Let \( \mathbf{v}_1, \ldots, \mathbf{v}_{n-1} \) be \( n - 1 \) independent vectors in \( \mathbb{R}^n \) (or \( \mathbb{C}^n \)). 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 \( \xi \) is normalized if it has mean 0 and variance 1. A complex random variable \( \xi \) is normalized if \( \xi = \frac{1}{\sqrt{2}} \xi_1 + \frac{1}{\sqrt{2}} i \xi_2 \), where \( \xi_1, \xi_2 \) are iid copies of a real normalized random variable.

Example 1.1. Some popular normalized variables

- real standard Gaussian \( g_R = \mathcal{N}(0, 1) \), or real Bernoulli \( b_R \) which takes value \( \pm 1 \) with probability 1/2;
- complex standard Gaussian \( g_C = \frac{1}{\sqrt{2}} g_{1,R} + \frac{1}{\sqrt{2}} i g_{2,R} \), or complex Bernoulli \( b_C = \frac{1}{\sqrt{2}} b_{1,R} + \frac{1}{\sqrt{2}} i b_{2,R} \).

Received April 17 2016; Revised September 5, 2016; Accepted October 25, 2016

© The Author(s) 2016. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permission@oup.com.
Fix a normalized random variable $\xi$ and consider the random vector $\mathbf{v} = (\xi_1, \ldots, \xi_n)$, whose entries are iid copies of $\xi$. 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 $\xi$; the $\mathbf{v}_i$ are the row vectors of $A$. Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$ be a unit vector that is orthogonal to the $\mathbf{v}_i$ (Here and later $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, depending on the support of $\xi$.) First, recent studies in the singularity probability of random non-Hermitian matrices (see for instance [6, 21]) show that under very general conditions on $\xi$, with extremely high probability $A$ has rank $n-1$. In this case $\mathbf{x}$ is uniquely determined up to the sign $\pm 1$ when $\mathbb{F} = \mathbb{R}$ or by a uniformly chosen rotation $\exp(i\theta)$ when $\mathbb{F} = \mathbb{C}$. Throughout the article, we use asymptotic notation under the assumption that $n$ tends to infinity. In particular, $X = O(Y), X \ll Y, \text{ or } Y \gg X$ means that $|X| \leq CY$ for some fixed $C$.

When the entries of $A$ are iid standard Gaussian $g_{\mathbb{F}}$, 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 $\mathbb{F}^n$. One then deduces the following properties (see for instance [20][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
  $$\mathbf{x} := \left(\frac{\xi_1}{S}, \ldots, \frac{\xi_n}{S}\right)$$
  where $\xi_i$ are iid standard Gaussian $g_{\mathbb{F}}$, 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,
  $$\sqrt{n} \mathbf{x}^\ast \mathbf{u} \xrightarrow{d} g_{\mathbb{F}};$$

- (the largest coordinate) for any $C > 0$, with probability at least $1 - n^{-c}$
  $$\|\mathbf{x}\|_\infty \leq \sqrt{\frac{8(C+1)^3 \log n}{n}};$$

- (the smallest coordinate) for $n \geq 2$, any $0 \leq c < 1$, and any $a > 1$,
  $$\|\mathbf{x}\|_{\min} = \min(|x_1|, \ldots, |x_n|) \geq \frac{c}{a} \frac{1}{n^{3/2}}$$
  with probability at least $\exp(-2c) - \exp\left(-\frac{a^2 - \sqrt{2a^2 - 1}}{2} n\right)$.\qed
Motivated by the universality phenomenon (see, for instance [31]), it is natural to ask if these properties are universal, namely that they hold if $\xi$ 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 $\xi$ is sub-gaussian if there exists a parameter $K_0 > 1$ such that for all $t$

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

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

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

**Theorem 1.4** (Main result). Suppose that $A = (a_{ij})_{1 \leq i \leq n-1, 1 \leq j \leq n}$, where $a_{ij}$ are iid copies of a normalized sub-gaussian random variable $\xi$. Let $x$ be the normal vector of the rows of $A$, then the followings hold.

