show that this is an isotropic body of bounded isotropic constant. Indeed, if $\mathbf{x}' = (x'_1, \dots, x'_n)$ is sampled uniformly from S' and if $\mathbf{\Theta} = (\theta_1, \dots, \theta_n)$ is any unit vector in H, then by (3) and (4)

$$\begin{aligned}
\mathbf{E}_{\mathbf{x}' \in S'} \langle \mathbf{x}', \mathbf{\Theta} \rangle^2 &= \mathbf{E}_{\mathbf{x}' \in S'} (\sum_i \theta_i x'_i)^2 \\
&= \mathbf{E}_{\mathbf{x} \in S} \sum_i \alpha^2 (\sum_i \theta_i (x_i - \frac{1}{n}))^2 \\
&= \alpha^2 \sum_i \theta_i^2 (x_i - \frac{1}{n})^2 + 2\alpha^2 \sum_{i \neq j} \theta_i \theta_j (x_i - \frac{1}{n})(x_j - \frac{1}{n}) \\
&= \alpha^2 (\frac{2}{n(n+1)} - \frac{1}{n^2}) \sum_i \theta_i^2 + 2\alpha^2 (\frac{1}{n(n+1)} - \frac{1}{n^2}) \theta_i \theta_j \\
&= \alpha^2 (\frac{1}{n(n+1)}) \sum_i \theta_i^2 + \alpha^2 (\frac{1}{n(n+1)} - \frac{1}{n^2}) (\sum_i \theta_i)^2 \\
&= \frac{\alpha^2}{n(n+1)}.
\end{aligned}$$

Thus the isotropic constant of S' is of constant order. Theorem 2.10 applied to $\mathbf{x}'$ yields the following for sufficiently large constant C

$$\mathbf{P}(\mathbf{x}' \in S', \|\mathbf{x}'\| \geq C\sqrt{n}) \leq \exp(-\epsilon\sqrt{n}).$$

Lemma 2.9 then follows because $\alpha\|\mathbf{x} - (1/n, \dots, 1/n)\| = \|\mathbf{x}'\|$.

3. The singularity of $\bar{X}$

In order to justify Theorem 1.10, one of the key steps is to bound the singularity probability of the matrix $\sqrt{n}\bar{X} - z_0 I_{n-1}$. This problem is of interest of its own.

We will show the following general result regarding the least singular value σ_{n-1}.

Theorem 3.1. *Let $F = (f_{ij})_{2 \leq i,j \leq n}$ be a deterministic matrix where $|f_{ij}| \leq n^\gamma$ with some positive constant γ. Let X be an $n \times n$ matrix chosen uniformly from the set of doubly stochastic matrices. Then for any positive constant B there exists a positive constant A such that*

$$\mathbf{P}(\sigma_{n-1}(\bar{X} + F) \leq n^{-A}) \leq n^{-B}.$$

Combine with Theorem 2.7 we obtain the following important corollary which will be reserved for later applications.

Corollary 3.2. *Let $F=(f_{ij})_{2\le i,j\le n}$ be a deterministic matrix where $|f_{ij}|\le n^{\gamma}$ with some positive constant γ. Let $\tilde{X}=(x_{ij})$ be a random doubly stochastic matrix where $x_{ij}\le 10\log n/n$ for all $1\le i,j\le n$. Then there exists a positive constant A such that*

$$\mathbf{P}(\sigma_{n-1}(\bar{\tilde{X}}+F)\le n^{-A})=O(n^{-3}).$$

Here $\bar{\tilde{X}}$ is obtained from $\tilde{X}$ in the same way as how $\bar{X}$ was defined from X.

We remark that a similar version of Theorem 3.1 had appeared in [34] to deal with random matrices of i.i.d. entries (see also [6, 25] and the references therein). However, our task here looks much harder as the entries in each row and each column are not independent. We will now sketch the proof of Theorem 3.1, more details will be presented in Section 4.

Assume that $\sigma_{n-1}(\bar{X}+F)\le n^{-A}$. Then, by letting $C=(c_{ij})_{2\le i,j\le n}$ be the cofactor matrix of $\bar{X}+F$, there exist vectors $\mathbf{x}$ and $\mathbf{y}$ such that $\|\mathbf{x}\|=1$ and $\|\mathbf{y}\|\le n^{-A}$ and

$$C\mathbf{y}=\det(\bar{X}+F)\mathbf{x}.$$

So

$$\|C\mathbf{y}\|=|\det(\bar{X}+F)|.$$

Thus by Cauchy-Schwarz inequality, with a loss of a factor of n in probability and without loss of generality we can assume that

$$\sum_{j=2}^{n}|c_{2j}|^2\ge n^{2A-1}|\det(\bar{X}+F)|^2. \tag{7}$$

In what follows we fix the matrix $X_{(n-2)\times(n-1)}$ generated by the last $(n-2)$ rows and the last $(n-1)$ columns of X (equivalently, we fix the last $(n-2)$ rows of $\bar{X}$).

Let $s_2,\dots,s_n$ be the column sums of $X_{(n-2)\times(n-1)}$. By Theorem 2.6, the probability that all $x_{11},\dots,x_{1n},x_{21},\dots,x_{2n}$ are greater than n^{-2B-2} is bounded from below by $1-O(n^{-B})$, in which case we have

$$s_i\le 1-n^{-2B-2} \text{ for all } i\ge 2, \text{ and } 0\le s_1:=(n-2)-(s_2+\dots+s_n)\le 1-n^{-2B-2}. \tag{8}$$

Thus it is enough to justify Theorem 3.1 conditioning on this event.

Next, given a sequence $s_2,\dots,s_n$ satisfying (8), we will choose $x_2:=x_{22},\dots,x_n:=x_{2n}$ uniformly and respectively from the interval $[0,1-s_2],\dots,[0,1-s_n]$ such that

$$s_1 \le x_2 + \cdots + x_n \le 1. \tag{9}$$

The upper bound guarantees that $x_1 := x_{21} = 1 - (x_2 + \cdots + x_n) \ge 0$, while the lower bound ensures that $x_{11} = 1 - s_1 - x_{21} = x_2 + \cdots + x_n - s_1 \ge 0$.

We now express $\det(\bar{X}+F)$ as a linear form of its first row $(x_2 - x_1 + f_{22}, \dots, x_n - x_1 + f_{2n})$,

$$\det(\bar{X} + F) = \sum_{2\le j\le n} c_{2j}(\bar{X}+F)(x_j - x_1 + f_{2j}).$$

By using the fact that $x_1 = 1 - \sum_{2\le j\le n} x_j$ we can rewrite the above as

$$\det(\bar{X} + F) = \sum_{2\le j\le n} (c_{2j} + \sum_{2\le i\le n} c_{2i}) x_j + c, \tag{10}$$

where c is a constant depending on c_{2j}'s and f_{2j}'s.

Observe that

$$\sum_{2\le j\le n} |c_{2j} + \sum_{2\le i\le n} c_{2i}|^2 = \sum_{2\le j\le n} |c_{2j}|^2 + (n+1)|\sum_{2\le j\le n} c_{2j}|^2 \ge \sum_{2\le j\le n} |c_{2j}|^2.$$

Thus, by increasing A if needed, we obtain from (7) and (10) the following

$$|\sum_{2\le j\le n} x_j a_j + c| \le n^{-A},$$

where

$$a_j := \frac{c_{2j} + \sum_{2\le i\le n} c_{2i}}{(\sum_{2\le j\le n} |c_{2j} + \sum_{2\le i\le n} c_{2i}|^2)^{1/2}}. \tag{11}$$

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

- *Inverse step.* Given the matrix $X_{(n-2)\times(n-1)}$ for which all the column sums s_i satisfy (8), assume that

  $$\mathbf{P}_{x_2,\dots,x_n}\Big(|\sum_{2\le j\le n} a_j x_j + c)| \le n^{-A}\Big) \ge n^{-B},$$

  where the probability is taken over all $x_i, 2 \le i$ which satisfy (9). Then there is a strong structure among the cofactors c_{2j} of $X_{(n-2)\times(n-1)}$.

- *Counting step.* With respect to $X_{(n-2)\times(n-1)}$, the probability that there is a strong structure among the cofactors c_{2j} 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

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

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 $B := \{(k_1, \dots, k_r) \in \mathbf{Z}^r | K_i \le k_i \le K_i'\}$ 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 $i \ge 1$ and $g_0 = 0$, we say that Q is *symmetric*.

We are now ready to state our steps in details.

Theorem 3.3 (Inverse Step)**.** *Let $0 < \epsilon < 1$ and $B > 0$ be given constants. Assume that*

$$\mathbf{P}_{x_2,\dots,x_n}\Big(|\sum_{2\le j\le n} a_j x_j + c)| \le n^{-A}\Big) \ge n^{-B}.$$

for some sufficiently large integer A, where a_j are defined in (11)*, and x_j are chosen uniformly from the intervals $[0, 1 - s_i]$ such that the constraint* (9) *holds. Then there exists a vector $\mathbf{u} = (u_2, \dots, u_n) \in \mathbf{C}^{n-1}$ which satisfies the following properties.*

- $1/2 \le \|\mathbf{u}\| \le 2$ *and* $|\langle \mathbf{u}, \mathbf{r}_i(\bar{X} + F)\rangle| \le n^{-A+\gamma+2}$ *for all but the first row* $\mathbf{r}_1(\bar{X} + F)$ *of* $\bar{X} + F$.
- *All but n' components u_i belong to a GAP Q (not necessarily symmetric) of rank $r = O_{B,\epsilon}(1)$, and of cardinality $|Q| = n^{O_{B,\epsilon}(1)}$.*
- *All the real and imaginary parts of u_i and of the generators of Q are rational numbers of the form p/q, where $|p|, |q| \le n^{2A+3/2}$.*

In the second step of the approach we show that the probability for $X_{(n-2)\times(n-1)}$ having the above properties is negligible.

