NORMAL VECTOR OF A RANDOM HYPERPLANE

HOI H. NGUYEN AND VAN H. VU

ABSTRACT. Let $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ be $n-1$ independent vectors in $\mathbf{R}^n$ (or $\mathbf{C}^n$). We study $\mathbf{x}$, the unit normal vector of the hyperplane spanned by the $\mathbf{v}_i$. Our main finding is that $\mathbf{x}$ resembles a random vector chosen uniformly from the unit sphere, under some randomness assumption on the $\mathbf{v}_i$.

Our result has applications in random matrix theory. Consider an $n \times n$ random matrix with iid entries. We first prove an exponential bound on the upper tail for the least singular value, improving the earlier linear bound by Rudelson and Vershynin. Next, we derive optimal delocalization for the eigenvectors corresponding to eigenvalues of small modulus.

1. INTRODUCTION

A real random variable ξ is *normalized* if it has mean 0 and variance 1. A complex random variable ξ is normalized if $\xi = \frac{1}{\sqrt{2}}\xi_1 + \frac{1}{\sqrt{2}}\mathbf{i}\xi_2$, where ξ_1, ξ_2 are iid copies of a real normalized random variable.

Example 1.1. *Some popular normalized variables*

- *real standard Gaussian* $\mathbf{g}_R = \mathbf{N}(0, 1)$, *or real Bernoulli* $\mathbf{b}_R$ *which takes value* ± 1 *with probability* $1/2$;

- *complex standard Gaussian* $\mathbf{g_C} = \frac{1}{\sqrt{2}}\mathbf{g}_{1,\mathbf{R}} + \frac{1}{\sqrt{2}}\mathbf{i}\mathbf{g}_{2,\mathbf{R}}$, *or complex Bernoulli* $\mathbf{b_C} = \frac{1}{\sqrt{2}}\mathbf{b}_{1,\mathbf{R}} + \frac{1}{\sqrt{2}}\mathbf{i}\mathbf{b}_{2,\mathbf{R}}$.

Fixed a normalized random variable ξ and consider the random vector $\mathbf{v} = (\xi_1, \ldots, \xi_n)$, whose entries are iid copies of ξ. Sample $n-1$ iid copies $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ of $\mathbf{v}$. We would like to study the normal vector of the hyperplane spanned by the $\mathbf{v}_i$.

In matrix term, we let $A = (a_{ij})_{1 \leq i \leq n-1, 1 \leq j \leq n}$ be a random matrix of size $n-1$ by n where the entries a_{ij} are iid copies of ξ; the $\mathbf{v}_i$ are the row vectors of A. Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{F}^n$ be a unit vector that is orthogonal to the $\mathbf{v}_i$ (Here and later $\mathbf{F}$ is either $\mathbf{R}$ or $\mathbf{C}$, depending on the support of ξ.) First note that recent studies in the singularity probability of random non-Hermitian matrices (see for instance [6, 22]) show that under very general conditions on ξ, with extremely high probability A has rank $n-1$. In this case $\mathbf{x}$ is uniquely determined up to the sign ± 1 when $\mathbf{F} = \mathbf{R}$ or by a uniformly chosen rotation $\exp(\mathbf{i}\theta)$ when $\mathbf{F} = \mathbf{C}$. Throughout the paper, we use asymptotic notation under the assumption that n tends to

infinity. In particular, $X = O(Y)$, $X \ll Y$, or $Y \gg X$ means that $|X| \leq CY$ for some fixed C.

When the entries of A are iid standard gaussian $\mathbf{g_F}$, it is not hard to see that $\mathbf{x}$ is distributed as a random unit vector sampled according to the Haar measure in S^{n-1} of $\mathbf{F}^n$. One then deduces the following properties (see for instance [21][Section 2])

Theorem 1.2 (Random gaussian vector)**.** *Let* $\mathbf{x}$ *be a random vector uniformly distributed on the unit sphere* S^{n-1}. *Then,*

- *(joint distribution of the coordinates)* $\mathbf{x}$ *can be represented as*

$$\mathbf{x} := (\frac{\xi_1}{S}, \ldots, \frac{\xi_n}{S}) \tag{1}$$

   where ξ_i *are iid standard gaussian* $\mathbf{g_F}$, *and* $S = \sqrt{\sum_{i=1}^n |\xi_i|^2}$;
- *(inner product with a fixed vector) for any fixed vector* $\mathbf{u}$ *on the unit sphere,*

$$\sqrt{n}\mathbf{x}^*\mathbf{u} \xrightarrow{d} \mathbf{g_F}; \tag{2}$$

- *(the largest coordinate) for any* $C > 0$, *with probability at least* $1 - n^{-C}$

$$\|\mathbf{x}\|_\infty \leq \sqrt{\frac{8(C+1)^3 \log n}{n}}; \tag{3}$$

- *(the smallest coordinate) for* $n \geq 2$, *any* $0 \leq c < 1$, *and any* $a > 1$,

$$\|\mathbf{x}\|_{\min} = \min\{|x_1|, \ldots, |x_n|\} \geq \frac{c}{a}\frac{1}{n^{3/2}} \tag{4}$$

   with probability at least $\exp(-2c) - \exp\left(-\frac{a^2 - \sqrt{2a^2 - 1}}{2}n\right)$.

Motivated by the *universality phenomenon* (see, for instance [34]), it is natural to ask if these properties are universal, namely that they hold if ξ is non-gaussian. Our result confirms this prediction in a strong sense. They also have applications in the theory of random matrices, which we will discuss after stating the main result.

Let us introduce some notations. We say that ξ is sub-gaussian if there exists a parameter $K_0 > 1$ such that for all t

$$\mathbf{P}(|\xi| \geq t) = O(\exp(-\frac{t^2}{K_0})). \tag{5}$$

Definition 1.3 (Frequent events)**.** Let $\mathcal{E}$ be an event depending on n (which is assumed to be sufficiently large).

- $\mathcal{E}$ holds asymptotically almost surely if $\mathbf{P}(\mathcal{E}) = 1 - o(1)$.
- $\mathcal{E}$ holds with high probability if there exists a positive constant δ such that $\mathbf{P}(\mathcal{E}) \geq 1 - n^{-\delta}$.
- $\mathcal{E}$ holds with overwhelming probability, and write $P(\mathcal{E}) = 1 - n^{-\omega(1)}$, if for any $K > 0$, with sufficiently large n $\mathbf{P}(\mathcal{E}) \geq 1 - n^{-K}$.

Theorem 1.4 (Main result). *Suppose that a_{ij} are iid copies of a normalized sub-gaussian random variable ξ, then the followings hold.*

- *(the largest coordinate) There are constants $C, C_1 > 0$ such that for any $m \geq C_1 \log n$*

$$\mathbf{P}(\|\mathbf{x}\|_\infty \geq \sqrt{m/n}) \leq Cn^2 \exp(-m/C). \qquad (6)$$

  In particularly, with overwhelming probability

$$\|\mathbf{x}\|_\infty = O\left(\sqrt{\frac{\log n}{n}}\right).$$

- *(the smallest coordinate) with high probability*

$$\|\mathbf{x}\|_{\min} \geq \frac{1}{n^{3/2} \log^{O(1)} n}. \qquad (7)$$

- *(joint distribution of the coordinates) There exists a positive constant c such that the following holds: for any d-tuple $(i_1, \ldots, i_m)$, with $d = n^c$, the joint law of the tuple $(\sqrt{n} x_{i_1}, \ldots, \sqrt{n} x_{i_d})$ is asymptotically independent standard normal. More precisely, there exists a positive constant c' such that for any measurable set $\Omega \in \mathbf{F}^d$,*

$$|\mathbf{P}((\sqrt{n} x_{i_1}, \ldots, \sqrt{n} x_{i_d}) \in \Omega) - \mathbf{P}(\mathbf{g}_{\mathbf{F},1}, \ldots, \mathbf{g}_{\mathbf{F},d}) \in \Omega)| \leq d^{-c'}, \qquad (8)$$

  where $\mathbf{g}_{\mathbf{F},1}, \ldots, \mathbf{g}_{\mathbf{F},d}$ are iid standard gaussian.
- *(inner product with a fixed vector) Assume furthermore that ξ is symmetric, then for any fixed vector $\mathbf{u}$ on the unit sphere,*

$$\sqrt{n} \mathbf{x}^* \mathbf{u} \xrightarrow{d} \mathbf{g}_{\mathbf{F}}. \qquad (9)$$

It also follows easily from (6) and (8) that with high probability $\|\mathbf{x}\|_\infty = \Theta\left(\sqrt{\frac{\log n}{n}}\right)$. Indeed, it is clear that with high probability, with $m = n^c$ for some sufficiently small c, $\max\{|\mathbf{g}_{\mathbf{F},1}|, \ldots, |\mathbf{g}_{\mathbf{F},m}|\} \gg \sqrt{\log m} = \sqrt{c \log n}$. Thus by (8), with high probability $\max\{|x_1|, \ldots, |x_m|\} \gg \sqrt{\frac{\log n}{n}}$.