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

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

In particularly, for any $A > 0$ there exists a constant $C_A$ such that

$$P(\|x\|_\infty \geq C_A \sqrt{\log n} n) \leq n^{-A}.$$

- (the smallest coordinate) with high probability

$$\|x\|_{\text{min}} \geq \frac{1}{n^{3/2} \log^{o(1)} n}. \quad (7)$$

- (joint distribution of the coordinates) There exists a positive constant $c$ such that the following holds: for any $d$-tuple $(i_1, \ldots, i_d)$, 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 $\Omega \in \mathcal{F}^d$,

$$|\mathbf{P}(\langle \sqrt{n}x_i, \ldots \sqrt{n}x_d \rangle \in \Omega) - \mathbf{P}(\langle g_{F,1}, \ldots , g_{F,d} \rangle \in \Omega)| \leq d^{-c'},$$  \hspace{1cm} (8)

where $g_{F,1}, \ldots , g_{F,d}$ are iid standard Gaussian.

- (inner product with a fixed vector) Assume furthermore that $\xi$ is symmetric, then for any fixed vector $u$ on the unit sphere,

$$\sqrt{n}x^* u \xrightarrow{d} g_F. \hspace{1cm} (9)$$

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

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

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 direct applications (we also refer the reader to the works of Garnaev–Gluskin [16] and of Kashin [18] for a different version of delocalization, which has found fundamental applications in compressive sensing.)

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 $\xi$. Let $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ be its singular values. The two extremal $\sigma_1$ and $\sigma_n$ 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 [15], Goldstein and von Neumann speculated that $\sigma_n$ 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 [7] in the gaussian case.
Theorem 1.6. For any $t \geq 0$ we have

\[ P(\sigma_n(M_{\text{gr}}) \leq tn^{-1/2}) = \int_0^t \frac{1 + \sqrt{x}}{2\sqrt{x}} e^{-x/2 + \sqrt{x}} dx + o(1) \]

as well as

\[ P(\sigma_n(M_{\text{gc}}) \leq tn^{-1/2}) = \int_0^t e^{-x} dx. \]

In other words, $P(\sigma_n(M_{\text{gr}}) \leq tn^{-1/2}) = 1 - e^{-t/2 + \sqrt{t}} + o(1)$ and $P(\sigma_n(M_{\text{gc}}) \leq tn^{-1/2}) = 1 - e^{-t}$. These distributions have been confirmed to be universal (in the asymptotic sense) by Tao and the second author [28].

In applications, one usually needs large deviation results, which show that the probability that $\sigma_n$ is far from its mean is very small. For the lower bound, Rudelson and Vershyn [21] proved that for any $t > 0$

\[ P(\sigma_n \leq tn^{-1/2}) \leq Ct + .999^n, \quad (10) \]

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

\[ P(\sigma_n \geq tn^{-1/2}) \leq C \frac{\log t}{t}. \quad (11) \]

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 sub-gaussian random variable $\xi$ in either $\mathbb{R}$ or $\mathbb{C}$. Then there exist absolute constants $C_1, C_2$ depending on $K_0$ such that

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

Our proof of Theorem 1.7 is totally different from that of [? ]. 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 [14] by Erdős et al. and and continued by Tao et al. and by many others in [2–4, 8–12, 29, 32, 33]. Thanks to new universality techniques, one also proved normality of the eigenvectors; see for instance the work [19] by Knowles and Yin, [30] 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 $v_1, \ldots, v_n$ be the corresponding unit eigenvectors (where $v_i$ are chosen according to the Haar measure from the eigensphere if the corresponding roots are multiple). Recently, Rudelson and Vershynin [24] proved that with overwhelming probability all of the eigenvectors satisfy

$$\|v_i\|_\infty = O\left(\frac{\log^{9/2} n}{\sqrt{n}}\right).$$ (12)

By modifying the proof of Theorem 1.4, we are able to 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 sub-gaussian random variable $\xi$ in either $\mathbb{R}$ or $\mathbb{C}$. Then for any fixed $\varepsilon > 0$, with overwhelming probability the following holds for any unit eigenvector $x$ corresponding to an eigenvalue $\lambda$ of $A$ with $|\lambda| = O(1)$

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

We believe that the individual eigenvector in Theorem 1.9 satisfies the normality property (8), which would imply that the bound $O\left(\sqrt{\frac{\log n}{n}}\right)$ 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 hold (with logarithmic correction) under a weaker assumption that the variable $\xi$ is sub-exponential, namely there are positive constants $C, C'$ and $\alpha$ such that for all $t$ $P(|\xi| \geq t) \leq C \exp(-C't^\alpha)$; see Remark 2.3.