Theorem 3.4 (Counting Step)**.** *With respect to $X_{(n-2)\times(n-1)}$, or equivalently, with respect to the last $(n-2)$ rows of $\bar{X}$, the probability that there exists a vector $\mathbf{u}$ as in Theorem 3.3 is* $\exp(-\Theta(n))$.

Proof. (of Theorem 3.4) Firstly, we show that the number of structural vectors $\mathbf{u}$ described in Theorem 3.3 is bounded by $n^{O_{B,\epsilon}(n)+O_A(n^\epsilon)}$. Indeed, because each GAP is determined by its generators and its dimensions, and because all the real and complex parts of the genrators are of the form p/q where $|p|, |q| \le n^{2A+3/2}$, there are $n^{O_{A,B,\epsilon}(1)}$ GAPs which have rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$. Next, for each determined GAP Q of size $n^{O_{B,\epsilon}(1)}$, there are $|Q|^n = n^{O_{B,\epsilon}(n)}$ ways to choose the u_i as its elements. For the remaining $O(n^\epsilon)$ exceptional u_i that may not belong to Q, there are $n^{O_A(n^\epsilon)}$ ways to choose them as numbers of the form p/q where $|p|,|q| \le n^{2A+3/2}$. Putting these together we obtain the bound as claimed.

Secondly, as for each fixed structural vector $\mathbf{u}$ from Theorem 3.3 we have $|\langle \mathbf{u}, \mathbf{r}_i(\bar{X}+F)\rangle| = O(n^{-A+\gamma+2})$ for all $2 \le i \le n-1$. So

$$\Big|\sum_{2\le j} u_j(x_{ij} - x_{i1} + f_{ij})\Big| = \Big|\sum_{2\le j} x_{ij}(u_j + \sum_{2\le k} u_k) - \sum_{2\le j} u_j + \sum_{2\le j} u_j f_{ij}\Big| = O(n^{-A+\gamma+2}). \quad (12)$$

We next view this inequality as for the matrix model Y and $\bar{Y}$, where Y was introduced in Section 2 and $\bar{Y}$ is obtained from Y in the same way as how $\bar{X}$ was defined from X,

$$|\sum_{2\le j} \frac{1}{n} y_{ij}(u_j + \sum_{2\le k} u_k) - \sum_{2\le j} u_j + \sum_{2\le j} u_j f_{ij}| = O(n^{-A+\gamma+2}). \quad (13)$$

Observe that

$$\sum_{2\le j\le k} |u_j + \sum_{2\le k\le n} u_k|^2 \ge \sum_{2\le k\le n} u_k^2 \ge 1/4.$$

Thus there exits j_0 such that

$$|u_{j_0} + \sum_{2\le k\le n} u_k| \ge 1/2\sqrt{n}.$$

It then follows that for each i, with room to spare

$$\begin{aligned}
&\mathbf{P}\Big(|\sum_{2\le j} \frac{1}{n} y_{ij}(u_j + \sum_{2\le k} u_k) - \sum_{2\le j} u_j + \sum_{2\le j} u_j f_{ij}| = O(n^{-A+\gamma+2})\Big)\\
&= \mathbf{P}_{y_{ij}, j\ne j_0}\mathbf{P}_{y_{ij_0}}\Big(|\frac{1}{n} y_{ij_0}(u_{j_0} + \sum_{2\le k\le n} u_k) + \sum_{j\ne j_0} \frac{1}{n} y_{ij}(u_j + \sum_{2\le k\le n} u_k) - \dots| = O(n^{-A+\gamma+2}) | y_{ij, j\ne j_0}\Big)\\
&= O(n^{-A+\gamma+10}),
\end{aligned}$$

where in the last conditional probability estimate we used the fact that y_{ij} are i.i.d exponentials of mean one.

Hence, for each fixed structural vector $\mathbf{u}$, the probability $\mathbf{P_u}$ that (13) holds for all rows $\mathbf{r}_i(\bar{Y}+F), 2 \le i \le n-1$, is bounded by

$$\mathbf{P}_{\mathbf{u}} \le n^{(-A+\gamma+10)(n-2)}.$$

Summing over structural vectors $\mathbf{u}$, we thus obtain the following upper bound for the probability that there exists a structural vector $\mathbf{u}$ for which (13) holds for all rows $\mathbf{r}_i(\bar{Y}+F), 2 \le i \le n-1$

$$\sum_{\mathbf{u}} \mathbf{P}_{\mathbf{u}} \le n^{O_{B,\epsilon}(n)+O_A(n^{\epsilon})} n^{(-A+\gamma+10)(n-2)} = O(n^{-An/2}),$$

provided that A is large enough.

To conclude the proof of Theorem 3.4, we use Lemma 2.2 to pass from Y and $\bar{Y}$ back to X and $\bar{X}$. The probability that there exists a structural vector $\mathbf{u}$ for which (12) holds for all rows $\mathbf{r}_i(\bar{X}+F), 2 \le i \le n-1$, is then bounded by $O(n^{-An/2+4n}) = O\big(\exp(-\Theta(n))\big)$, provided that A is sufficiently large. $\square$

4. PROOF OF THEOREM 3.3

We recall from the assumption of Theorem 3.3 that

$$\mathbf{P}_{x_2,\dots,x_n}\Big(|\sum_{j\ge 2} a_j x_j + c| \le n^{-A}\Big) \ge n^{-B}, \tag{14}$$

where $x_2,\dots,x_n$ are uniformly sampled from the interval $[0,1-s_2],\dots,[0,1-s_n]$ respectively so that (9) holds.

This is a large concentration of linear form of mildly dependent random variables. Our first goal is to relax these dependencies.

4.1. **A simple reduction step.** Let E_n be the set of all $(x_2,\dots,x_n)$ uniformly sampled from $[0,1-s_2]\times\dots\times[0,1-s_n]$ so that (9) holds. We recall from (8) that $s_i \le 1-n^{-2B-2}$.

Consider the event $s_1 \le x_2' + \dots + x_n' \le 1$, where x_i' are independently and uniformly sampled from the interval $[0,1-s_i]$ respectively.

Note that $\mathbf{E}(x_2'+\dots+x_n') = \sum_{2\le i\le n}(1-s_i)/2 = (1-s_1)/2$. Since the random variables $x_i'-(1-s_i)/2$ are symmetric and uniform, the density function $f(x)$ of $x_2'+\dots+x_n'$ is maximized at $(1-s_1)/2$ and decreases as $|x-(1-s_1)/2|$ increases. Thus we have

$$\begin{aligned}\mathbf{P}\big((x_2',\dots,x_n')\in E_n\big) &= \mathbf{P}(s_1 \le x_2'+\cdots+x_n' \le 1)\\ &= \int_{s_1}^1 f(x)dx = \frac{\int_{s_1}^1 f(x)dx}{\int_0^{(1-s_2)+\cdots+(1-s_n)} f(x)dx}\\ &\ge \frac{1-s_1}{(1-s_2)+\cdots+(1-s_n)} = \frac{1-s_1}{1+s_1}\\ &= \Omega(n^{-2B-2}),\end{aligned}$$

where we noted from (8) that $s_1 \le 1-n^{-2B-2}$.

Observe that if we condition on $s_n \le x_2'+\cdots+x_n' \le 1$, then the distribution of $(x_2',\dots,x_n')$ is uniform over the set E_n. It thus follows from (14) that

$$\mathbf{P}_{x_2',\dots,x_n'}\Big(|\sum_{j\ge 2} a_j x_j' + c| \le n^{-A}\Big) \ge n^{-3B-2}. \tag{15}$$

In the next step of the reduction, we divide the intervals $[0,1-s_i]$ into disjoint intervals $I_{i1},\dots,I_{ik_i}$ of length n^{-3B-2}, where $k_i=(1-s_i)/n^{-3B-2}$ (without loss of generality, we assume that k_i are integers). Next, to sample x_i' uniformly from the interval $[0,1-s_i]$ we first choose at random an interval from $\{I_{i1},\dots,I_{ik_i}\}$ and then sample x_i' from it. By this way, (15) implies that there exist intervals $I_{ij_i}, 2\le i\le n$, such that if x_i' are chosen uniformly from I_{ij_i} then

$$\mathbf{P}_{x_2',\dots,x_n'}\Big(|\sum_{j\ge 2} a_j x_j' + c| \le n^{-A}\Big) \ge n^{-3B-2}. \tag{16}$$

Observe furthermore that, by shifting c if needed, we can assume that $I_{ij_i}=[0,n^{-3B-2}]$ for all i. Finally, by passing to $x_i'':=n^{3B+2}x_i'$ and by decreasing A to $A-(3B+2)$, we can assume that all x_i' are uniformly sampled from the interval $[0,1]$.