Our approach can be extended to unit vectors orthogonal to the rows of an iid matrices A of size $(n - k) \times n$, for any fixed k or even k grows slowly with n; the details will appear in a later paper.

As random hyperplanes appear frequently in various areas, including random matrix theory, high dimensional geometry, statistics, and theoretical computer science, we expect that Theorem 1.4 will be useful. For the rest of this section, we discuss two applications.

1.5. **Tail bound for the least singular value of a random iid matrix.** Given an $n \times n$ random matrix $M_n(\xi)$ with entries being iid copies of a normalized variable ξ. Let $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ be its singular values. The two extremal σ_1 and σ_n are of special interest, and was studied by Goldstein and von Neumann, as they tried to analyze the running time of solving a system of random equations $M_n x = b$.

In [17], Goldstein and von Neumann speculated that σ_n is of order $n^{-1/2}$, which turned out to be correct. In particular, $\sqrt{n}\sigma_n$ tends to a limiting distribution, which was computed explicitly by Edelman in [8] in the gaussian case.

Theorem 1.6. *For any $t \geq 0$ we have*

$$\mathbf{P}(\sigma_n(M_{\mathbf{gR}}) \leq tn^{-1/2}) = \int_0^t \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-x/2+\sqrt{x}} dx + o(1)$$

as well as

$$\mathbf{P}(\sigma_n(M_{\mathbf{gC}}) \leq tn^{-1/2}) = \int_0^t e^{-x} dx.$$

In other words, $\mathbf{P}(\sigma_n(M_{\mathbf{gR}}) \leq tn^{-1/2}) = 1 - e^{-t/2+\sqrt{t}} + o(1)$ and $\mathbf{P}(\sigma_n(M_{\mathbf{gC}}) \leq tn^{-1/2}) = 1 - e^{-t}$. These distributions have been confirmed to be universal (in the asymptotic sense) by Tao and the second author [31].

In applications, one usually needs large deviation results, which show that the probability that σ_n is far from its mean is very small. For the lower bound, Rudelson and Vershyin [22] proved that for any $t > 0$

$$\mathbf{P}(\sigma_n \leq tn^{-1/2}) \leq Ct + .999^n, \tag{10}$$

which is sharp up to the constant C. For the upper bound, in a different paper [24], the same authors showed

$$\mathbf{P}(\sigma_n \geq tn^{-1/2}) \leq C \frac{\log t}{t}. \tag{11}$$

Using Theorem 1.4, we improve this result significantly by proving an exponential tail bound,

Theorem 1.7 (Exponential upper tail for the least singular values). *Assume that the entries of $M_n = (m_{ij})_{1 \leq i,j \leq n}$ are iid copies of a normalized subgaussian random variable ξ in either $\mathbf{R}$ or $\mathbf{C}$. Then there exist absolute constants C_1, C_2 depending on K_0 such that*

$$\mathbf{P}(\sigma_n \geq tn^{-1/2}) \leq C_1 \exp(-C_2 t).$$

Our proof of Theorem 1.7 is totally different from that of [24]. As showed in the gaussian case, the exponential bound is sharp, up to the value of C_2.

1.8. **Eigenvectors of random iid matrices.** Our theorem is closely related to (and in fact was motivated by) recent results concerning delocalization and normality of eigenvectors of random matrices. For random Hermitian matrices, there have been many results achieving almost optimal delocalization of eigenvectors, starting with the work [16] by Erdős et al.

and and continued by Tao et al. and by many others in [32, 36, 9, 10, 11, 12, 13, 35, 2, 3, 4]. Thanks to new universality techniques, one also proved normality of the eigenvectors; see for instance the work [19] by Knowles and Yin, [33] by Tao and Vu, and [5] by Bourgade and Yau.

For non-Hermitian random matrix $M_n(\xi) = (m_{ij})_{1 \leq i,j \leq n}$, much less is known. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues with $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the corresponding unit eigenvectors (where $\mathbf{v}_i$ are chosen according to the Haar measure from the eigensphere if the corresponding roots are multiple). Recently, Rudelson and Vershynin [26] proved that with overwhelming probability all of the eigenvectors satisfy

$$\|\mathbf{v}_i\|_\infty = O(\frac{\log^{9/2} n}{\sqrt{n}}). \tag{12}$$

By modifying the proof of Theorem 1.4, we are able sharpen this bound for eigenvectors of eigenvalues with small modulus.

Theorem 1.9 (Optimal delocalization for small eigenvectors). *Assume that the entries of $M_n = (m_{ij})_{1 \leq i,j \leq n}$ are iid copies of a normalized subgaussian random variable ξ in either $\mathbf{R}$ or $\mathbf{C}$. Then for any fixed $\varepsilon > 0$, with overwhelming probability the following holds for any unit eigenvector $\mathbf{x}$ corresponding to an eigenvalue λ of A with $|\lambda| = O(1)$*

$$\|\mathbf{x}\|_\infty = O(\sqrt{\frac{\log n}{n}}).$$

We believe that the individual eigenvector in Theorem 1.9 satisfies the normality property (8), which would imply that the bound $O(\sqrt{\frac{\log n}{n}})$ is optimal up to a multiplicative constant. Figure 1 below shows that the first coordinate of the eigenvector corresponding to the smallest eigenvalue behaves like a gaussian random variable.

Finally, let us mention that all of our results holds (with logarithmic correction) under a weaker assumption that the variable ξ is sub-exponential, namely there are positive constants C, C' and α such that for all t $\mathbf{P}(|\xi| \geq t) \leq C \exp(-C't^\alpha)$; see Remark 2.3.

The rest of the paper is organized as follows. After introducing supporting lemmas in Section 2, we will prove (6) and Theorem 1.9 in Section 3. Section 6 and Section 7 are devoted to proving (8) and (9) correspondingly, while (7) will be shown in Section 4. Finally, we prove Theorem 1.7 in Section 5.

2. THE LEMMAS

We will use the following well-known concentration result of distances in random non-Hermitian matrices (see for instance [32, Lemma 43], [28, Corollary 2.19] or [36]).
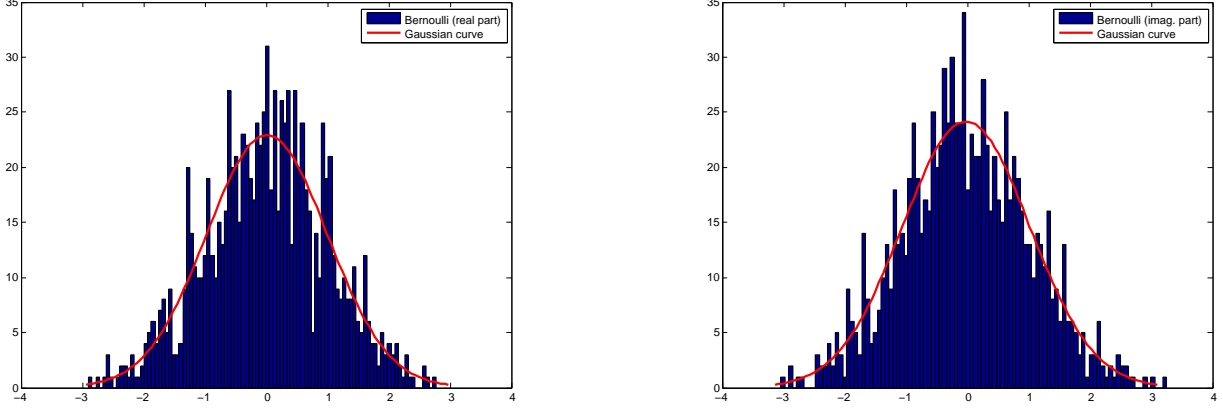
FIGURE 1. We sampled 1000 random complex iid Bernoulli matrices of size $n = 500$. The histograms represent the normalized real and imaginary parts $\sqrt{2n}\mathrm{Re}(\mathbf{v}(1))$ and $\sqrt{2n}\mathrm{Im}(\mathbf{v}(1))$ of the first coordinate of the unit eigenvector $\mathbf{v}$ associated with the eigenvalue of smallest modulus.