The rest of the article is organized as follows. After introducing supporting lemmas in Section 2, we will prove (6) and Theorem 1.9 in Section 3. Sections 6 and 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.
Fig. 1. We sampled 1000 random complex iid Bernoulli matrices of size \( n = 500 \). The histograms represent the normalized real and imaginary parts \( \sqrt{2n} \text{Re}(\mathbf{v}(1)) \) and \( \sqrt{2n} \text{Im}(\mathbf{v}(1)) \) of the first coordinate of the unit eigenvector \( \mathbf{v} \) associated with the eigenvalue of smallest modulus.

2 The Lemmas

We will use the following well-known concentration result of distances in random non-Hermitian matrices (see for instance [29, Lemma 43], [25, Corollary 2.19] or [33]).

Lemma 2.1. Let \( H \) be a subspace of co-dimension \( m \) in \( \mathbb{F}^l \) 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 \( \mathbb{F} \)-normalized sub-gaussian random variable \( \xi \). Then the following holds.

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

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

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

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

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 \( \mathbb{F} \)-normalized random variable \( \xi \). 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 $\xi$. Then

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

In particularly, for any $t > 0$

$$\mathbb{P}(|\mathbf{x}^* \mathbf{A} \mathbf{x} - \mathbb{E} \mathbf{x}^* \mathbf{A} \mathbf{x}| > t \|\mathbf{A}\|_{HS}) \leq O\left(\exp\left(-c \frac{t^2}{K_0^3}\right) + \exp\left(-c \frac{t}{K_0^2}\right)\right).$$

This lemma was first proved by Hanson and Wright in a special case [17]. The above general version is due to Rudelson and Vershynin [23]; see also [33] 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 article hold (with logarithmic correction) for sub-exponential variables. One can achieve this by repeating the proofs, using the results from [33] (such as [33, 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 [28]. 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 $\xi$ be a $\mathbf{F}$-normalized random variable with finite third moment $\mathbb{E}|\xi|^3 < \infty$. Let $S$ denote the random sum

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

where $\xi_i$ are iid copies of $\xi$. Then for any $t > 0$ we have

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

where the implied constant depends on the third moment of $\xi$. In particular,

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

□
Lemma 2.5 (Berry–Esséen theorem for frames). [28, Proposition D.2] Let $1 \leq k \leq l$, and let $\xi$ be $F$-normalized and have finite third moment. Let $v_1, \ldots, v_l \in F^k$ be a normalized tight frame for $F^k$, in other words

$$v_1v_1^* + \cdots + v_nv_n^* = I_k,$$

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

$$S = \xi_1v_1 + \cdots + \xi_nv_n,$$

where $\xi_1, \ldots, \xi_n$ are iid copies of $\xi$. Similarly, let $G := (g_1, \ldots, g_k) \in F^k$ be formed from $k$ iid copies of the standard Gaussian random variable $g_F$. Then for any measurable $\Omega \subset F^k$ and for any $\varepsilon = \varepsilon(k, n) > 0$ we have

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

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

\[ \square \]

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 $C$

$$P(|x_1| \ll \sqrt{\frac{m}{n}}) = 1 - O\left(n \exp\left(-\frac{m}{C}\right)\right). \quad (13)$$

Let $c_i$, $1 \leq i \leq n$ be the columns of $A$. Because $\sum_{i=2}^n |x_i|^2 \leq 1$, 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

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

Let $H$ be the subspace generated by $c_j, j \geq m+1$. Let $P_H$ be the orthogonal projection from $F^{n-1}$ 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 [6, 21, 28]) that with probability $1 - \exp(-cn)$ we have $\dim(H^\perp) = m - 1$, which implies $\text{tr}(P_H) = m - 1$. 