4.2. **High concentration of linear form.** A classical result of Erdős [12] and Littlewood-Offord [21] asserts that if b_i are real numbers of magnitude $|b_i|\ge 1$, then the probability that the random sum $\sum_{i=1}^n b_i x_i$ concentrates on an interval of length one is of order $O(n^{-1/2})$, where x_i are i.i.d. copies of a Bernoulli random variable. This remarkable inequality has generated an impressive way of research, particularly from the early 1960s to the late 1980s. We refer the reader to [18, 20] and the references therein for these developments.

Motivated by inverse theorems from additive combinatorics, Tao and Vu studied the underlying reason as to why the concentration probability of $\sum_{i=1}^n b_i x_i$ on a short interval is large. A closer look at the definition of GAPs defined in the previous section reveals that if b_i are very *close* to the elements of a *GAP* of rank $O(1)$ and size $n^{O(1)}$, then the concentration probability of $\sum_{i=1}^n b_i x_i$ on a short interval is of order $n^{-O(1)}$, where x_i are i.i.d. copies of a Bernoulli random variable.

It has been shown by Tao and Vu [32, 34, 35] in an implicit way, and by the current author and Vu [24] in a more explicit way that these are essentially the only examples that have high concentration probability.

We say that a complex number a is δ-close to a set $Q \subset \mathbf{C}$ if there exists $q \in Q$ such that $|a-q| \leq \delta$.

Theorem 4.3 (Inverse Littlewood-Offord theorem for linear forms)**.** [24, Corollary 2.10] *Let $0 < \epsilon < 1$ and $C > 0$. Let $\beta > 0$ be an arbitrary real number that may depend on n. Suppose that $b_i = (b_{i1}, b_{i2})$ are complex numbers such that $\sum_{i=1}^n \|b_i\|^2 = 1$, and*

$$\sup_a \mathbf{P}_{\mathbf{x}}(|\sum_{i=1}^n b_i x_i - a| \leq \beta) = \rho \geq n^{-C},$$

where $\mathbf{x} = (x_1, \dots, x_n)$, and x_i are i.i.d. copies of random variable ξ satisfying $\mathbf{P}(c_1 \leq \xi - \xi' \leq c_2) \geq c_3$ for some positive constants c_1, c_2 and c_3 . Then, for any number n' between n^ϵ and n, there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r k_i g_i : k_i \in \mathbf{Z}, |k_i| \leq L_i\}$ such that

- *at least $n - n'$ numbers b_i are β-close to Q;*
- *Q has small rank, $r = O_{C,\epsilon}(1)$, and small cardinality*
$$|Q| \leq \max\left(O_{C,\epsilon}(\frac{\rho^{-1}}{\sqrt{n'}}), 1\right);$$
- *there exists a non-zero integer $p = O_{C,\epsilon}(\sqrt{n'})$ such that all generators $g_i = (g_{i1}, g_{i2})$ of Q have the form $g_{ij} = \beta\frac{p_{ij}}{p}$, with $p_{ij} \in \mathbf{Z}$ and $|p_{ij}| = O_{C,\epsilon}(\beta^{-1}\sqrt{n'})$.*

Theorem 4.3 was proved in [24] with $c_1 = 1, c_2 = 2$ and $c_3 = 1/2$, but the proof there automatically extends to any constants $0 < c_1 < c_2$ and $0 < c_3$.

The interested reader is invited to read also [23],[28],[39] for other variants and further developments of the inverse results.

We now prove Theorem 3.3. Theorem 4.3 applied to (16), with $n' = n^\epsilon, C = 3B + 2$ and x_i being independently and uniformly distributed over the interval $[0, 1]$, implies that there exists a vector $\mathbf{v} = (v_2, \dots, v_n)$ such that

- $|a_i - v_i| \leq n^{-A}$ for all indices i from $\{2, \dots, n\}$;
- all but n' numbers v_i belong to a GAP Q of small rank, $r = O_{B,\epsilon}(1)$, and of small cardinality $|Q| = O(n^{O_{B,\epsilon}(1)})$;
- all the real and imaginary parts of v_i and of the generators of Q are rational numbers of the form p/q, with $p, q \in \mathbf{Z}$ and $|p|, |q| = O_{B,\epsilon}(n^{A+1/2})$.

Recall that

$$a_j = \frac{c_{2j} + \sum_{2 \le i \le n} c_{2i}}{(\sum_{2 \le j \le n} |c_{2j} + \sum_{2 \le i \le n} c_{2i}|^2)^{1/2}}.$$

We will translate the above useful information on a_j's to c_j's. To do so we fist find a number of the form p/n^A, where $p \in \mathbf{Z}$ and $-n^A \le p \le n^A$ such that

$$\Big|\frac{p}{n^A} - \frac{\sum_{2 \le j \le n} c_{2j}}{(\sum_j |c_{2j} + \sum_{2 \le i \le n} c_{2i}|^2)^{1/2}}\Big| \le \frac{1}{n^A}.$$

Thus, by shifting the GAP Q by p/n^A, we obtain $|a_j' - v_j'| \le 2n^{-A}$, and so

$$\|\mathbf{a}' - \mathbf{v}'\| = O(n^{-A+1/2}),$$

where $\mathbf{a}' = (a_2', \dots, a_n'), \mathbf{v}' = (v_2', \dots, v_n')$ and

$$a_j' = \frac{c_{2j}}{(\sum_j |c_{2j} + \sum_{2 \le i \le n} c_{2i}|^2)^{1/2}}, \text{ as well as } v_j' = v_j - \frac{p}{n^A}.$$

By definition, $1/2n^2 \le \sum |a_j'|^2 \le 1$, so by the triangle inequality

$$\|\mathbf{v}'\| \ge \|\mathbf{a}'\| - O(n^{-A+1/2}) \ge 1/\sqrt{2}n - O(n^{-A+1/2})$$

and

$$\|\mathbf{v}'\| \le \|\mathbf{a}'\| + O(n^{-A+1/2}) \le 1 + O(n^{-A+1/2}).$$

More importantly, as $\mathbf{a}'$ is proportional to $(c_{22}, \dots, c_{2n})$ (which are the cofactors of $\bar{X} + F$), $\mathbf{a}'$ is orthogonal to all but the first row of $\bar{X} + F$. In other words, $|\langle \mathbf{a}', \mathbf{r}_i(\bar{X} + F)\rangle| = 0$ for all $i \ge 2$. It is thus implied that

$$|\langle \mathbf{v}', \mathbf{r}_i(\bar{X} + F)\rangle| \le n^{-A+\gamma+1}.$$

In the last step of the proof, we find nonzero numbers $p', q' \in \mathbf{Z}, |p'|, |q'| = O(n)$ so that $\|\mathbf{v}'\|/2 \le p'/q' \le 2\|\mathbf{v}'\|$.

Set

$$\mathbf{u} := \frac{q'}{p'}\mathbf{v}',$$

we then have

- $1/2 \le \|\mathbf{u}\| \le 2$ and $\langle \mathbf{u}, \mathbf{r}_i(\bar{X}+F)\rangle \le n^{-A+\gamma+2}$ for all but the first rows of $\bar{X}+F$;
- all but n' components u_i belong to a GAP Q' (not necessarily symmetric) of small rank, $r = O_{B,\epsilon}(1)$, and of small cardinality $|Q'| = O(n^{O_{B,\epsilon}(1)})$;
- all the real and imaginary parts of u_i and of the generators of Q' are rational numbers of the form p/q, with $p,q \in \mathbf{Z}$ and $|p|,|q| = O_{B,\epsilon}(n^{2A+3/2})$.

5. Spectral concentration of i.i.d. random covariance matrices

From now on we will mainly focus on the bounded model $\tilde{X}$ rather than on X. This is the model where we can relate to $\tilde{Y}$, a matrix of bounded i.i.d entries (defined in Section 2) for which concentration results may easily apply. Furthermore, by Corollary 2.7, there is not much difference between the two models X and $\tilde{X}$.

Having learned from Corollary 3.2 that $|\det(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1})|$ is bounded away from zero, we will show that $\frac{1}{n}\log|\det(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1})|$ is well concentrated around its mean. This result will then immediately imply Theorem 1.10.

In order to study the concentration of $\det(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1})$, we might first relate it to the counterpart $\bar{\tilde{Y}}$. However, the entries of the later model are not independent, and so certain well-known concentration results for i.i.d matrices are not applicable. To avoid this technical issue, we will modify $\sqrt{n}\bar{\tilde{X}}$ as follows. Observe that

$$\det(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1}) = \frac{1}{\sqrt{n}}\det(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}), \tag{17}$$

where F_{z_0} is the deterministic matrix obtained from $z_0 I_{n-1}$ by attaching $(-\sqrt{n},\dots,-\sqrt{n})$ and $(-\sqrt{n},0,\dots,0)^T$ as its first row and first column respectively, and $\tilde{X}_{(n-1)\times n}$ is the matrix obtained from $\tilde{X}$ by replacing its first row by a zero vector,

$$\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0} := \begin{pmatrix} \sqrt{n} & \sqrt{n} & \cdots & \sqrt{n} \\ \sqrt{n}\tilde{x}_{21} & \sqrt{n}\tilde{x}_{22} - z_0 & \cdots & \sqrt{n}\tilde{x}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{n}\tilde{x}_{n1} & \sqrt{n}\tilde{x}_{n2} & \cdots & \sqrt{n}\tilde{x}_{nn} - z_0 \end{pmatrix}.$$

As it turns out, it is more pleasant to work with $\tilde{X}_{(n-1)\times n}$ because the entries of its counterpart $\tilde{Y}_{(n-1)\times n}$ are now independent. To relate the singularity of $\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1}$ to that of $\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}$, we have a crucial observation below.