Lemma 2.1. *Let H be a subspace of co-dimension m in $\mathbf{F}^l$ and let P_H be the projection matrix onto the complement $H^\perp$ of H. Let $\mathbf{u} = (u_1, \ldots, u_l)$ and $\mathbf{v} = (v_1, \ldots, v_l)$ be independent random vectors where u_i, v_i are iid copies of an $\mathbf{F}$-normalized sub-gaussian random variable ξ. Then the following holds.*

(1) the distance from $\mathbf{u}$ to H is well concentrated around its mean,

$$\mathbf{P}\left(|\|P_H\mathbf{u}\|_2 - \sqrt{m}| \geq t\right) \leq \exp(-t^2/K_0^4);$$

(2) the correlation $\mathbf{v}^T P_H \mathbf{u}$ is small,

$$\mathbf{P}\left(|\mathbf{v}^T P_H \mathbf{u}| \geq t\right) \leq \exp(-t^2/K_0^4).$$

More generally, we have

Lemma 2.2 (Hanson-Wright inequality). *There exists an absolute constant c such that the following holds for any sub-gaussian $\mathbf{F}$-normalized random variable ξ. Let A be a fixed $l \times l$ Hermitian matrix. Consider a random vector $\mathbf{x} = (x_1, \ldots, x_l)$ where the entries are iid copies of ξ. Then*

$$\mathbf{P}(|\mathbf{x}^*A\mathbf{x} - \mathbf{E}\mathbf{x}^*A\mathbf{x}| > t) \leq 2\exp\left(-c\min(\frac{t^2}{K_0^4\|A\|_{HS}^2}, \frac{t}{K_0^2\|A\|_2})\right).$$

In particularly, for any $t > 0$

$$\mathbf{P}\left(|\mathbf{x}^*A\mathbf{x} - \mathbf{E}\mathbf{x}^*A\mathbf{x}| > t\|A\|_{HS}\right) \leq O\left(\exp(-c\frac{t^2}{K_0^4}) + \exp(-c\frac{t}{K_0^2})\right).$$

This lemma was first proved by Hanson and Wright in a special case [18]. The above general version is due to Rudelson and Vershynin [25]; see also [36] for related results which hold (with logarithmic correction) for sub-exponential variables.

Remark 2.3. *As mentioned at the end of the introduction, the results of this paper hold (with logarithmic correction) for sub-exponential variables. One can achieve this by repeating the proofs, using the results from [36] (such as [36, Corollary 1.6]) instead of Lemmas 2.1 and 2.2. We leave the details as an exercise.*

The next tool is Berry-Esséen theorem for frames, proved by Tao and Vu in [31]. As the statement is technical, let us first warm the reader up by the classical Berry-Esséen theorem.

Lemma 2.4 (Berry-Esséen theorem). *Let $v_1, \ldots, v_l \in \mathbf{F}$ be real numbers with $\sum_i |v_i|^2 = 1$ and let ξ be a $\mathbf{F}$-normalized random variable with finite third moment $\mathbf{E}|\xi|^3 < \infty$. Let S denote the random sum*

$$S = \sum_i v_i \xi_i,$$

where ξ_i are iid copies of ξ. The for any $t \in \mathbf{F}$ we have

$$\mathbf{P}(|S| \leq t) = \mathbf{P}(|\mathbf{g_F}| \leq t) + O(\sum_i |v_i|^3),$$

where the implied constant depends on the third moment of ξ. In particularly,

$$\mathbf{P}(|S| \leq t) = \mathbf{P}(|\mathbf{g_F}| \leq t) + O(\max_i |v_i|).$$

Lemma 2.5 (Berry-Esséen theorem for frames). [31, Proposition D.2] *Let $1 \leq k \leq l$, and let ξ be an $\mathbf{F}$-normalized and have finite third moment. Let $v_1, \ldots, v_l \in \mathbf{F}^k$ be a normalized tight frame for $\mathbf{F}^k$, in other words*

$$v_1 v_1^* + \cdots + v_n v_n^* = I_k,$$

where I_k is the identity matrix on $\mathbf{F}^k$. Let $S \in \mathbf{F}^k$ denote the random variable

$$S = \xi_1 v_1 + \cdots + \xi_n v_n,$$

where $\xi_1, \ldots, \xi_n$ are iid copies of ξ. Similarly, let $G := (\mathbf{g}_1, \ldots, \mathbf{g}_k) \in F^k$ be formed from k iid copies of the standard gaussian random variable $\mathbf{g_F}$. Then for any measurable $\Omega \subset F^k$ and for any $\varepsilon = \varepsilon(k, n) > 0$ we have

$$\mathbf{P}\Big(G \in \Omega/\partial_\varepsilon \Omega\Big) - O\Big(k^{5/2}\varepsilon^{-3} \max_j \|v_j\|_\infty\Big) \le P(S \in \Omega) \le \mathbf{P}\Big(G \in \Omega \cup \partial_\varepsilon \Omega\Big) + O\Big(k^{5/2}\varepsilon^{-3} \max_j \|v_j\|_\infty\Big),$$

where $\partial_\varepsilon \Omega$ is the collection of $\mathbf{x} \in \mathbf{F}^k$ such that $\mathrm{dist}_{\|.\|_\infty}(\mathbf{x}, \partial\Omega) \le \varepsilon$.

3. TREATMENT FOR THE LARGEST COORDINATE: PROOF OF (6) AND THEOREM 1.9

3.1. **Proof of** (6).

By a union bound, it suffices to show that for sufficiently large C

$$\mathbf{P}\Big(|x_1| \ll \sqrt{\frac{m}{n}}\Big) = 1 - O\Big(n \exp(-\frac{m}{C})\Big). \tag{13}$$

Let $\mathbf{c}_i$, $1 \le i \le n$ be the columns of A. Because $\sum_{i=2}^n |x_i|^2 \le 1$, among the $(n-1)/m$ subset sums $|x_2|^2 + \cdots + |x_m|^2, |x_{m+1}|^2 + \cdots + |x_{2m-1}|^2, \ldots, |x_{n-m+2}|^2 + \cdots + |x_n|^2$, there is a subset sum which is smaller than m/n. With a loss of a factor n/m in probability, without loss of generality we will assume that

$$|x_2|^2 + \cdots + |x_m|^2 \le \frac{m}{n-1}.$$

Let H be the subspace generated by $c_j, j \ge m + 1$. Let P_H be the orthogonal projection from $\mathbf{F}^{n-1}$ onto $H^\perp$. We view P_H as a Hermitian matrix of size $(n-1) \times (n-1)$ satisfying $P_H^2 = P_H$. It is known (see for instance [22, 31, 6]) that with probability $1 - \exp(-cn)$ we have $\dim(H^\perp) = m - 1$, which implies $\mathrm{tr}(P_H) = m - 1$.

Recall that by definition,

$$x_1 \mathbf{c}_1 + x_2 \mathbf{c}_2 + \cdots + x_m \mathbf{c}_m + \sum_{i \ge m+1} x_i \mathbf{c}_i = 0. \tag{14}$$

Applying P_H, we have

$$x_1 P_H \mathbf{c}_1 = -P_H(x_2 \mathbf{c}_2 + \cdots + x_m \mathbf{c}_m),$$

which implies

$$|x_1|^2 \|P_H \mathbf{c}_1\|_2^2 = \| \sum_{j=2}^{m} x_j P_H \mathbf{c}_j \|_2^2 = \sum_{2 \le j_1 \le j_2 \le m} x_{j_1} \bar{x}_{j_2} \mathbf{c}_{j_1}^T P_H \mathbf{c}_{j_2} := \|Q\mathbf{x}'\|_2^2, \qquad (15)$$

where $\mathbf{x}' = (x_2, \ldots, x_m)$ and $Q\mathbf{x}' := \sum_{j=2}^{m} x_j P_H \mathbf{c}_j$. We remark that the x_i here are not deterministic but depend on the column vectors $\mathbf{c}_i$.

As $Q\mathbf{x}'$ is linear, and as $|x_2|^2 + \cdots + |x_m|^2 \le m/(n-1)$, we have

$$\|Q\mathbf{x}'\|_2 \le \sup_{\mathbf{y} \in \mathbf{F}^{m-1}, \|\mathbf{y}\|=1} \|Q\mathbf{y}\|_2 \sqrt{\frac{m}{n-1}}.$$

Thus

$$|x_1|^2 \|P_H \mathbf{c}_1\|_2^2 \le \sup_{\mathbf{y} \in \mathbf{F}^{m-1}, \|\mathbf{y}\|=1} \|Q\mathbf{y}\|_2^2 \frac{m}{n-1}. \qquad (16)$$

We are going to estimate the operator norm $\|Q\|_2$ basing the randomness of $\mathbf{c}_j, 2 \le j \le m$.

Lemma 3.2. *There exists a sufficiently large constant C such that*
$$\mathbf{P}_{\mathbf{c}_2, \ldots, \mathbf{c}_m}(\|Q\|_2^2 \ge Cm) = O(\exp(-2(m-1))).$$

Assume Lemma 3.2 for the moment, we can complete the proof of (13) as follows. First, by Lemma 2.1, $\|P_H \mathbf{c}_1\|_2^2 \ge m/2$ with probability at least $1 - \exp(-\frac{m}{4K_0^4})$. We then deduce from (16) and from Lemma 3.2 that

$$\mathbf{P}(|x_1|^2 \gg \frac{m}{n}) \le O\Big(\frac{n}{m} \exp(-\frac{m-1}{4K_0^4}) + \exp(-2(m-1)) \Big),$$

completing the proof.