Recall that by definition,
\[
x_1 c_1 + x_2 c_2 + \cdots + x_m c_m + \sum_{i \geq m+1} x_i c_i = 0. \tag{14}
\]
Applying \( P_H \), we have
\[
x_1 P_H c_1 = -P_H (x_2 c_2 + \cdots + x_m c_m),
\]
which implies
\[
|x_1|^2 \| P_H c_1 \|^2_2 = \left\| \sum_{j=2}^{m} x_j P_H c_j \right\|^2_2 = \sum_{2 \leq j \leq m} |x_j|^2 c_j^T P_H c_j + 2 \sum_{2 \leq j_1 < j_2 \leq m} x_{j_1} x_{j_2} c_{j_1}^T P_H c_{j_2} := \| Qx' \|^2_2, \tag{15}
\]
where \( x' = (x_2, \ldots, x_m) \) and
\[
Qx' := \sum_{j=2}^{m} x_j P_H c_j.
\]
We remark that the \( x_i \) here are not deterministic but depend on the column vectors \( c_i \).

As \( Qx' \) is linear, and as \(|x_2|^2 + \cdots + |x_m|^2 \leq m/(n-1)\), we have
\[
\| Qx' \|^2_2 \leq \sup_{y \in F^{m-1}, \| y \|^2_2 = 1} \| Qy \|^2_2 \sqrt{\frac{m}{n-1}}.
\]
Thus
\[
|x_1|^2 \| P_H c_1 \|^2_2 \leq \sup_{y \in F^{m-1}, \| y \|^2_2 = 1} \| Qy \|^2_2 \frac{m}{n-1}. \tag{16}
\]
We are going to estimate the operator norm \( \| Q \|_2 \) basing on the randomness of \( c_j, 2 \leq j \leq m \).

**Lemma 3.2.** There exists a sufficiently large constant \( C \) such that
\[
P_{c_2, \ldots, c_m} (\| Q \|^2_2 \geq Cm) = O(\exp(-2(m - 1))). \quad \square
\]
Assume Lemma 3.2 for the moment, we can complete the proof of (13) as follows. First, by Lemma 2.1, \( \| P_H c_1 \|^2_2 \geq m/2 \) with probability at least \( 1 - \exp(-m/4K_0^2) \). We then deduce from (16) and from Lemma 3.2 that
\[
P(|x_1|^2 \gg \frac{m}{n}) \leq O \left( \frac{n}{m} \exp(-\frac{m-1}{4K_0^2}) + \exp(-2(m - 1)) \right),
\]
completing the proof.
To prove Lemma 3.2, we first estimate \( \|Qy\|_2 \) for any fixed \( y \in S^{m-2} \). We will show

**Lemma 3.3.** There exists a sufficiently large constant \( C \) such that for any fixed \( y \in F^{m-1} \) with \( \|y\|_2 = 1 \),

\[
P_{c_2 \ldots c_m} (\|Qy\|_2^2 \geq Cm) = O(\exp(-4(m - 1))).
\]

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

\[
P\left( \exists y \in N, \|Qy\|_2^2 \geq 2m \right) = O\left( |N| \exp(-4(m - 1)) \right) = O\left( \exp(-2(m - 1)) \right).
\]

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

\[
\|Qy'\|_2 \leq \|Qy\|_2 + \|Q(y - y')\|_2 \leq \|Qy\|_2 + \|Q\|_2/2.
\]

This implies that \( \|Q\|_2 \leq \sup_{y \in N} \|Qy\|_2 + \|Q\|_2/2 \), and hence

\[
\|Q\|_2 \leq 2 \sup_{y \in N} \|Qy\|_2.
\]

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

\[
\|P\|_2 = \|y\|_2^2 = 1.
\]

Applying Lemma 2.2 to \( S = c^*Pc \), we have

\[
P(|S - \text{tr}P| \geq t) \leq O\left( \exp(-c \frac{t^2}{K_0^2 \|P\|_2^2}) + \exp(-c \frac{t}{K_0^2 \|P\|_2}) \right).
\]

It is easy to show that

\[
\text{tr}P = (m - 1) \sum_{j=0}^{m-1} |y_j|^2 = m - 1.
\]
Taking $t = 4(c^{-1} + 1)K_0^2(m - 1) := \alpha(m - 1)$, we obtain