Claim 5.1. *Suppose that A is sufficiently large constant. We have*

$$\sigma_n(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}) \geq \frac{1}{n}\min\big(\frac{1}{\sqrt{2n}}\sigma_{n-1}(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1}) - O(n^{-A}), n^{-A}\big).$$

To prove this claim, let $\mathbf{c}_1, \dots, \mathbf{c}_n$ be the columns of $\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}$. Let $\mathbf{v} = (v_1, \dots, v_n)$ be any unit vector. If $|v_1 + \cdots + v_n| \geq n^{-A-1/2}$, then it is clear that $\|(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0})\mathbf{v}\| \geq |\sqrt{n}(v_1 + \cdots + v_n)| \geq n^{-A}$. Otherwise, as $|v_1|^2 + \cdots + |v_n|^2 = 1$, we can easily deduce that $|v_2|^2 + \cdots + |v_n|^2 \geq 1/2n$. Next, by the triangle inequality,

$$\begin{aligned}
\|(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0})v\| &= \|\sum_{2\leq i\leq n} v_i \mathbf{c}_i\| = \|\sum_{2\leq i\leq n} v_i(\mathbf{c}_i - \mathbf{c}_1) + (v_1 + \cdots + v_n)\mathbf{c}_1\| \\
&\geq \|\sum_{2\leq i\leq n} v_i \mathbf{c}_i\| - n^{-A-1/2}\|\mathbf{c}_1\| \\
&\geq (|v_2|^2 + \cdots + |v_n|^2)^{1/2}\sigma_{n-1}(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1}) - \sqrt{2}n^{-A} \\
&\geq \frac{1}{\sqrt{2n}}\sigma_{n-1}(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1}) - O(n^{-A}).
\end{aligned}$$

Claim 5.1 guarantees that the polynomial probability bound for $\sigma_{n-1}(\sqrt{n}\bar{\tilde{X}} - z_0 I_{n-1})$ from Corollary 3.2 continues to hold for $\sigma_n(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0})$ (with probably worse A).

Theorem 5.2. *There exists a positive constant A such that*

$$\mathbf{P}(\sigma_n(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}) \leq n^{-A}) = O(n^{-3}).$$

Our goal is then to establish a large concentration of $\frac{1}{n}\log|\det(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0})|$ around its mean. We now pass to consider $\tilde{Y}$.

5.3. **Large concentration for $\tilde{Y}$.** Consider the i.i.d matrices $\tilde{Y}$ defined from Section 2, and let $\tilde{Y}_{(n-1)\times n}$ be the matrix obtained from $\tilde{Y}$ by replacing its first row by the zero vector.

We first observe from Claim 5.1 that

$$\sigma_n(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0}) \geq \frac{1}{n}\min\big(\frac{1}{\sqrt{2n}}\sigma_{n-1}(\frac{1}{\sqrt{n}}\bar{\tilde{Y}} - z_0 I_{n-1}) - O(n^{-A}), n^{-A}\big),$$

where

$$\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0} = \begin{pmatrix} \sqrt{n} & \sqrt{n} & \cdots & \sqrt{n} \\ \frac{1}{\sqrt{n}}\tilde{y}_{21} & \frac{1}{\sqrt{n}}\tilde{y}_{22} - z_0 & \cdots & \frac{1}{\sqrt{n}}\tilde{y}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n}}\tilde{y}_{n1} & \frac{1}{\sqrt{n}}\tilde{y}_{n2} & \cdots & \frac{1}{\sqrt{n}}\tilde{y}_{nn} - z_0 \end{pmatrix}.$$

On the other hand, conditioning on $\tilde{y}_{21},\dots,\tilde{y}_{n1}$, the entries $\tilde{y}_{ij}-\tilde{y}_{i1}$ of the matrix $\bar{\tilde{Y}}$ are independent, and so we can apply known singularity bounds, for instance [31, Theorem 2.1], for i.i.d matrices to conclude that for any positive constant B, there exists a positive constant A such that $\mathbf{P}(\sigma_{n-1}(\frac{1}{\sqrt{n}}\bar{\tilde{Y}}-z_0I_{n-1}\le n^{-A})=O(n^{-B})$. Returning to $\tilde{Y}_{(n-1)\times n}$, we hence obtain the following.

Theorem 5.4. *For any positive constant B, there exists a positive constant A such that*

$$\mathbf{P}(\sigma_n(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})\le n^{-A})=O(n^{-B}).$$

This bound will be exploited later on.

Next, let H denote the following Hermitian matrix

$$H:=(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})^*(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}).$$

It is clear that the eigenvalues $\lambda_1(H),\dots,\lambda_n(H)$ of H can be written as

$$\lambda_1(H)=\sigma_1^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}),\dots,\lambda_n(H)=\sigma_n^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}),$$

where $\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})$ are the singular values of $\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}$.

The following concentration result will serve as our main lemma.

Lemma 5.5. *Assume that f is a function so that $g(x):=f(x^2)$ is convex and has finite Lipshitz norm $\|g\|_L$. Then for any $\delta\ge CK\|g\|_L/n$, where $K=10\log n$ is the upper bound for the entries of $\tilde{Y}_{(n-1)\times n}$ and C is a sufficiently large absolute constant, we have*

$$\mathbf{P}\left(\Big|\sum_{i=1}^n f(\lambda_i(H))-\mathbf{E}(\sum_{i=1}^n f(\lambda_i(H)))\Big|\ge\delta n\right)=O\left(\exp(-C'\frac{n^2\delta^2}{K^2\|g\|_L^2})\right),$$

here C' and the implied constant depend on C.

Remark that when F_{z_0} vanishes, Lemma 5.5 is essentially [17, Corollary 1.8] of Guionnet and Zeitouni. We will show that the method there can be easily extended for any deterministic matrix F_{z_0}.

Proof. (of Lemma 5.5) Consider the following Hermitan matrix K_{2n} of size $2n\times 2n$

$$K_{2n}=\begin{pmatrix}0 & (\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})^*\\ \frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0} & 0\end{pmatrix}.$$

Apparently,

$$K_{2n}^2 = \begin{pmatrix} (\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})^*(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0}) & 0 \\ 0 & (\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})^* \end{pmatrix}.$$

So to prove Lemma 5.5, it is enough to show that

$$\mathbf{P}\left(\Big|\sum_{i=1}^{2n} g(\lambda_i(K_{2n})) - \mathbf{E}(\sum_{i=1}^{2n} g(\lambda_i(K_{2n})))\Big| \geq 2\delta n\right) = O\left(\exp(-C'\frac{n^2\delta^2}{K^2\|g\|_L^2})\right), \tag{18}$$

where $\lambda_i(K_{2n})$ are the eigenvalues of K_{2n}.

Next, by following [17, Lemma 1.2] we obtain the following.

Lemma 5.6. *The function $M \mapsto \operatorname{tr}(g(\frac{1}{\sqrt{n}}M+F))$ of Hermitian matrices $M = (m_{ij})_{1\leq i,j\leq n}$, where F is a deterministic Hermitian matrix whose entries may depend on n, is a*

- *convex function;*
- *Lipschitz function of constant bounded by $2\|g\|_L$.*

We refer the reader to Appendix A for a proof of Lemma 5.6. To deduce (18) from Lemma 5.6, we apply the following well-known Talagrand concentration inequality [29].

Lemma 5.7. *Let $\mathbf{D}$ be the disk $\{z \in \mathbf{C}, |z| \leq K\}$. For every product probability μ in $\mathbf{D}^N$, every convex function $F : \mathbf{C}^N \mapsto \mathbf{R}$ of Lipschitz norm $\|F\|_L$, and every $r \geq 0$,*

$$\mathbf{P}(|F - M(F)| \geq r) \leq 4\exp(-r^2/16K^2\|F\|_L^2),$$

where $M(F)$ denotes the median of F.

Indeed, let F be the function : $\tilde{Y}' \mapsto \operatorname{tr}(g(K_{2n})) = \operatorname{tr}(g(\frac{1}{\sqrt{n}}\tilde{Y}' + F'))$, where

$$\tilde{Y}' = \begin{pmatrix} 0 & \tilde{Y}^*_{(n-1)\times n} \\ \tilde{Y}_{(n-1)\times n} & 0 \end{pmatrix}$$

and

$$F' = \begin{pmatrix} 0 & -F^*_{z_0} \\ -F_{z_0} & 0 \end{pmatrix}.$$

Observe that the entries of $\tilde{Y}'$ are supported on $|x| \le K = 10\log n$. By Lemma 5.6, F is convex function with Lipschitz constant bounded by $2\|g\|_L$. The conclusion (18) of Lemma 5.5 then follows by applying Lemma 5.7.

$\square$

In what follows we will apply Lemma 5.5 for two functions, one gives an almost complete control on the large spectra of H and one yields a good bound on the number of small spectra of H. We will choose c to be a sufficiently small constant, and with room to spare we set

$$\epsilon = \delta = \Theta(n^{-c}).$$

5.8. **Concentration of large spectra for i.i.d matrices.** Following [10] and [13], we first apply Lemma 5.5 to the cut-off function $f_\epsilon(x) := \log(\max(\epsilon, x))$. Note that $f_\epsilon(x^2)$ has Lipschitz constant $2\epsilon^{-1/2}$. Although the function is not convex, it is easy to write it as a difference of two convex functions of Lipschitz constant $O(\epsilon^{-1/2})$, and so Lemma 5.5 applies because $\delta = \Theta(n^{-c}) \ge C\epsilon^{1/2}K/n$.