To prove Lemma 3.2, we first estimate $\|Q\mathbf{y}\|_2$ for any fixed $\mathbf{y} \in S^{m-2}$. We will show

Lemma 3.3. *There exists a sufficiently large constant C such that for any fixed $\mathbf{y} \in \mathbf{F}^{m-1}$ with $\|\mathbf{y}\|_2 = 1$,*
$$\mathbf{P}_{\mathbf{c}_2, \ldots, \mathbf{c}_m}(\|Q\mathbf{y}\|_2^2 \ge Cm) = O(\exp(-4(m-1))).$$

The deduction of Lemma 3.2 from Lemma 3.3 is standard, we present it here for the sake of completeness.

Proof. (of Lemma 3.2) Let $\mathcal{N}$ be a $(1/2)$-net for the set of unit vectors in $\mathbf{F}^{m-1}$. As is well known, one can assume that $|\mathcal{N}| \le 4^m$. Applying Lemma 3.3,

$$\mathbf{P}\Big(\exists \mathbf{y} \in \mathcal{N}, \|Q\mathbf{y}\|_2^2 \ge 2m \Big) = O\Big(|\mathcal{N}| \exp(-4(m-1)) \Big) = O\Big(\exp(-2(m-1)) \Big).$$

Now for any unit vector $\mathbf{y}'$, there exists $\mathbf{y} \in \mathcal{N}$ such that $\|\mathbf{y}' - \mathbf{y}\|_2 \leq 1/2$, and thus by the triangle inequality

$$\|Q\mathbf{y}'\|_2 \leq \|Q\mathbf{y}\|_2 + \|Q(\mathbf{y} - \mathbf{y}')\|_2 \leq \|Q\mathbf{y}\|_2 + \|Q\|_2/2.$$

This implies that $\|Q\|_2 \leq \sup_{\mathbf{y} \in \mathcal{N}} \|Q\mathbf{y}\|_2 + \|Q\|_2/2$, and hence

$$\|Q\|_2 \leq 2 \sup_{\mathbf{y} \in \mathcal{N}} \|Q\mathbf{y}\|_2.$$

$\square$

Proof. (of lemma 3.3) Let $\mathbf{c}$ be the concatenation of $(\mathbf{c}_{i_1}, \ldots, \mathbf{c}_{i_{m-1}})$, then $\|Q\mathbf{y}\|_2^2$ can be written as a bilinear form $S = \mathbf{c}^* P \mathbf{c}$ where P is the tensor product of $\mathbf{y}\mathbf{y}^*$ and P_H, with $\mathbf{y} = (y_1, \ldots, y_{m-1})$. By construction, P consists of $(m-1)^2$ blocks where the kl-th block is the matrix $y_k \bar{y}_l P_H$. It thus follows that

$$\|P\|_2 = \|\mathbf{y}\|_2^2 = 1.$$

Applying Lemma 2.2 to $S = \mathbf{c}^* P \mathbf{c}$, we have

$$\mathbf{P}(|S - \mathrm{tr}P| \geq t) \leq O\Big(\exp(-c\frac{t^2}{K_0^4 \|P\|_{HS}^2}) + \exp(-c\frac{t}{K_0^2 \|P\|_2}) \Big).$$

It is easy to show that

$$\mathrm{tr}P = (m-1) \sum_{j=0}^{m-1} |y_j|^2 = m - 1.$$

Taking $t = 4(c^{-1} + 1)K_0^2(m-1) := \alpha(m-1)$, we obtain

$$\mathbf{P}\Big(S \geq (\alpha+1)(m-1) \Big) \leq O\Big(\exp(-16\frac{(m-1)^2}{\|P\|_{HS}^2}) + \exp(-4(m-1)) \Big).$$

To this end, by properties of a tensor product,

$$\|P\|_{HS}^2 = \|\mathbf{y}\mathbf{y}^T\|_{HS}^2 \|P_H\|_{HS}^2 = m - 1,$$

which implies that

$$\mathbf{P}\Big(S \geq (\alpha+1)(m-1) \Big) = O\Big(\exp(-4(m-1)) \Big). \tag{17}$$

$\square$

We now turn to the eigenvectors.

3.4. **Proof of Theorem 1.9.** We will be working with the perturbed matrix $M_n - \lambda_0$ where $(M_n - \lambda_0)_{ii} = m_{ii} - \lambda_0, 1 \leq i \leq n$ and $(M_n - \lambda_0)_{ij} = m_{ij}, i \neq j$. By a standard net argument, it suffices to show the following

Theorem 3.5. *For any fixed λ_0 with $|\lambda_0| \leq O(1)$, the following holds with overwhelming probability with respect to M_n: if $\|(M_n - \lambda_0)\mathbf{x}\|_2 \leq 1/n$ then $\mathbf{x}$ satisfies (6).*

Equivalently, we show that for any unit vector $\mathbf{x} \in \mathbf{F}^n$ satisfying the condition of Theorem 3.5, then

$$\mathbf{P}(|x_1| \ll \sqrt{\frac{m}{n}}) = 1 - O\Big(n \exp(-\frac{m}{C})\Big). \tag{18}$$

We will proceed as in Subsection 3.1 by assuming that $|x_2|^2 + \cdots + |x_m|^2 \leq m/(n-1)$, where instead of (14) we have

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_m\mathbf{c}_m + \sum_{i \geq m+1\}} x_i\mathbf{c}_i = \mathbf{r} \tag{19}$$

for some vector $\mathbf{r}$ with norm $\|\mathbf{r}\|_2 \leq 1/n$, where $\mathbf{c}_i$ is the i-th column of the matrix $M_n - \lambda_0$.

Projecting onto $H^\perp$, we obtain

$$|x_1|^2\|P_H\mathbf{c}_1\|_2^2 \leq 2\|\sum_{j=2}^m x_j P_H\mathbf{c}_j\|_2^2 + 2\|\mathbf{r}\|_2^2 \leq 2\sum_{2 \leq j_1 \leq j_2 \leq m} x_{j_1}\bar{x}_{j_2}\mathbf{c}_{j_1}^* P_H\mathbf{c}_{j_2} + \frac{2}{n^2}.$$

Note that here as $|\lambda_0| = O(1)$, Lemma 2.1 is still effective, which yields $\|P_H\mathbf{c}_1\|_2^2 \geq m/2$ with probability at least $1 - \exp(-\frac{m}{4K_0^4})$.

To estimate the right hand side, set $Q(\mathbf{x}') := \sum_{j=2}^m x_j P_H\mathbf{c}_j$. Similarly to Lemma 3.2, we will establish

Lemma 3.6. *There exists a sufficiently large constant C such that*

$$\mathbf{P}_{\mathbf{c}_2,\ldots,\mathbf{c}_m}(\|Q\|_2^2 \geq Cm) = O(\exp(-2(m-1))).$$

It is clear that (18) follows from Lemma 3.6. Furthermore, similarly to our treatment in the previous subsection, for this lemma it suffices to show the following analog of Lemma 3.3 for any fixed $\mathbf{y}$.

Lemma 3.7. *There exists a sufficiently large constant C such that for any fixed $\mathbf{y} \in \mathbf{F}^{m-1}$ with $\|\mathbf{y}\|_2 = 1$,*

$$\mathbf{P}_{\mathbf{c}_2,\ldots,\mathbf{c}_m}(\|Q\mathbf{y}\|_2^2 \geq Cm) = O(\exp(-4(m-1))).$$

It remains to prove Lemma 3.7. Write $\mathbf{c}_j = \mathbf{c}'_j - \lambda_0 \mathbf{f}_j$, where $\mathbf{f}_j$ is a $\{0,1\}$-vector with at most one non-zero entry and $\mathbf{c}'_j$ is a random vector of iid entries. Thus

$$
\sum_{1 \leq i,j \leq m-1} y_i \bar{y}_j \mathbf{c}_i^* P_H \mathbf{c}_j = \sum_{1 \leq i,j \leq m-1} y_i \bar{y}_j \mathbf{c}_i'^* P_H \mathbf{c}'_j
$$
$$
+ \lambda_0 \sum_{1 \leq i,j \leq m-1} y_i \bar{y}_j \mathbf{c}_i'^* P_H \mathbf{f}_j
$$
$$
+ \lambda_0 \sum_{1 \leq i,j \leq m-1} y_i \bar{y}_j \mathbf{f}_i^* P_H \mathbf{c}'_j
$$
$$
+ |\lambda_0|^2 \sum_{1 \leq i,j \leq m-1} y_i \bar{y}_j \mathbf{f}_i^* P_H \mathbf{f}_j
$$
$$
:= S + S' + S'' + S'''.
$$