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

To this end, by properties of a tensor product,

$$\|P\|_{HS}^2 = \|yy^T\|_{HS}^2 \|PH\|_{HS}^2 = m - 1,$$

which implies that

$$P(S \geq (\alpha + 1)(m - 1)) = O\left( \exp(-4(m - 1)) \right). \quad (17)$$

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 $\lambda_0$ with $|\lambda_0| \leq O(1)$, the following holds with overwhelming probability with respect to $M_n$: if $\|(M_n - \lambda_0)x\|_2 \leq 1/n$ then $x$ satisfies (6). $\square$

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

$$P(|x_1| \ll \sqrt{\frac{m}{n}}) = 1 - O\left( n \exp\left( -\frac{m}{C} \right) \right). \quad (18)$$

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

$$x_1c_1 + x_2c_2 + \cdots + x_mc_m + \sum_{i=m+1} x_ic_i = r \quad (19)$$

for some vector $r$ with norm $\|r\|_2 \leq 1/n$, where $c_i$ is the $i$-th column of the matrix $M_n - \lambda_0$. Projecting onto $H^\perp$, we obtain

$$|x_1|^2\|PHc_1\|_2^2 \leq 2\sum_{j=2}^m x_j\|PHc_j\|_2^2 + 2\|r\|_2^2 \leq 2\sum_{1 \leq j \leq m} x_j\bar{x}_jc_j^*PHc_j + \frac{2}{n^2}.$$
Here as $|\lambda_0| = O(1)$, Lemma 2.1 is still effective, which yields $\|P_h c_1\|_2^2 \geq m/2$ with probability at least $1 - \exp(-\frac{m}{4K^2})$.

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

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

$$P_{c_2, \ldots, c_m}(\|Q\|_2^2 \geq Cm) = O(\exp(-2(m - 1))).$$

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 $y$.

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

$$P_{c_2, \ldots, c_m}(\|Qy\|_2^2 \geq Cm) = O(\exp(-4(m - 1))).$$

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

$$\sum_{1 \leq i, j \leq m-1} y_i \tilde{y}_j c_i^{*} P_h c_j = \sum_{1 \leq i, j \leq m-1} y_i \tilde{y}_j c_i^{*} P_h c_j^{'}$$

$$- \lambda_0 \sum_{1 \leq i, j \leq m-1} y_i \tilde{y}_j c_i^{*} P_h f_j$$

$$- \lambda_0 \sum_{1 \leq i, j \leq m-1} y_i \tilde{y}_j f_i^{*} P_h c_j^{'}$$

$$+ |\lambda_0|^2 \sum_{1 \leq i, j \leq m-1} y_i \tilde{y}_j f_i^{*} P_h f_j$$

$$:= S + S' + S'' + S'''.$$

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

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

Next, we have

$$|S'| = |\lambda_0 \sum_{1 \leq i, j \leq m-1} y_i \tilde{y}_j c_i^{*} P_h f_j| = |\lambda_0 (\sum_{1 \leq i \leq m-1} y_i c_i^{*}) (\sum_{1 \leq j \leq m} y_j P_h f_j)|$$

$$= |\lambda_0 (\sum_{1 \leq i \leq m-1} y_i c_i^{*}) P_h (\sum_{1 \leq j \leq m-1} \tilde{y}_j f_j)|.$$
Additionally, as \( \|P_H\|_2 \leq 1 \) and \( \|y\|_2 = 1 \), by the properties of \( f_i \) the vector \( z := P_H(\sum_{1 \leq j \leq m-1} \tilde{y}_j f_j) \) has norm at most \( \|z\|_2 \leq 1 \). As such, the sub-gaussian random variable \( (\sum_{1 \leq i \leq m-1} y_i c_i^*)z \) has variance at most one, and hence
\[
P\left(\left| \sum_{1 \leq i \leq m-1} y_i c_i^* \right| \geq m - 1 \right) = o\left( \exp(-4(m - 1)) \right).
\]

We can argue similarly for \( S'' \) to obtain the same bound. Finally,
\[
|S'''| = |\lambda_0|^2 \|P_H(\sum_{1 \leq j \leq m-1} y_j f_j)\|_2^2 \leq |\lambda_0|^2.
\]