Theorem 5.9. *We have*

$$\mathbf{P}\left(\Big|\sum_{\sigma_i^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})\in S_\epsilon} \log\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0}) - \mathbf{E}\big(\sum_{\sigma_i^2(\dots)\in S_\epsilon}\log\sigma_i(\dots)\big)\Big| \ge \delta n\right)$$

$$= O\big(\exp(-n^2\delta^2\epsilon/K^2)\big) = O(\exp(-n\log^2 n)),$$

where $S_\epsilon := \{x \in \mathbf{R}, x \ge \epsilon\}$.

For short, from now on we set

$$h_{\epsilon,\tilde{Y}_{(n-1)\times n}}(z_0) := \frac{1}{n}\mathbf{E}\big(\sum_{\sigma_i^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})\in S_\epsilon} \log\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})\big).$$

Serving as the main term, $h_{\epsilon,\tilde{Y}_{(n-1)\times n}}(z_0)$ will play a key role in our analysis. In our next subsection we apply Lemma 5.5 to another function f.

5.10. **Concentration of the number of small eigenvalues for i.i.d matrices.** Let I be the interval $[0,\epsilon]$. We are going to show that the number N_I of the eigenvalues $\lambda_i(H)$ which belong to I is small with very high probability.

It is not hard to construct two functions f_1, f_2 such that $(f_1 - f_2) - \mathbf{1}_I$ is non-negative and supported on an interval of length ϵ/C, and so that both of $g_1(x) = f_1(x^2)$ and $g_2(x) = f_2(x^2)$ are convex functions of Lipschitz constant $O(\epsilon^{-1/2})$. (For instance one may construct $f_1(x), f_2(x)$ in such a way that the even function $g_1(x) = f_1(x^2)$ is identical to 1 on the interval $[\epsilon^{1/2}, \epsilon^{1/2}]$ and being straight concave down from both edges with a slope of $O(\epsilon^{-1/2})$, while the graph of the function $g_2(x) = f_2(x^2)$ is obtained from that of $g_1(x)$ by replacing its positive part with zero).

Next, by Lemma 5.5 we have

$$\mathbf{P}\left(\Big|\sum_{\lambda_i(H)} f_1(\lambda_i(H)) - \mathbf{E}(\sum_{\lambda_i(H)} f_1(\lambda_i(H)))\Big| \geq \delta n\right) = O\big(\exp(-n\log^2 n)\big),$$

and

$$\mathbf{P}\left(\Big|\sum_{\lambda_i(H)} f_2(\lambda_i(H)) - \mathbf{E}(\sum_{\lambda_i(H)} f_2(\lambda_i(H))\Big| \geq \delta n\right) = O\big(\exp(-n\log^2 n)\big).$$

By the triangle inequality, we thus have

$$\mathbf{P}\left(\Big|\sum_{\lambda_i(H)} (f_1 - f_2)(\lambda_i(H)) - \mathbf{E}(\sum_{\lambda_i(H)} (f_1 - f_2)(\lambda_i(H)))\Big| \geq 2\delta n\right) = O\big(\exp(-n\log^2 n)\big).$$

Because the error-function $f = (f_1 - f_2) - \mathbf{1}_I$ is nonnegative, it follows that with probability $1 - O(\exp(-n\log^2 n))$

$$\sum_{\lambda_i(H)} \mathbf{1}_I(\lambda_i(H)) + \sum_{\lambda_i(H)} f(\lambda_i(H)) \leq \mathbf{E}\big(\sum_{\lambda_i(H)} (f_1 - f_2)(\lambda_i(H))\big) + 2\delta n,$$

and hence

$$\begin{aligned} N_I = \sum_{\lambda_i(H)} \mathbf{1}_I(\lambda_i(H)) &\leq \mathbf{E}\big(\sum_{\lambda_i(H)} (f_1 - f_2)(\lambda_i(H))\big) + 2\delta n \\ &\leq 2\mathbf{E}\big(\sum_{\lambda_i(H)} \mathbf{1}_J(\lambda_i(H))\big) + 2\delta n \\ &\leq 2\mathbf{E}(N_J) + 2\delta n, \end{aligned}$$

where J is the interval $[0, \epsilon+\epsilon/C]$ and N_J is the number of eigenvalues of H in J. (Strictly speaking, we have to set $J=[-\epsilon/C, \epsilon+\epsilon/C]$. However, as λ_i are non-negative, we can omit its negative interval.)

To exploit the above information furthermore, we apply a result saying that N_J has small expected value (see also [37, Proposition 28] and the references therein).

Lemma 5.11. *For all $J \subset \mathbf{R}$ with $|J| \geq K^2 \log^2 n/n^{1/2}$, one has*

$$N_J \ll n|J|$$

with probability $1-\exp(-\omega(\log n))$. In particular,

$$\mathbf{E}(N_J) \leq Cn|J|,$$

where C is a sufficiently large constant.

Remark that this result holds for any deterministic matrix F_0 in the definition of H. We defer the proof of Lemma 5.11 to Appendix B.

In summary, we have obtained the following result.

Theorem 5.12. *With probability $O(\exp(-n\log^2 n))$, we have*

$$N_I \geq 2C\epsilon n + 2\delta n,$$

where N_I is the number of $\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})$ such that $\sigma_i^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0}) \in [0,\epsilon]$.

Consequently, it follows from Theorems 5.4 and 5.12 that with probability $1-O(n^{-B})$ the following holds

$$\frac{1}{n}\sum_{\sigma_i^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})\in[0,\epsilon]} \log \sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0}) = O((\epsilon+\delta)\log n) = O(n^{-c}\log n).$$

Thus, combining with Theorem 5.9, we infer the following.

Theorem 5.13. *Let z_0 be fixed and let B be a positive constant. Then the following holds with probability $1-O(n^{-B})$*

$$\left|\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})| - h_{\epsilon,\tilde{Y}_{(n-1)\times n}}(z_0)\right| \leq 2\delta + O(n^{-c}\log n) = O(n^{-c}\log n),$$

where the implied constants depend on B.

5.14. **Asymptotic formula for $h_{\epsilon,\tilde{Y}_{(n-1)\times n}}(z_0)$.** We next claim that $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})|$ also converges to the corresponding part of the circular law, and so giving an asymptotic formula for $h_{\epsilon,\tilde{Y}_{(n-1)\times n}}(z_0)$.

Theorem 5.15. *For almost all z_0, the following holds with probability one*

$$\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0})| - \int_{\mathbf{C}} \log|w-z_0| d\mu_{\mathbf{cir}}(w) = o(1). \tag{19}$$

Note that this result is more or less a circular law for random matrices of i.i.d. entries. To prove it we just simply rely on [34].

Proof. (of Theorem 5.15) We first pass to $\bar{\tilde{Y}}$

$$\bar{\tilde{Y}} = \begin{pmatrix} \tilde{y}_{22} - \tilde{y}_{21} & \cdots & \tilde{y}_{2n} - \tilde{y}_{21} \\ \tilde{y}_{32} - \tilde{y}_{31} & \cdots & \tilde{y}_{3n} - \tilde{y}_{31} \\ \vdots & \vdots & \vdots \\ \tilde{y}_{n2} - \tilde{y}_{n1} & \cdots & \tilde{y}_{nn} - \tilde{y}_{n1} \end{pmatrix},$$

where $\tilde{y}_{ij}$ are i.i.d. copies of $\tilde{y}$.

As

$$\det(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n} - F_{z_0}) = \sqrt{n}\det(\frac{1}{\sqrt{n}}\bar{\tilde{Y}} - z_0 I_{n-1}),$$

it is enough to prove the claim for $\det(\frac{1}{\sqrt{n}}\bar{\tilde{Y}} - z_0 I_{n-1})$.

View $\bar{\tilde{Y}}$ as a sum of the matrix $(\tilde{y}_{ij})_{2\le i,j\le n}$ and R, the $(n-1)\times(n-1)$ matrix formed by $(-\tilde{y}_{i1},\dots,-\tilde{y}_{i1})$ for $2\le i\le n$. Because R has rank one and the average square of its entries $\frac{1}{n-1}\sum_i \tilde{y}_{i1}^2$ is bounded almost surely (with respect to $\tilde{y}_{21},\dots,\tilde{y}_{n1}$), [34, Corollary 1.15] applied to $\bar{\tilde{Y}}$ implies that the ESD of $\frac{1}{\sqrt{n}}\bar{\tilde{Y}}$ converges almost surely to the circular law.

Finally, thanks to [34, Theorem 1.20], for almost all z_0 the following holds with probability one

$$\frac{1}{n}\log\big|\det(\frac{1}{\sqrt{n}}\bar{\tilde{Y}} - z_0 I_{n-1})\big| - \int_{\mathbf{C}} \log|w-z_0| d\mu_{\mathbf{cir}}(w) = o(1).$$

□

Theorems 5.13 and 5.15 immediately imply that for almost all z_0

$$h_{\epsilon,\tilde{Y}_{(n-1)\times n}}(z_0) - \int_{\mathbf{C}} \log|w-z_0| d\mu_{\mathbf{cir}}(w) = o(1). \tag{20}$$

By substituting (20) back to Theorem 5.9, we have

$$\mathbf{P}\left(\Big|\frac{1}{n}\sum_{\sigma_i^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})\in S_\epsilon} \log\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}) - \int_{\mathbf{C}}\log|w-z_0|d\mu_{\mathbf{cir}}(w)\Big| \ge \delta + o(1)\right)$$

$$= O(\exp(-n\log^2 n)). \tag{21}$$

6. Large concentration for $\tilde{X}$, proof of Theorem1.10

In this section we will apply the transference principle of Lemma 2.3 to pass the results of Section 5 back to $\tilde{X}$. Our treatment here is similar to [8, Section 4].