For S, argue similarly as in the proof of Lemma 3.3, we obtain the following analog of (17)

$$\mathbf{P}\Big(S \geq (\alpha + 1)(m - 1)\Big) = O(\exp(-4(m - 1))).$$

Next, we have

$$
|S'| = |\lambda_0 \sum_{1 \leq i,j \leq m} y_i \bar{y}_j \mathbf{c}_i'^* P_H \mathbf{f}_j| = |\lambda_0 (\sum_{1 \leq i \leq m-1} y_i \mathbf{c}_i'^*)(\sum_{1 \leq j \leq m-1} y_j P_H \mathbf{f}_j)|
$$
$$
= |\lambda_0 (\sum_{1 \leq i \leq m-1} y_i \mathbf{c}_i'^*) P_H (\sum_{1 \leq j \leq m-1} \bar{y}_j \mathbf{f}_j)|,
$$

Additionally, as $\|P_H\|_2 \leq 1$ and $\|\mathbf{y}\|_2 = 1$, by the properties of $\mathbf{f}_i$ the vector $\mathbf{z} := P_H(\sum_{1 \leq j \leq m-1} \bar{y}_j \mathbf{f}_j)$ has norm at most $\|\mathbf{z}\|_2 \leq 1$. As such, the subgaussian random variable $(\sum_{1 \leq i \leq m-1} y_i \mathbf{c}_i'^*)\mathbf{z}$ has variance at most one, and hence

$$\mathbf{P}\Big(|(\sum_{1 \leq i \leq m-1} y_i \mathbf{c}_i'^*)\mathbf{z}| \geq m - 1\Big) = o\Big(\exp(-4(m - 1))\Big).$$

We can argue similarly for S'' to obtain the same bound. Finally, notice that

$$|S'''| = |\lambda_0|^2 \|P_H(\sum_{1 \leq j \leq m-1} y_j \mathbf{f}_j)\|_2^2 \leq |\lambda_0|^2.$$

Putting all the estimates together, we obtain Lemma 3.7 as long as $|\lambda_0| = O(1)$.

4. Treatment for the smallest coordinate: proof of (7)

Let M be the random matrix of size $(n-1) \times (n-1)$ obtained from A by deleting its first column. Set $\mathbf{x}' = (x_2, \ldots, x_n)$, we have

$$A\mathbf{x} = x_1 \mathbf{c}_1 + M\mathbf{x}' = 0.$$

As it is known that with probability at least $1 - \exp(-cn)$ the matrix M is invertible; in this case, we can write

$$x_1 M^{-1} \mathbf{c}_1 = -\mathbf{x}'.$$

Since

$$|x_1|^2 \|M^{-1}\mathbf{c}_1\|_2^2 = \|\mathbf{x}'\|_2^2 = 1 - |x_1|^2,$$

we obtain

$$|x_1|^2 = \frac{1}{1 + \|M^{-1}\mathbf{c}_1\|_2^2} = \frac{1}{1 + \sum_{j=1}^{n-1} \sigma_j^{-2} |\mathbf{c}_1^T \mathbf{u}_j|^2},$$

where $\sigma_1 \geq \cdots \geq \sigma_{n-1}$ are the singular values of M with corresponding left-singular vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$.

We now condition on M. By the sub-gaussian property of the entries, we can easily show that there is a constant C such that with overwhelming probability (with respect to $\mathbf{c}_1$)

$$|\mathbf{c}_1^T \mathbf{u}_1| \leq C \log n \wedge \cdots \wedge |\mathbf{c}_1^T \mathbf{u}_{n-1}| \leq C \log n. \tag{20}$$

We will need the following estimate

Claim 4.1. *With respect to M we have*

$$\mathbf{P}(\sum_{i=1}^{n-1} \sigma_i^{-2} \leq n^3 \log^8 n) \geq 1 - \frac{1}{n \log n}.$$

Proof. (of Claim 4.1) By (10)

$$\mathbf{P}(\sigma_{n-1}^{-1} \leq n^{3/2} \log^3 n) \geq 1 - \frac{1}{n \log^3 n}.$$

Thus by the union bound

$$\mathbf{P}\Big(\sum_{i=1}^{\log^2 n} \sigma_{n-i}^{-2} \leq n^3 \log^8 n\Big) \geq 1 - \frac{1}{n \log n}.$$

For the remaining sum $\sum_{j=1}^{n-\log n-1} \sigma_j^{-2}$, by the Cauchy-interlacing law,

$$\sum_{j=1}^{n-\log^2 n-1} \sigma_j^{-2}(M) \leq \sum_{j=1}^{n-\log n-1} \sigma_j^{-2}(M'),$$

where M' is obtained from M by deleting its first $\log^2 n$ columns.

On the other hand, by the negative second moment identity (see [30, Lemma A.4])

$$\sum_{j=1}^{n-\log^2 n} \sigma_j^{-2}(M') = \sum_{j=1}^{n-\log^2 n} d_j^{-2}, \tag{21}$$

where d_j is the distance from the jth row of M' to the hyperplane H_j spanned by the remaining rows of M'. Using Theorem 2.1 and the union bound, we obtain, for some constant c and with overwhelming probability, that $d_j \geq c \log n$ simultaneously for all $1 \leq j \leq n - \log^2 n$. This implies that with overwhelming probability with respect to M

$$\sum_{j=1}^{n-\log^2 n} \sigma_j^{-2} \ll \frac{n}{\log^2 n}.$$

$\square$

Now by (20) and Claim 4.1, we have

$$\mathbf{P}\Big(|x_1| \gg \frac{1}{n^3 \log^{10} n}\Big) \geq 1 - \frac{1}{n \log n}.$$

By the union by, we have with probability at least $1 - \frac{1}{\log n}$,

$$|x_1| \geq \frac{1}{n^3 \log^{10} n} \wedge \cdots \wedge |x_n| \geq \frac{1}{n^3 \log^{10} n},$$

proving the desired statement.

5. EXPONENTIAL UPPER TAIL BOUNDS: PROOF OF THEOREM 1.7

Using [31, Theorem 1.3] we can compare $\mathbf{P}(\sigma_n \geq tn^{-1/2})$ with $\mathbf{P}(\sigma_n(\mathbf{g_F}) \geq tn^{-1/2})$, where $\sigma_n(\mathbf{g_F})$ is the least singular value of an $\mathbf{F}$-normalized gaussian matrix. More precisely, it shows that there exists a positive constant c such that

$$\mathbf{P}(\sigma_n \geq tn^{-1/2}) \leq \mathbf{P}(\sigma_n(\mathbf{g_F}) \geq tn^{-1/2}) + n^{-c}.$$

In the complex case, Theorem 1.6 has $\mathbf{P}(\sigma_n(\mathbf{g_C}) \geq tn^{-1/2}) = \exp(-t)$. Since $n^{-c} = \exp(-c \log n)$, this implies the claim for $t \leq C \log n$ for any fixed C and properly chosen constants C_1, C_2.

In the real case, one cannot apply Theorem 1.6 directly because of the error term is just plainly $o(1)$. However, in [31] Tao and the second author proved that this error term is at most $n^{-c'}$ for some constant $c' > 0$. Thus, one can conclude in the same manner as in the complex case.

From here we assume $t > C \log n$, where C is a sufficiently large constant. By the proof of (6) of Theorem 1.4 (applied for matrices of size $n \times (n+1)$ instead of $(n-1) \times n$) we have, for all $m \geq C \log n$ that

$$\mathbf{P}(|x_1| \geq m^{1/2} n^{-1/2}) = O(\exp(-m)).$$

Equivalently, for all $t = m \geq C \log n$

$$\mathbf{P}(|x_1| \geq t^{1/2} n^{-1/2}) = O(\exp(-t)).$$

One the other hand, similarly to our treatment in Section 4

$$|x_1|^2 = \frac{1}{1 + \sum_{j=1}^n \sigma_j^{-2}(\mathbf{c}_1^T \mathbf{u}_j)^2},$$

where σ_j are the singular values of the random square matrix M_n formed by the last n columns, $\mathbf{c}_1$ is the first column, and $\mathbf{u}_j$ are the corresponding unit eigenvector of σ_j^2 in $M_n M_n^*$.

Thus with probability at least $1 - O(\exp(-t))$ we have

$$1 + \sum_{j=1}^n \sigma_j^{-2}(\mathbf{c}_1^T \mathbf{u}_j)^2 \geq \frac{n}{t}. \tag{22}$$

Next, again by following the argument in Section 4 (using the negative-moment identity (21), the Cauchy-interlacing law, and Theorem 2.1), we can prove

Claim 5.1. *With probability at least* $1 - n\exp(-\frac{t}{K_0^2})$ *one has*

$$\sum_{j=1}^{n-100t} \sigma_j^{-2} \leq \frac{n}{2t},$$

with K_0 *from* (5).

To handle the coefficients $|\mathbf{c}^T\mathbf{u}_j|$, we use the following concentration result from [36].