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 \( x' = (x_2, \ldots, x_n) \), we have
\[
Ax = x_1 c_1 + Mx' = 0.
\]
As it is known that with probability at least \( 1 - \exp(-cn) \) the matrix \( M \) is invertible; in this case, we can write
\[
x_1 M^{-1} c_1 = -x'.
\]

Since
\[
|x_1|^2 \|M^{-1} c_1\|_2^2 = \|x'\|_2^2 = 1 - |x_1|^2,
\]
we obtain
\[
|x_1|^2 = \frac{1}{1 + \|M^{-1} c_1\|_2^2} = \frac{1}{1 + \sum_{j=1}^{n-1} \sigma_j^{-2}|c_i^* u_j|^2},
\]
where \( \sigma_1 \geq \cdots \geq \sigma_{n-1} \) are the singular values of \( M \) with corresponding left-singular vectors \( u_1, \ldots, u_{n-1} \).

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 \( c_1 \))
\[
|c_i^* u_1| \leq C \log n \wedge \cdots \wedge |c_i^* u_{n-1}| \leq C \log n.
\]
We will need the following estimate

**Claim 4.1.** With respect to \(M\) we have

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

**Proof.** (of Claim 4.1) By (10)

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

Thus by the union bound

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

For the remaining sum \(\sum_{i=1}^{n-\log^2 n-1} \sigma_i^{-2}\), by the Cauchy-interlacing law,

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

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 [27, Lemma A.4])

\[
\sum_{j=1}^{n-\log^2 n-1} \sigma_j^{-2}(M') = \sum_{j=1}^{n-\log^2 n-1} d_j^{-2},
\]

where \(d_j\) is the distance from the \(j\)th 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 \(M\)

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

Now by (20) and Claim 4.1, we have

\[
P(|x_1| \gg \frac{1}{n^3 \log^{10} n}) \geq 1 - \frac{1}{n \log n}.
\]
By the union bound, 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 [28, Theorem 1.3], we can compare $P(\sigma_n \geq tn^{-1/2})$ with $P(\sigma_n(g_F) \geq tn^{-1/2})$, where $\sigma_n(g_F)$ is the least singular value of an $F$-normalized gaussian matrix. More precisely, it shows that there exists a positive constant $c$ such that
\[ P(\sigma_n \geq tn^{-1/2}) \leq P(\sigma_n(g_F) \geq tn^{-1/2}) + n^{-c}. \]

In the complex case, Theorem 1.6 has $P(\sigma_n(g_C) \geq tn^{-1/2}) = \exp(-t)$. 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 [28] 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
\[ P(|x_1| \geq m^{1/2}n^{-1/2}) = O(\exp(-m)). \]

Equivalently, for all $t = m \geq C \log n$
\[ P(|x_1| \geq t^{1/2}n^{-1/2}) = O(\exp(-t)). \]

One the other hand, similarly to our treatment in Section 4
\[ |x_1|^2 = \frac{1}{1 + \sum_{j=1}^{n} \sigma_j^{-2}(c_1^T u_j)^2}, \]
where $\sigma_j$ are the singular values of the random square matrix $M_n$ formed by the last $n$ columns, $c_1$ is the first column, and $u_j$ are the corresponding unit eigenvector of $\sigma_j^2$ in $M_n^* M_n$. 

---

**Note:** The text provided is a continuous fragment of a mathematical document discussing exponential upper tail bounds and related theorems. The full context includes more detailed proofs and definitions that are not fully captured in this snippet. The snippet is self-contained but may require additional background knowledge to fully understand.
Thus with probability at least $1 - O(\exp(-t))$ we have

$$1 + \sum_{j=1}^{n} \sigma_j^{-2} (c_i^T u_j)^2 \geq \frac{n}{t}. \tag{22}$$

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 $|c^T u_j|$, we use the following concentration result from [33].