By Lemma 2.3 and (21), conditioning on $\tilde{Y}\in\tilde{D}_n$ we have

$$\mathbf{P}\left(\Big|\frac{1}{n}\sum_{\sigma_i^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})\in S_\epsilon} \log\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}) - \int_{\mathbf{C}}\log|w-z_0|d\mu_{\mathbf{cir}}(w)\Big| \ge \delta + o(1)|\tilde{Y}\in\tilde{D}_n\right)$$

$$= O(n^{10n}\exp(-n\log^2 n)) = O(\exp(-n\log^2 n/2)). \tag{22}$$

Next, for each $\tilde{Y}\in\tilde{D}_n$ we will compare the singular values of $\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}$ with those of $\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0}$, where $\tilde{X}$ is determined by $\Phi(\frac{1}{n}\tilde{Y})$, i.e. $\tilde{x}_{ij}=\frac{1}{n}\tilde{y}_{ij}$ for all $2\le i,j\le n$.

By definition, as $\tilde{Y}\in\tilde{D}_n$, we have $|\frac{1}{n}\tilde{y}_{i1}-\tilde{x}_{i1}|\le n^{-4}$, and so the operator norm of the difference matrix is bounded by

$$\Big\|(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}) - (\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})\Big\| \le \frac{1}{n^2}.$$

This leads to a similar bound for the singular values for every i (see for instance [19])

$$\Big|\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0}) - \sigma_i(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})\Big| \le \frac{1}{n^2}. \tag{23}$$

Notice furthermore that, conditioning on $\tilde{Y}\in\tilde{D}_n$, $\Phi(\frac{1}{n}\tilde{Y})$ is uniformly distributed on the set $\tilde{S}_n$ of bounded doubly stochastic matrices $\tilde{X}$. Thus, by a slight modification of ϵ by an amount of n^{-2} (thus the order of ϵ remains $\Theta(n^{-c})$), we obtain from (22) the following upper tail bound with respect to $\tilde{X}$

$$\mathbf{P}\left(\frac{1}{n}\sum_{\sigma_i^2(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})\in S_{\epsilon+n^{-2}}}\log\sigma_i(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})-\int_{\mathbf{C}}\log|w-z_0|d\mu_{\mathbf{cir}}(w)\geq\delta+o(1)\right)$$

$$=O(\exp(-n\log^2 n/2)).$$

Also, we obtain a similar probability bound for the lower tail

$$\mathbf{P}\left(\frac{1}{n}\sum_{\sigma_i^2(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})\in S_{\epsilon-n^{-2}}}\log\sigma_i(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})-\int_{\mathbf{C}}\log|w-z_0|d\mu_{\mathbf{cir}}(w)\leq-(\delta+o(1))\right)$$

$$=O(\exp(-n\log^2 n/2)).$$

Notice that these bounds hold for any $\epsilon=\Theta(n^{-c})$. By gluing them together we infer the following variant of (22).

Theorem 6.1. *With respect to $\tilde{X}$ we have*

$$\mathbf{P}\left(\Big|\frac{1}{n}\sum_{\sigma_i^2(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})\in S_{\epsilon}}\log\sigma_i(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})-\int_{\mathbf{C}}\log|w-z_0|d\mu_{\mathbf{cir}}(w)\Big|\geq\delta+o(1)\right)$$

$$=O(\exp(-n\log^2 n/2)).$$

Next, conditioning on $\tilde{Y}\in\tilde{D}_n$, by Theorem 5.12 and Lemma 2.3, with probability $O(n^{10n}\exp(-n\log^2 n))=O(\exp(-n\log^2 n/2))$ we have

$$N_I\geq 2C\epsilon n+2\delta n,$$

where N_I is the number of $\sigma_i(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})$ such that $\sigma_i^2(\frac{1}{\sqrt{n}}\tilde{Y}_{(n-1)\times n}-F_{z_0})\in[0,\epsilon]$.

Because $\Phi(\frac{1}{n}\tilde{Y})$ is uniformly distributed on the set $\tilde{S}_n$ conditioning on $\tilde{Y}\in\tilde{D}_n$, and also because of (23), we imply the following.

Theorem 6.2. *With probability $O(\exp(-n\log^2 n))$ with respect to $\tilde{X}$, we have*

$$N_I\geq 2C(\epsilon+\frac{1}{n^2})n+2\delta n,$$

where N_I is the number of $\sigma_i(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0})$ such that $\sigma_i^2(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}) \in [0,\epsilon]$.

We now gather the ingredients together to complete the proof of our main result.

Proof. (of Theorem 1.10 for $\tilde{X}$) By Theorems 5.2 and 6.2, we have that

$$\mathbf{P}\Big(\frac{1}{n}\sum_{\sigma_i^2(\sqrt{n}\tilde{X}_{(n-1)\times n}-F_{z_0})\in[0,\epsilon]} \log\sigma_i(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}) = O((\epsilon+\delta)\log n)\Big) = 1 - O(n^{-3}).$$

A combination of this fact with Theorem 6.1 implies that for almost all z_0,

$$\mathbf{P}\left(\Big|\frac{1}{n}\log|\det(\sqrt{n}\tilde{X}_{(n-1)\times n} - F_{z_0}) - \int_{\mathbf{C}}\log|w-z_0|d\mu_{\mathbf{cir}}(w)\Big| = o(1)\right) = 1 - O(n^{-3}).$$

Hence, by (17),

$$\mathbf{P}\left(\Big|\frac{1}{n}\log|\det(\sqrt{n}\bar{\bar{X}} - z_0 I_{n-1}) - \int_{\mathbf{C}}\log|w-z_0|d\mu_{\mathbf{cir}}(w)\Big| = o(1)\right) = 1 - O(n^{-3}),$$

completing the proof.

□

Appendix A. Proof of Lemma 5.6

The main goal of this section is to justify Lemma 5.6. Although our proof is identical to [17, Theorem 1.1] and [17, Corollary 1.8], let us present it here for the sake of completeness.

A.1. **Convexity.** For simplicity, we first show that the function $M \mapsto \mathrm{tr}(g(M+F))$ is convex. It then follows that the function $M \mapsto \mathrm{tr}(g(\frac{1}{\sqrt{n}}M+F))$ is also convex.

For any Hermitian matrices U and V

$$g(V+F) - g(U+F) = \int_0^1 Dg\Big(U+F+\eta(V-U)\Big)\sharp(V-U)d\eta$$

where

$$Dg(U+F)\sharp(V) = \lim_{\epsilon\to 0}\epsilon^{-1}\Big(g(U+F+\epsilon V) - g(U+F)\Big).$$

For polynomial functions g, the non-commutative derivation D can be computed and one finds in particular that for any $p \in \mathbf{N}$,

$$(V+F)^p-(U+F)^p = \int_0^1 \left(\sum_{k=0}^{p-1}(U+F+\eta(V-U))^k(V-U)(U+F+\eta(V-U))^{p-k-1}\right) d\eta. \tag{24}$$

For such a polynomial function, by taking the trace and using $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, one deduces that

$$\mathrm{tr}\big((U+F)^p\big)-\mathrm{tr}\Big((\frac{U+V}{2}+F)^p\Big) = p\int_0^1 \mathrm{tr}\left((\frac{U+V}{2}+F+\eta\frac{U-V}{2})^{p-1}\frac{U-V}{2}\right) d\eta, \tag{25}$$

$$\mathrm{tr}\big((V+F)^p\big)-\mathrm{tr}\big((\frac{U+V}{2}+F)^p\big) = p\int_0^1 \mathrm{tr}\left((\frac{U+V}{2}+F-\eta\frac{U-V}{2})^{p-1}\frac{V-U}{2}\right) d\eta. \tag{26}$$

It follows from (24),(25) and (26) that

$$\begin{aligned}\Delta &:= \mathrm{tr}\big((U+F)^p\big)+\mathrm{tr}\big((V+F)^p\big)-2\mathrm{tr}\big((\frac{U+V}{2}+F)^p\big)\\ &= \frac{p}{2}\sum_{k=0}^{p-2}\int_0^1\int_0^1 \eta d\eta d\theta \mathrm{tr}\Big((U-V)Z_{\eta,\theta}^k(U-V)Z_{\eta,\theta}^{p-2-k}\Big)\end{aligned} \tag{27}$$

with

$$Z_{\eta,\theta} := \frac{U+V}{2}+F-\eta\frac{U-V}{2}+\eta\theta(U-V).$$

Next, for fixed $\eta, \theta \in [0,1]^2$, and fixed U, V, F Hermitian matrices, $Z_{\eta,\theta}$ is also Hermitian, and so we can find a unitary matrix $U_{\eta,\theta}$ and a diagonal matrix $D_{\eta,\theta}$ with real diagonal entries $\lambda_{\eta,\theta}(1), \ldots, \lambda_{\eta,\theta}(n)$ so that

$$Z_{\eta,\theta} = U_{\eta,\theta}D_{\eta,\theta}U_{\eta,\theta}^*.$$

Let $W_{\eta,\theta} = U_{\eta,\theta} = U_{\eta,\theta}^*(U-V)U_{\eta,\theta}$. Then

$$\begin{aligned}\Delta &= \frac{p}{2}\sum_{k=0}^{p-2}\int_0^1\int_0^1 \eta d\eta d\theta \mathrm{tr}\Big(W_{\eta,\theta}D_{\eta,\theta}^k W_{\eta,\theta}D_{\eta,\theta}^{p-2-k}\Big)\\ &= \frac{p}{2}\sum_{k=0}^{p-2}\int_0^1\int_0^1 \eta d\eta d\theta \sum_{k=0}^{p-2}\sum_{1\le i,j\le n}\lambda_{\eta,\theta}^k(i)\lambda_{\eta,\theta}^{p-2-k}(j)|W_{\eta,\theta}(ij)|^2. \qquad (28)\end{aligned}$$