Lemma 5.2. [36, Lemma 1.2] *Let* $\mathbf{c} = (x_1, \ldots, x_n)$ *be a random vector where* x_i *are iid copies of* ξ. *Then there exists a constant* $C' > 0$ *such that the following holds. Let* H *be a subspace of dimension* d *with an orthonormal basis* $\{\mathbf{u}_1, \ldots, \mathbf{u}_d\}$. *Then for any* $0 \leq c_1, \ldots, c_d \leq 1$ *and any* s

$$\mathbf{P}\left(|\sqrt{\sum_{j=1}^{d} c_j|\mathbf{c}^T\mathbf{u}_j|^2} - \sqrt{\sum_{j=1}^{d} c_j}| \geq s\right) \leq 2\exp(-C'\frac{s^2}{K_0^4}).$$

Remark 5.3. *There is a strong relation between this lemma and Lemma 2.2. First, one can give a short proof of this lemma using Lemma 2.2. Second, one can also prove a generalization of Lemma 2.2 to sub-exponential variables (with logarithmic correction) using this lemma. See Remark 2.3.*

In particular, by squaring, it follows that

$$\mathbf{P}\left(|\sum_{j=1}^{d} c_j|\mathbf{c}^T\mathbf{u}_j|^2 - \sum_{j=1}^{d} c_j| \geq 2s\sqrt{\sum_{j=1}^{d} c_j} + s^2\right) \leq 2\exp(-C'\frac{s^2}{K_0^4}). \tag{23}$$

Next, Lemma 5.2, applied to $\sum_{j=1}^{n} \sigma_j^{-2}(\mathbf{c}_1^T\mathbf{u}_j)^2$ (with $c_j = \frac{\sigma_j^{-2}}{\sigma_n^{-2}}$ and $s = t^{1/2}$), implies that

$$\mathbf{P}\left(|\sum_{j=1}^{n} \sigma_j^{-2}(\mathbf{c}_1^T\mathbf{u}_j)^2 - \sum_{j=1}^{n-1} \sigma_j^{-2}| \geq 2t^{1/2}\sigma_{n-1}^{-1}\sqrt{\sum_{j=1}^{n} \sigma_j^{-2}} + t\sigma_n^{-2}\right) \leq 2\exp(-C'\frac{t}{K_0^4}).$$

Thus, with probability at least $1 - 2\exp(-C'\frac{t}{K_0^4})$, we have

$$\sum_{j=1}^{n} \sigma_j^{-2}(\mathbf{c}_1^T\mathbf{u}_j)^2 \leq \sum_{j=1}^{n} \sigma_j^{-2} + 2t^{1/2}\sigma_n^{-1}\sqrt{\sum_{j=1}^{n} \sigma_j^{-2}} + t\sigma_n^{-2}. \tag{24}$$

Now we can conclude from (22), (24) and Claim 5.1 (noting that $t \geq C\log^{3/2} n$) that with probability at least $1 - 2\exp(-C'\frac{t}{K_0^4})$

$$\frac{n}{t} \leq (\frac{n}{2t} + 100t\sigma_n^{-2}) + 2t^{1/2}\sigma_n^{-1}\sqrt{\frac{n}{2t} + 100t\sigma_n^{-2}} + t\sigma_n^{-2}.$$

This event guarantees that $\frac{n}{4t} \leq 100t\sigma_n^{-2}$, or equivalently $\sigma_n \leq 20\frac{t}{\sqrt{n}}$. Our proof is complete.

6. NORMALITY OF VECTORS: PROOF OF (8)

We will show that

$$|\mathbf{P}((\sqrt{n}x_1 \in \Omega) - \mathbf{P}(\mathbf{g}_{\mathbf{F},1} \in \Omega)| \leq n^{-c'}. \tag{25}$$

The general case with joint distribution of d components, with d chosen to be a small power of n, can be treated similarly; see also (36) below.

Our method follows that of [31]. First, by (6) of Theorem 1.4, it suffices to work with the event $\mathcal{E}$

$$|x_i| = O(\frac{\log^{1/2} n}{\sqrt{n}}), 1 \leq i \leq n. \tag{26}$$

We will need the following result (see for instance [1, Theorem 3.1]).

Theorem 6.1. *Assume that $\sum_{i=1}^n n|x_i|^2 = n$ and $n|x_i|^2 \leq L$ for all i. Then there exists an absolute constant c such that for a uniformly randomly chosen $(m-1)$-set $\{i_1, \ldots, i_{m-1}\}$ from the index set $\{2, \ldots, n\}$*

$$\mathbf{P}\Big(|n|x_{i_1}|^2 + \cdots + n|x_{i_{m-1}}|^2 - (m-1)| \geq t\Big) \leq 2\exp(-ct^2/L^2),$$

where the probability is with respect to $\{i_1, \ldots, i_{m-1}\}$.

For convenience, denote by $\mathcal{F}_{i_1,\ldots,i_{m-1}}$ the event

$$\Big|n|x_{i_1}|^2 + \cdots + n|x_{i_{m-1}}|^2 - (m-1)\Big| \leq \log^2 n. \tag{27}$$

By Theorem 6.1, with $L = O(\log^{1/2} n)$ and $t = \log^2 n$

$$\mathbf{P}(\mathcal{E} \wedge \mathcal{F}_{i_1,\ldots,i_{m-1}}) = 1 - n^{-\omega(1)}.$$

We are conditioning on these two events for the rest of the argument.

With foresight, we choose m slightly larger than the value in Section 3, in particularly m will take the form n^{1/C_0} for some sufficiently large constant C_0 to be chosen later. We next exploit (14) once more by projecting onto the orthogonal complement $H^\perp$ of H. This time we view the projection as $\pi_H : \mathbf{R}^{n-1} \to \mathbf{R}^{m-1}$,

$$x_1 \pi_H(\mathbf{c}_1) + x_{i_1} \pi_H(\mathbf{c}_{i_1}) + \cdots + x_{i_{m-1}} \pi_H(\mathbf{c}_{i_{m-1}}) = \sum_{j=0}^{m-1} x_{i_j} \pi_H(\mathbf{c}_{i_j}) = 0.$$

By a normalization $y_{i_j} := x_{i_j} / \sqrt{\sum_{j=0}^{m-1} |x_{i_j}|^2}$, we rewrite as (with $i_0 = 1$)

$$y_{i_0} \pi_H(\mathbf{c}_{i_0}) + y_{i_1} \pi_H(\mathbf{c}_{i_1}) + \cdots + y_{i_{m-1}} \pi_H(\mathbf{c}_{i_{m-1}}) = 0. \tag{28}$$

For $1 \le i \le n-1$, let $\mathbf{u}_i = \pi_H(\mathbf{e}_i) \in \mathbf{R}^{m-1}$ be the projection of the standard unit vector $\mathbf{e}_i$. Then for $1 \le j \le m-1$

$$\pi_H(\mathbf{c}_{i_j}) = \sum_{1 \le i \le n-1} a_{i i_j} \mathbf{u}_i,$$

where $a_{i i_j}$ are the entries of our matrix A.

In other words, one can view the $(m-1) \times m$ matrix $M = (\pi_H(\mathbf{c}_{i_0}), \pi_H(\mathbf{c}_{i_1}) \ldots, \pi(\mathbf{c}_{i_{m-1}}))$ as

$$M = \sum_{1 \le i \le n-1, 0 \le j \le m-1} a_{i i_j} M_{i i_j},$$

where $M_{i i_j}$ is the $(m-1) \times m$ matrix whose columns are zero except the i_j-th one, which is $\mathbf{u}_i$. Next we record a useful lemma about the matrix M, which can be proved by standard techniques from [22, 29].

Lemma 6.2. *With high probability with respect to a_{ij}, the least singular value of MM^* is at least m^{-2} and the largest singular value of MM^* is at most m^2.*

Let $\mathcal{E} = \mathcal{E}_{i_1, \ldots, i_{m-1}}$ be this event. By the property of projection $\pi_H^* \pi_H = I_{m-1}$,

$$\sum_{0 \le j \le m-1} \sum_{1 \le i \le n-1} M_{i i_j} M_{i i_j}^* = I_{m(m-1)}, \tag{29}$$

where we view $M_{i i_j}$ as vectors in $\mathbf{R}^{m(m-1)}$.