**Lemma 5.2.** [33, Lemma 1.2] Let $c = (x_1, \ldots, x_n)$ be a random vector where $x_i$ are iid copies of $\xi$. Then there exists a constant $C' > 0$ such that the following holds. Let $H$ be a subspace of dimension $d$ with an orthonormal basis $\{u_1, \ldots, u_d\}$. Then for any $0 \leq c_1, \ldots, c_d \leq 1$ and any $s$

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

**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

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

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

$$P\left(\left|\sum_{j=1}^{n} \sigma_j^{-2} (c_i^T u_j)^2 - \sum_{j=1}^{n} \sigma_j^{-2} \right| \geq 2t^{1/2} \sigma_n^{-1} \sqrt{\sum_{j=1}^{n} \sigma_j^{-2} + t \sigma_n^{-2}} \right) \leq 2 \exp\left(-C' \frac{t}{K_0^4}\right).$$
Thus, with probability at least $1 - 2\exp(-C' t/\sqrt{n})$, we have

$$\sum_{j=1}^{n} \sigma_j^{-2} (c^*_j u_j)^2 \leq \sum_{j=1}^{n} \sigma_j^{-2} + 2t^{1/2} \sigma_n^{-1} \sqrt{\sum_{j=1}^{n} \sigma_j^{-2} + t \sigma_n^{-2}}. \quad (24)$$

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

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

This event guarantees that $\frac{n}{4t} \leq 100t \sigma_n^{-2}$, or equivalently $\sigma_n \leq 20 \frac{1}{\sqrt{n}}$. Our proof is complete.

6 Normality of Vectors: Proof of (8)

We will show that

$$|P(\sqrt{n}x_1 \in \Omega) - P(g_{F,1} \in \Omega)| \leq n^{-c'.} \quad (25)$$

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

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

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 $\{2, \ldots, n\}$

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

where the probability is with respect to $\{i_1, \ldots, i_{m-1}\}$. \qed
For convenience, denote by $F_{i_1\ldots i_{m-1}}$ the event
\[
\left| n|x_{i_1}|^2 + \cdots + n|x_{i_{m-1}}|^2 - (m - 1) \right| \leq \log^2 n. \tag{27}
\]

By Theorem 6.1, with $L = O(\log n)$ and $t = \log^2 n$
\[
\mathbb{P}(E \land F_{i_1\ldots i_{m-1}}) = 1 - n^{-\omega(1)}.
\]

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 $\pi_H : \mathbb{R}^{n-1} \to \mathbb{R}^{m-1}$,
\[
x_1\pi_H(c_1) + x_i\pi_H(c_{i_1}) + \cdots + x_{i_{m-1}}\pi_H(c_{i_{m-1}}) = \sum_{j=0}^{m-1} x_{i_j}\pi_H(c_{i_j}) = 0.
\]

By a normalization $y_{i_0} := x_{i_0}/\sqrt{\sum_{j=0}^{m-1} |x_{i_j}|^2}$, we rewrite as (with $i_0 = 1$)
\[
y_{i_0}\pi_H(c_{i_0}) + y_{i_1}\pi_H(c_{i_1}) + \cdots + y_{i_{m-1}}\pi_H(c_{i_{m-1}}) = 0. \tag{28}
\]

For $1 \leq i \leq n - 1$, let $u_i = \pi_H(e_i) \in \mathbb{R}^{m-1}$ be the projection of the standard unit vector $e_i$. Then for $1 \leq j \leq m - 1$
\[
\pi_H(c_{i_j}) = \sum_{1 \leq i \leq n-1} a_{i_{i_j}i} u_i,
\]
where $a_{i_{i_j}i}$ are the entries of our matrix $A$.

In other words, one can view the $(m - 1) \times m$ matrix $M = (\pi_H(c_{i_0}), \pi_H(c_{i_1}), \ldots, \pi_H(c_{i_{m-1}}))$ as
\[
M = \sum_{1 \leq i \leq n-1, 0 \leq j \leq m-1} a_{i_{i_j}i} M_{i_{i_j}i},
\]
where $M_{i_{i_j}i}$ is the $(m - 1) \times m$ matrix whose columns are zero except the $i_j$-th one, which is $u_i$. Next we record a useful lemma about the matrix $M$, which can be proved by standard techniques from [21, 26].