But

$$\sum_{k=0}^{p-2}\lambda_{\eta,\theta}^k(i)\lambda_{\eta,\theta}^{p-2-k}(j) = \frac{\lambda_{\eta,\theta}^{p-1}(i)-\lambda_{\eta,\theta}^{p-1}(j)}{\lambda_{\eta,\theta}(i)-\lambda_{\eta,\theta}(j)} = (p-1)\int_0^1\Big(\alpha\lambda_{\eta,\theta}(j)+(1-\alpha)\lambda_{\eta,\theta}(i)\Big)^{p-2}d\alpha.$$

Hence, substituting in (28) gives,

$$\Delta = \frac{1}{2}\sum_{1\le i,j\le n}\int_0^1\int_0^1\int_0^1 d\alpha\eta d\eta d\theta |W_{\eta,\theta}(ij)|^2 g''(\alpha\lambda_{\eta,\theta}(j)+(1-\alpha)\lambda_{\eta,\theta}(i)) \ge 0 \qquad (29)$$

for the polynomial $g(x) = x^p$.

Now, with U, V, F being fixed, the eigenvalues $\lambda_{\eta,\theta}(1),\dots,\lambda_{\eta,\theta}(n)$ and the entries of $W_{\eta,\theta}$ are uniformly bounded. Hence, by Runge's theorem, we can deduce by approximation that (29) holds for any twice continuously differentiable function g. As a consequence, for any such convex function, $g'' \ge 0$ and

$$\Delta = \mathrm{tr}\big(g(U+F)\big) + \mathrm{tr}\big(g(V+F)\big) - 2\mathrm{tr}\big(g(\frac{U+V}{2}+F)\big) \ge 0.$$

A.2. **Boundedness.** Now we show that the function $M \mapsto \mathrm{tr}(g(\frac{1}{\sqrt{n}}M+F))$ has Lipschitz constant bounded by $2\|g\|_L$.

First, for any bounded continuously differentiable function g we will show that

$$\sum_{1\le i,j\le n}\Big(d_{\Re(x_{ij})}\mathrm{tr}(g(\frac{1}{\sqrt{n}}M+F))\Big)^2 + \sum_{1\le i,j\le n}\Big(d_{\Im(x_{ij})}\mathrm{tr}(g(\frac{1}{\sqrt{n}}M+F))\Big)^2 \le 4\|g\|_L^2.$$

We can verify that

$$d_{\Re(x_{ij})}\mathrm{tr}\Big(g(\frac{1}{\sqrt{n}}M+F)\Big) = \frac{1}{\sqrt{n}}\mathrm{tr}\Big(g'(\frac{1}{\sqrt{n}}M+F)\Delta_{ij}\Big) \qquad (30)$$

where $\Delta_{ij}(kl) = 1$ if $kl = ij$ or ji and zero otherwise.

Indeed, (30) is a consequence of (24) for polynomial functions, and it can be extended for bounded continuously differentiable functions by approximations. In other words, we have

$$d_{\Re(x_{ij})}\mathrm{tr}\big(g(\frac{1}{\sqrt{n}}M+F)\big) = \begin{cases} \frac{1}{\sqrt{n}}\Big(g'(\frac{1}{\sqrt{n}}M+F)(ij)+g'(\frac{1}{\sqrt{n}}M+F)(ji)\Big) & i\neq j; \\ \\ \frac{1}{\sqrt{n}}g'(\frac{1}{\sqrt{n}}M+F)(ii) & i=j. \end{cases}$$

Hence,

$$\sum_{i,j}\Big(d_{\Re(x_{ij})}\mathrm{tr}\big(g(\frac{1}{\sqrt{n}}M+F)\big)\Big)^2 \le \frac{2}{n}\sum_{i,j}|g'(\frac{1}{\sqrt{n}}M+F)(ij)|^2 = \frac{2}{n}\mathrm{tr}\Big(g'(\frac{1}{\sqrt{n}}M+F)g'(\frac{1}{\sqrt{n}}M+F)^*\Big).$$

But if $\lambda_1,\dots,\lambda_n$ denote the eigenvalues of $\frac{1}{\sqrt{n}}M+F$ then

$$\mathrm{tr}\Big(g'(\frac{1}{\sqrt{n}}M+F)g'(\frac{1}{\sqrt{n}}M+F)^*\Big) = \frac{1}{n}\sum(g'(\lambda_i))^2 \le \|g'\|_\infty^2.$$

Thus we have

$$\sum_{i,j}\Big(d_{\Re(x_{ij})}\mathrm{tr}\big(g(\frac{1}{\sqrt{n}}M+F)\big)\Big)^2 \le 2\|g'\|_\infty^2.$$

The same argument applies for derivatives with respect to $\Im(x_{ij})$, and so by integration by parts and by Cauchy-Schwarz inequality

$$\Big|\mathrm{tr}\big(g(\frac{1}{\sqrt{n}}U+F)\big)-\mathrm{tr}\big(g(\frac{1}{\sqrt{n}}V+F)\big)\Big| \le 2\|g\|_L\|U-V\|$$

for any U and V.

Observe that this last result for bounded continuously differentiable function g naturally extends to Lipschitz functions by approximations, completing the proof.

Appendix B. Proof of Lemma 5.11

Note that if F_{z_0} vanishes then this is [37, Proposition 28] (see also [2]). We show that the method there extends easily to any deterministic F_{z_0}.

Assume for contradiction that

$$|N_J| \ge Cn|J|$$

for some large constant C to be chosen later. We will show that this will lead to a contradiction with high probability.

We will control the eigenvalue counting function N_J via the Stieltjes transform

$$s(z) := \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j(H) - z}.$$

Fix J and let x be the midpoint of J. Set $\eta := |J|/2$ and $z := x + i\eta$, we then have

$$\Im(s(z))\frac{4}{5}\frac{N_J}{\eta n}.$$

Hence,

$$\Im(s(z)) \gg C. \tag{31}$$

Next, with $H' := (\frac{1}{\sqrt{n}}\Phi(\tilde{Y}) - F_{z_0})(\frac{1}{\sqrt{n}}\Phi(\tilde{Y}) - F_{z_0})^* = \frac{1}{n}MM^*$ where $M := \Phi(\tilde{Y}) - \sqrt{n}F_{z_0}$, we have (see also [2, Chapter 11])

$$s(z) = \frac{1}{n}\sum_{k\le n} \frac{1}{h'_{kk} - z - \mathbf{a}_k^*(H'_k - zI)^{-1}\mathbf{a}_k},$$

where h'_{kk} is the kk entry of H'; H'_k is the $n-1$ by $n-1$ matrix with the k-th row and k-th column of H' removed; and $\mathbf{a}_k$ is the k-th column of H' with the k-th entry removed.

Note that $\Im(\frac{1}{z}) \le \frac{1}{\Im(z)}$, one concludes from (31) that

$$\frac{1}{n}\sum_{k\le n} \frac{1}{\left|\eta + \Im(\mathbf{a}_k^*(H'_k - zI)^{-1}\mathbf{a}_k)\right|} \gg C.$$

By the pigeonhole principle, there exists k such that

$$\frac{1}{\left|\eta + \Im(\mathbf{a}_k^*(H'_k - zI)^{-1}\mathbf{a}_k)\right|} \gg C. \tag{32}$$

Fix such k, note that

$$\mathbf{a}_k = \frac{1}{n}M_k\mathbf{r}_k^*, \text{ and } H'_k = \frac{1}{n}M_kM_k^*$$

where $\mathbf{r}_k = \mathbf{r}_k(M)$ and M_k is the $(n-1) \times n$ matrix formed by removing $\mathbf{r}_k(M)$ from M. Thus if we let $\mathbf{v}_1 = \mathbf{v}_1(M_k), \dots, \mathbf{v}_{n-1} = \mathbf{v}_{n-1}(M_k)$ and $\mathbf{u}_1 = \mathbf{u}_1(M_k), \dots, \mathbf{u}_{n-1} = \mathbf{u}_{n-1}(M_k)$ be the orthogonal systems of left and right singular vectors of M_k, and let $\lambda_j = \lambda_j(H_k') = \frac{1}{n}\sigma_j^2(M_k)$ be the associated eigenvalues, one has

$$\mathbf{a}_k^*(H_k' - zI)^{-1}\mathbf{a}_k = \sum_{1 \le j \le n-1} \frac{|\mathbf{a}_k^* \mathbf{v}_j|^2}{\lambda_j - z}.$$