Now for any fixed $(t_0, \ldots, t_d) \in \mathbf{R}_+^{d+1}$, with $d \le m^{1/2}$, let $\Omega \subset \mathbf{R}^{m(m-1)}$ be the set of matrices M of size $m \times (m-1)$ satisfying Lemma 6.2 such that the normal vector $(y_0, \ldots, y_{m-1})$ satisfies $\sqrt{m}|y_{i_0}| \le t_0, \ldots, \sqrt{m}|y_{i_d}| \le t_d$. For convenience, define

$$p_{t_0,\ldots,t_d} := \mathbf{P}\Big(\sqrt{m}|y_{i_0}| \le t_0, \ldots, \sqrt{m}|y_{i_d}| \le t_d \Big| \mathbf{c}_j, j \notin \{i_0,\ldots,i_{m-1}\} \wedge \mathcal{E}\Big)$$

$$= \mathbf{P}\Big(\sum_{1 \le i \le n-1, 0 \le j \le m-1} a_{ii_j} M_{ii_j} \in \Omega \Big| \mathbf{c}_j, j \notin \{i_0,\ldots,i_{m-1}\} \wedge \mathcal{E}\Big). \tag{30}$$

As with (29) we are ready to apply Lemma 2.5. It is crucial to notice that conditioning on $\mathbf{c}_j, j \notin \{i_0,\ldots,i_{m-1}\}$, the approximating matrix $\sum_{1 \le i \le n-1, 0 \le j \le m-1} \mathbf{g}_{ii_j} M_{ii_j}$ is a gaussian iid matrix of size $(m-1) \times m$, and hence Theorem 1.2 applies to the normal vector $(y_{i_0,\mathbf{g}}, \ldots, y_{i_d,\mathbf{g}})$ of this matrix

$$\mathbf{P}\Big(\sum_{1 \le i \le n-1, 0 \le j \le m-1} \mathbf{g}_{ii_j} M_{ii_j} \in \Omega/\partial_\varepsilon \Omega\Big) - O\Big(m^5 \varepsilon^{-3} \max_{ii_j} \|M_{ii_j}\|_\infty\Big) \le p_{t_0,\ldots,t_d} \le$$

$$\le \mathbf{P}\Big(\sum_{1 \le i \le n-1, 0 \le j \le m-1} \mathbf{g}_{ii_j} M_{ii_j} \in \Omega \cup \partial_\varepsilon \Omega\Big) + O\Big(m^5 \varepsilon^{-3} \max_{ij} \|M_{ij}\|_\infty\Big). \tag{31}$$

For $\|M_{ii_j}\|_\infty$, we apply the following crucial lemma from [31, Proposition 3.5].

Lemma 6.3 (flatness of orthogonal projection). *There exists a positive constant c (independently of C_0) such that the following holds with overwhelming probability with respect to $\mathbf{c}_j, j \notin \{i_0,\ldots,i_{m-1}\}$: for any unit vector $\mathbf{v} \in H^\perp$ we have*

$$\|\mathbf{v}\|_\infty \le n^{-c}.$$

For short we let $\mathcal{G}_{i_1,\ldots,i_{m-1}}$ be the event considered in Lemma 6.3, thus

$$\mathbf{P}(\mathcal{G}_{i_1,\ldots,i_{m-1}}) = 1 - n^{-\omega(1)}.$$

Let us now consider the sets $\Omega/\partial_\varepsilon \Omega$ and $\Omega \cup \partial_\varepsilon \Omega$. Assume that $M, M' \in \Omega$ with normal vectors $\mathbf{y} = (y_0,\ldots,y_{m-1})$ and $\mathbf{y}' = (y'_0,\ldots,y'_{m-1})$ and such that $\|M - M'\|_\infty \le \varepsilon$. Then as $\|M\mathbf{y}'\|_2 = \|(M - M')\mathbf{y}'\|_2 \le \|M' - M\|_2 \le m\varepsilon$, we have (rather generously) $\|M^* M\mathbf{y}'\|_2 \le m^3\varepsilon$. By definition of Ω (which satisfies Lemma 6.2), it then follows that (again very generously)

$$\|\mathbf{y} - \mathbf{y}\|_\infty \le m^8 \varepsilon.$$

Hence it follows from (31) that

$$\mathbf{P}\Big(\sqrt{m}|y_{i_0,\mathbf{g}}| \le t_0 - m^8\varepsilon, \ldots, \sqrt{m}|y_{i_d,\mathbf{g}}| \le t_d - m^8\varepsilon\Big) - O\Big(m^5 \varepsilon^{-3} \max_{ii_j} \|M_{ii_j}\|_\infty\Big) \le p_{t_0,\ldots,t_d} \le$$

$$\le \mathbf{P}\Big(\sqrt{m}|y_{i_0,\mathbf{g}}| \le t_0 + m^8\varepsilon, \ldots, \sqrt{m}|y_{i_d,\mathbf{g}}| \le t_d + m^8\varepsilon\Big) + O\Big(m^5 \varepsilon^{-3} \max_{ij} \|M_{ij}\|_\infty\Big). \tag{32}$$

Now choose $\varepsilon = n^{-c/4}$ (with c from Lemma 6.3) and $m = n^{c/64}$. We have

$$\mathbf{P}\Big(\sqrt{m}|y_{i_0,\mathbf{g}}| \le t_0 - n^{-c/8}, \ldots, \sqrt{m}|y_{i_d,\mathbf{g}}| \le t_d - n^{-c/8}\Big) - O(n^{-c/8}) \le p_{t_0,\ldots,t_d} \le$$

$$\le \mathbf{P}\Big(\sqrt{m}|y_{i_0,\mathbf{g}}| \le t_0 + n^{-c/8}, \ldots, \sqrt{m}|y_{i_d,\mathbf{g}}| \le t_d + n^{-c/8}\Big) + O(n^{-c/8}). \qquad (33)$$

By Theorem 1.2, we have, for some constant c' sufficiently small depending on c

$$\Big| p_{t_0,\ldots,t_d} - \mathbf{P}\Big(\sqrt{m}|y_{i_0,\mathbf{g}}| \le t_0, \ldots, \sqrt{m}|y_{i_d,\mathbf{g}}| \le t_d\Big)\Big| = O(n^{-c'}). \qquad (34)$$

Now we pass from $\sqrt{m}y_{i_j}$ to $\sqrt{n}x_{i_j}$ conditioning on $\mathcal{E} \wedge \mathcal{F}_{i_1,\ldots,i_{m-1}}$. On this event, by (27)

$$\Big| |x_{i_0}|^2 + \cdots + |x_{i_{m-1}}|^2 - \frac{m}{n}\Big| \le \frac{\log^2 n}{n}.$$

In other words,

$$\Big| \sqrt{|x_{i_0}|^2 + \cdots + |x_{i_{m-1}}|^2} - \sqrt{\frac{m}{n}}\Big| \ll \frac{\log^2 n}{n}\Big/\sqrt{\frac{m}{n}} \ll \frac{\log^2 n}{\sqrt{mn}}.$$

Consequently,

$$|\sqrt{m}y_{i_j} - \sqrt{n}x_{i_j}| = \Big|\sqrt{m}x_{i_j}\Big(\frac{1}{\sqrt{|x_{i_0}|^2 + \cdots + |x_{i_{m-1}}|^2}} - \frac{1}{\sqrt{\frac{m}{n}}}\Big)\Big|$$

$$\ll |\sqrt{m}x_{i_j}|\frac{\log^2 n}{\sqrt{mn}}\Big/\frac{m}{n}$$

$$\ll \frac{\log^{5/2} n}{m}, \qquad (35)$$

where we used the bound $|x_i| = O(\frac{\log^{1/2} n}{\sqrt{n}})$ in the last estimate.

In summary, it follows from (34) and (35) that conditioning on $\mathcal{E} \wedge \mathcal{F}_{i_1,\ldots,i_{m-1}} \wedge \mathcal{G}_{i_1,\ldots,i_{m-1}}$

$$\mathbf{P}\Big(\sqrt{n}|x_{i_0}| \le t_0, \ldots, \sqrt{n}|x_{i_d}| \le t_d\Big) = \mathbf{P}\Big(|\mathbf{g}_{\mathbf{F},0}| \le t_0, \ldots, |\mathbf{g}_{\mathbf{F},d}| \le t_d\Big) + O(n^{-c''}). \qquad (36)$$

for some absolute constant c''. In particularly, this immediately implies (25) as all of the conditional events hold with high probability.

7. PROOF OF (9)

The treatment here follows closely [33, Proposition 25]. Let α be a number growing slowly to infinity that will be specified later. For each component x_i of $\mathbf{x}$ we decompose

$$x_i = x_i 1_{|\sqrt{n}x_i| \leq \alpha} + x_i 1_{|\sqrt{n}x_i| > \alpha} := x_{i,\leq} + x_{i,>}.$$

We then decompose $\mathbf{x} = \mathbf{x}_\leq + \mathbf{x}_>$ accordingly. For (9) it suffices to show

Claim 7.1. *With an appropriate choice of α we have*

(i) $\sqrt{n}\mathbf{x}_\leq^T \mathbf{u} \xrightarrow{d} N(0,1)$;

(ii) $\sqrt{n}\mathbf{x}_>^T \mathbf{u}$ *converges to zero in probability.*

For (ii), we will estimate the second moment

$$n\mathbf{E}(\mathbf{x}_>^T\mathbf{u})^2 = n \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq n} u_i u_j \mathbf{E}x_{i,>}x_{j,>}.$$

Because ξ is symmetric, $\mathbf{E}x_{i,>}x_{j,>} = 0$ if $i \neq j$, and hence

$$n\mathbf{E}(\mathbf{x}_>^T\mathbf{u})^2 = n \sum_{1 \leq i \leq n} u_i^2 \mathbf{E}x_{i,>}^2 = n\mathbf{E}x_{1,>}^2.$$

Now, by the exchangeability, $n\mathbf{E}x_1^2 = 1$. Also, by (8)

$$n\mathbf{E}x_1^2 1_{|\sqrt{n}x_1| \leq \alpha} = \mathbf{E}N(0,1)^2 1_{|N(0,1)| \leq \alpha} + O(n^{-c}).$$

It thus follows that, as $\alpha \to \infty$ together with n

$$n\mathbf{E}x_{1,>}^2 = n\mathbf{E}x_1^2 1_{|\sqrt{n}x_1| > \alpha} = o(1). \tag{37}$$

By Markov's bound, $|\sqrt{n}\mathbf{x}_>^T\mathbf{u}| \to 0$ in probability as claimed in (ii).