Thus

$$\Im\big(\mathbf{a}_k^*(H_k' - zI)^{-1}\mathbf{a}_k\big) \ge \eta \sum_{1 \le j \le n-1} \frac{|\mathbf{a}_k^* \mathbf{v}_j|^2}{\eta^2 + |\lambda_j - x|^2}.$$

We conclude from (32) that

$$\sum_{1 \le j \le n-1} \frac{|\mathbf{a}_k^* \mathbf{v}_j|^2}{\eta^2 + |\lambda_j - x|^2} \ll \frac{1}{C\eta}.$$

Note that $\mathbf{a}_k^* \mathbf{v}_j$ can be written as

$$\mathbf{a}_k^* \mathbf{v}_j = \frac{\sigma_j(M_k)}{n} \mathbf{r}_k \mathbf{u}_j.$$

Next, from the Cauchy interlacing law, one can find an interval $L \subset \{1, \dots, n-1\}$ of length

$$|L| \gg C\eta n$$

such that $\lambda_j \in L$. We conclude that

$$\sum_{j \in L} \frac{\sigma_j^2}{n^2} |\mathbf{r}_k \mathbf{u}_j|^2 \ll \frac{\eta}{C}.$$

Since $\lambda_j \in J$, one has $\sigma_j = \Theta(\sqrt{n})$, and thus

$$\sum_{j \in L} |\mathbf{r}_k \mathbf{u}_j|^2 \ll \frac{\eta n}{C}.$$

The LHS can be written as $\|\pi_V(\mathbf{r}_k^*)\|^2$, where V is the span of the eigenvectors $\mathbf{u}_j$ for $j \in L$ and $\pi_V(.)$ is the projection onto V. But from Talagrand inequality for distance (Lemma

B.1 below), we see that this quantity is $\gg \eta n$ with very high probability, giving the desired contradiction.

Lemma B.1. *Assume that $V \subset \mathbf{C}^n$ is a subspace of dimension $\dim(V) = d \le n - 10$. Let $\mathbf{f}$ be a fixed vector (whose coordinates may depend on n). Let $\mathbf{y} = (0, y_2, \dots, y_n)$, where $y = \tilde{y}_i - 1$ and $\tilde{y}_i$ are i.i.d. copies of $\tilde{y}$ defined from (2). Let $\sigma = \Theta(1)$ denote the standard deviation of $\tilde{y}$ and $K = 10 \log n$ denote the upper bound of $\tilde{y}$, then for any $t > 0$ we have*

$$\mathbf{P}_{\mathbf{y}}\left(\pi_V(\mathbf{y} + \mathbf{f}) \ge \sqrt{2}\sigma\sqrt{d}/2 - O(K) - t\right) \ge 1 - O\Big(\exp(-\frac{t^2}{16K^2})\Big).$$

We now give a proof of Lemma B.1. It is clear that the function $(y_2, \dots, y_n) \mapsto \pi_V(\mathbf{y} + \mathbf{f})$ is convex and 1-Lipschitz. Thus by Theorem 5.7 we have

$$\mathbf{P}_{\mathbf{y}}\Big(|\pi_V(\mathbf{y} + \mathbf{f}) - M(\pi_V(\mathbf{y} + \mathbf{f}))| \ge t\Big) = O\Big(\exp(-16t^2/K^2)\Big). \quad (33)$$

Hence, it is implied that

$$\begin{aligned}\mathbf{P}_{\mathbf{y},\mathbf{y}'}\Big(|\pi_V(\mathbf{y} + \mathbf{f}) + \pi_V(\mathbf{y}' + \mathbf{f}) - 2M(\pi_V(\mathbf{y} + \mathbf{f}))| \le 2t\Big) &= \Big(1 - O(\exp(-16t^2/K^2))\Big)^2 \\ &= 1 - O\Big(\exp(-16t^2/K^2)\Big), \quad (34)\end{aligned}$$

where $\mathbf{y}'$ is an independent copy of $\mathbf{y}$.

On the other hand, by the triangle inequality

$$\pi_V(\mathbf{y} + \mathbf{f}) + \pi_V(\mathbf{y}' + f) \ge \pi_V(\mathbf{y} - \mathbf{y}').$$

Applying Talagrand inequality once more for the random vector $\mathbf{y} - \mathbf{y}'$ (see for instance [36, Lemma 68]), we see that

$$\mathbf{P}_{\mathbf{y},\mathbf{y}'}\Big(|\pi_V(\mathbf{y} - \mathbf{y}') - \sqrt{2}\sigma\sqrt{d}| \ge t\Big) = O\Big(\exp(-t^2/16K^2)\Big).$$

Thus,

$$\mathbf{P}_{\mathbf{y},\mathbf{y}'}\Big(\pi_V(\mathbf{y}) + \pi_V(\mathbf{y}') \ge \sqrt{2}\sigma\sqrt{d} - t\Big) = 1 - O\Big(\exp(-t^2/16K^2)\Big).$$

By comparing with (34), we deduce that

$$M(\pi_V(\mathbf{y} + \mathbf{f})) \ge \sqrt{1/2}\sigma\sqrt{d} - O(K).$$

Substituting this bound back to (34), we obtain the one-sided estimate as desired.

Acknowledgment. The author is grateful to M. Meckes for pointing out the references [22] and [27] and to A. Guionnet for helpful e-mail exchange regarding Lemma 5.5. He is particularly thankful to R. Pemantle and V. Vu for helpful discussion and enthusiastic encouragement.

References

[1] Z. D. Bai, *Circular law*, Annals of Probability 25 (1997), 494-529.

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

[3] A. Barvinok and J. A. Hartigan, *An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums*, to appear in Transactions of the American Mathematical Society, `arxiv.org/abs/0910.2477`.

[4] A. Barvinok and J. A. Hartigan, *Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes*, Advances in Applied Mathematics, 45 (2010), 252-289.

[5] A. Barvinok and J. A. Hartigan, *The number of graphs and a random graph with a given degree sequence*, to appear in Random Structures & Algorithms, `arxiv.org/abs/1003.0356`.

[6] C. Bordenave, P. Caputo, D. Chafai, *Circular law theorem for random Markov matrices*, to appear in Probability Theory and Related Fields, `arxiv.org/abs/0808.1502`.

[7] D. Chafai, *The Dirichlet Markov Ensemble*, Journal of Multivariate Analysis 101, 555-567 (2010).

[8] S. Chaterjee, P. Diaconis and A. Sly, *on properties of random doubly stochastic matrices*, to appear in Annales de l'Institut Henri Poincaré, `arxiv.org/abs/1010.6136`.

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

[10] K. P. Costello and V. Vu, *Concentration of random determinants and permanent estimators*, Siam Journal of Discrete Mathematics 23 (2009), 1356-1371.

[11] E. R. Canfield and B. D. McKay, *The asymptotic volume of the Birkhoff polytope*, Online Journal of Analytic Combinatorics (2009), no. 4.

[12] P. Erdős, *On a lemma of Littlewood and Offord*, Bulletin of the American Mathematical Society 51 (1945), 898-902.

[13] S. Friedland, B. Rider, and O. Zeitouni, *Concentration of permanent estimators for certain large matrices*, Advances in Applied Mathematics 14 (2004) 1559-1576.

[14] V. L. Girko, *Circular law*, Theory of Probability and its Application (1984), 694-706.

[15] V. L. Girko, *The strong circular law, Twenty years later*, II. Random Operators and Stochastic Equations 12 (2004), no. 3, 255-312.

[16] F. Götze and A. N. Tikhomirov, *The circular law for random matrices*, Annals of Probability 38 (2010), no. 4, 1444-1491.

[17] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electronic Communications in Probability 5 (2000) 119-136.

[18] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Periodica Mathematica Hungarica 8 (1977), no. 3-4, 197-211.

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

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

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

[22] M. Meckes and E. Meckes, *The central limit problem for random vectors with symmetries* (with E. Meckes), Journal of Theoretical Probability 20 (2007), 697-720.

[23] H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, to appear in Duke Mathematics Journal, `arxiv.org/abs/1101.3074`.

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

[25] H. Nguyen and V. Vu, *Circular law for random combinatorial matrices*, submitted, `arxiv.org/abs/1203.5941`.

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

[27] G. Paouris, *Concentration of mass in convex bodies*, Geometric and Functional Analysis 16 (2006), 1021-1049.

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

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

[30] T. Tao, *Outliers in the spectrum of i.i.d. matrices with bounded rank perturbations*, to appear in Probability Theory and related Fields, `arxiv.org/abs/1012.4818`.

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

[32] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices,* Bulletin of the American Mathematical Society (N.S.) 46 (2009), no. 3, 377–396.

[33] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2) 169 (2009), no 2, 595-632.

[34] T. Tao, V. Vu and appendix by M. Krishnapur, *Random matrices: universality of ESDs and the circular law*, Annals of Probability 38 (2010), no. 5, 2023-2065.

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

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

[37] T. Tao and V. Vu, *Random covariance matrices: universality of local statistics of eigenvalues*, to appear in Annals of Probability, `arxiv.org/abs/0912.0966`.

[38] L. Tran, K. Wang and V. Vu, *Sparse random graphs: eigenvalues and eigenvectors*, to appear in Random Structures and Algorithms, `arxiv.org/abs/1011.6646`.

[39] R. Vershynin, *Invertibility of symmetric random matrices*, submitted, `arxiv.org/abs/1102.0300`.

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

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