For (i), by Carleman's criteria, it suffices to show that for every fix positive integer k the k-moment of $\sqrt{n}\mathbf{x}_\leq^T\mathbf{u}$ asymptotically matches with that of $\mathbf{g_R}$. We have

$$n^{k/2}\mathbf{E}(\mathbf{x}_\leq^T\mathbf{u})^k = n^{k/2} \sum_{i_1,\ldots,i_k} u_{i_1} \ldots u_{i_k} \mathbf{E}x_{i_1,\leq} \ldots x_{i_k,\leq}.$$

Now we make use of the symmetry assumption on ξ. By this assumption, the expectation vanishes unless each index i appears an even number of times. Furthermore, by (8)

$$n^{k/2}\mathbf{E}x_{i_1,\leq}\ldots x_{i_k,\leq} = \mathbf{Eg}_{\mathbf{R},i_1}\ldots\mathbf{g}_{\mathbf{R},i_k} + o(1).$$

Thus

$$n^{k/2}\sum_{i_1,\ldots,i_k} u_{i_1}\ldots u_{i_k}\mathbf{E}x_{i_1,\leq}\ldots x_{i_k,\leq} = \sum_{i_1,\ldots,i_k} u_{i_1}\ldots u_{i_k}\mathbf{Eg}_{\mathbf{R},i_1}\ldots\mathbf{g}_{\mathbf{R},i_k} + o(\sum_* |u_{i_1}\ldots u_{i_k}|)$$
$$= \mathbf{E}(\mathbf{g_R})^k + o(\sum_* |u_{i_1}\ldots u_{i_k}|),$$

where the implied constant can depend on k, and $\sum_*$ indicates over all k-tuples $i_1,\ldots,i_k$ in which each index i appears an even number of times. To complete the proof, we just note that

$$\sum_* |u_{i_1}\ldots u_{i_k}| \leq (\sum_i u_i^2)^{k/2} = 1.$$

References

[1] R. Adamczak, D. Chafai and P. Wolff, Circular law for random matrices with exchangeable entries, *Random Structures & Algorithms*, **48** (2016), 3, 454-479.

[2] F. Benaych-Georges and S. Péché, Localization and delocalization for heavy tailed band matrices, *Annales de l'Institut Henri Poincaré*, **50** (2014), 4, 1385-1403.

[3] A. Bloemendal, L Erdős, A. Knowles, H.T. Yau and J. Yin, Isotropic local laws for sample covariance and generalized Wigner matrices, *Electronic Journal of Probability*, **19** (2014), 33-53.

[4] C. Bordenave and A. Guionnet, Localization and delocalization of eigenvectors for heavy-tailed random matrices, *Probability Theory and Related Fields*, **157** (2013), 885-953.

[5] P. Bourgade and H.-T. Yau, The eigenvector moment flow and local quantum unique ergodicity, *to appear in Communications in Mathematical Physics*, `arxiv.org/abs/1312.05301`.

[6] J. Bourgain, P. Matchett and V. Vu, On the singularity probability of discrete random matrices, *Journal of Functional Analysis 258* (2010), no.2, 559-603.

[7] C. Cacciapuoti, A. Maltsev and B. Schlein, Local Marchenko-Pastur law at the hard edge of sample covariance matrices, *Journal of Mathematical Physics*, **54** (2013), 043302.

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

[9] L. Erdős and A. Knowles, Quantum diffusion and eigenfunction delocalization in a random band matrix model, *Communication in Mathematical Physics*, **303** (2011), 509-554.

[10] L. Erdős and A. Knowles, Quantum diffusion and delocalization for band matrices with general distribution, *Annales de l'Institut Henri Poincaré* , **12** (2011), 1227-1319.

[11] L. Erdős, A. Knowles, H.-T. Yau and J. Yin, Delocalization and diffusion profile for random band matrices, *Communication in Mathematical Physics*, **323** (2013), 1, 367-416.

[12] L. Erdős, A. Knowles, H.-T. Yau and J. Yin, Spectral statistics of Erdős-Rényi graphs I: local semicircle law, *Annals of Probability*, **41** (2013), no. 3B, 2279-2375.

[13] L. Erdős, A. Knowles, H.-T. Yau and J. Yin, Spectral statistics of Erdős-Rényi graphs II: eigenvalue spacing and the extreme eigenvalues, *Communication in Mathematical Physics*, **314** (2012), no. 3, 587-640.

[14] L. Erdős, B. Schlein and H.-T. Yau, Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices, *Annals of Probability*, **37** (2009), 815-852.

[15] L. Erdős, B. Schlein, and H.-T. Yau, Local semicircle law and complete delocalization for Wigner random matrices, *Communication in Mathematical Physics*, **287** (2009), 641-655.

[16] L. Erdős, B. Schlein and H.-T. Yau, Wegner estimate and level repulsion for Wigner random matrices, *International Mathematics Research Notices*, 2010, 436-479.

[17] H. Goldstine and J. von Neumann, Numerical inverting of matrices of high order, *Bull. Amer. Math. Soc.* **53** (1947), 1021-1099.

[18] D. L. Hanson and E. T. Wright, A bound on tail probabilities for quadratic forms in independent random variables, *Ann. Math. Statist.*, **42** (1971), 1079-1083.

[19] A. Knowles and J. Yin, Eigenvector distribution of Wigner matrices, *Probability Theory and Related Fields*, **155** (2013), No. 3, 543-582.

[20] H. Nguyen and V. Vu, Random matrices: law of the determinant, *Annals of Probability*, (2014), Vol. **42** (2014), No. 1, 146-167.

[21] S. O'Rourke, V. Vu and K. Wang, *Eigenvectors of random matrices: a survey*, `http://arxiv.org/abs/1601.03678`.

[22] M. Rudelson, R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices, *Advances in Mathematics*, **218** (2008), no. 2, 600-633.

[23] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix, *Communications on Pure and Applied Mathematics*, **62** (2009), 1707-1739.

[24] M. Rudelson, R. Vershynin, The least singular value of a random square matrix is $O(n^{-1/2})$, *Comptes rendus de l'Académie des sciences - Mathématique* 346 (2008), 893-896.

[25] M. Rudelson and R. Vershynin, Hanson-Wright inequality and sub-gaussian concentration, *Electronic Communications in Probability*, **18** (2013), 1-9.

[26] M. Rudelson and R. Vershynin, Delocalization of eigenvectors of random matrices with independent entries, *Duke Mathematical Journal*, to appear, `arXiv:1306.2887`.

[27] M. Shub and S. Smale, Complexity of Bezout's theorem II: volumes and probabilities, *Computational Algebraic Geometry*, in: Progr. Math., vol. **109**, Birkhouser, 1993, pp. 267-285.

[28] T. Tao, Topics in random matrix theory, *Graduate Studies in Mathematics*, **132**, American Mathematical Society, Providence, RI, 2012.

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

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

[31] T. Tao and V. Vu, Random matrices: the distribution of the smallest singular values, *Geometric and Functional Analysis* **20** (2010), no. 1, 260-297.

[32] T. Tao and V. Vu, Random matrices: universality of local eigenvalue statistics, *Acta Mathematica*, **206** (2011), 127-204.

[33] T. Tao and V. Vu, Random matrices: universal properties of eigenvectors, *Random Matrices Theory Application*, **1** (2012), no. 1.

[34] T. Tao and V. Vu, Random matrices: The Universality phenomenon for Wigner ensembles, *Modern Aspects of Random Matrix Theory*, Proceedings of Symposia in Applied Mathematics, 2014: Vol. 72.

[35] L. Tran, V. Vu and K. Wang, Sparse random graphs: eigenvalues and eigenvectors, *Random Structures & Algorithms*, **42** (2013), 110-134.

[36] V. Vu and K. Wang, Random weighted projections, random quadratic forms and random eigenvectors, *Random Structures & Algorithms*, **47**, 792-821, 2015.

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

SCHOOL OF MATHEMATICS, INSTITUTE FOR ADVANCED STUDIES, PRINCETON NJ 08540

E-mail address: `nguyen.1261@math.osu.edu`

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN CT 06520

E-mail address: `van.vu@yale.edu`