**Lemma 6.2.** With high probability with respect to $a_{i_{i_j}i}$, the least singular value of $MM^*$ is at least $m^{-2}$ and the largest singular value of $MM^*$ is at most $m^2$. $\square$
Let $E = E_{i_1, \ldots, i_{m-1}}$ be this event. By the property of projection $\pi_H \pi_H' = I_{m-1}$,

$$
\sum_{0 \leq j \leq m-1} \sum_{1 \leq i \leq n-1} M_{ij} M'_{ij} = I_{m(m-1)},
$$

(29)

where we view $M_{ij}$ as vectors in $\mathbb{R}^{m(m-1)}$.

Now for any fixed $(t_0, \ldots, t_d) \in \mathbb{R}_+^{d+1}$, with $d \leq m^{1/2}$, let $\Omega \subset \mathbb{R}^{m(m-1)}$ 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 $\sqrt{m} \{ y_0 \} \leq t_0, \ldots, \sqrt{m} \{ y_d \} \leq t_d$. For convenience, define

$$
p_{t_0, \ldots, t_d} := \mathbb{P} \left( \sqrt{m} \{ y_0 \} \leq t_0, \ldots, \sqrt{m} \{ y_d \} \leq t_d \mid c_j, j \notin \{i_0, \ldots, i_{m-1}\} \wedge E \right) = \mathbb{P} \left( \sum_{1 \leq i \leq n-1, 0 \leq j \leq m-1} a_{ij} M_{ij} \in \Omega \mid c_j, j \notin \{i_0, \ldots, i_{m-1}\} \wedge E \right).
$$

(30)

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

$$
\mathbb{P} \left( \sum_{1 \leq i \leq n-1, 0 \leq j \leq m-1} g_{ij} M_{ij} \in \Omega / \partial, \Omega \right) - O \left( m^5 \epsilon^{-3} \max_{i,j} \|M_{ij}\|_\infty \right) \leq p_{t_0, \ldots, t_d} \leq \mathbb{P} \left( \sum_{1 \leq i \leq n-1, 0 \leq j \leq m-1} g_{ij} M_{ij} \in \Omega \cup \partial, \Omega \right) + O \left( m^5 \epsilon^{-3} \max_{i,j} \|M_{ij}\|_\infty \right).
$$

(31)

For $\|M_{ij}\|_\infty$, we apply the following crucial lemma from [28, 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 $c_j, j \notin \{i_0, \ldots, i_{m-1}\}$: for any unit vector $v \in H^\perp$ we have

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

□

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

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

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

\[ \|y - y'\|_\infty \leq m^8 \epsilon. \]

Hence it follows from (31) that

\[
P\left( \sqrt{m}|y_{i_0, g}| \leq t_0 - m^8 \epsilon, \ldots, \sqrt{m}|y_{i_d, g}| \leq t_d - m^8 \epsilon \right) - O(m^{5} \epsilon^{-3} \max_{i_j} \|M_{ij}\|_\infty) \leq p_{t_0, \ldots, t_d} \leq \]

\[
\leq P\left( \sqrt{m}|y_{i_0, g}| \leq t_0 + m^8 \epsilon, \ldots, \sqrt{m}|y_{i_d, g}| \leq t_d + m^8 \epsilon \right) + O(m^{5} \epsilon^{-3} \max_{i_j} \|M_{ij}\|_\infty). \tag{32}
\]

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

\[
P\left( \sqrt{m}|y_{i_0, g}| \leq t_0 - n^{-c/8}, \ldots, \sqrt{m}|y_{i_d, g}| \leq t_d - n^{-c/8} \right) - O(n^{-c/8}) \leq p_{t_0, \ldots, t_d} \leq \]

\[
\leq P\left( \sqrt{m}|y_{i_0, g}| \leq t_0 + n^{-c/8}, \ldots, \sqrt{m}|y_{i_d, g}| \leq t_d + n^{-c/8} \right) + O(n^{-c/8}). \tag{33}
\]

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

\[
|p_{t_0, \ldots, t_d} - P\left( \sqrt{m}|y_{i_0, g}| \leq t_0, \ldots, \sqrt{m}|y_{i_d, g}| \leq t_d \right)| = O(n^{-c'}). \tag{34}
\]

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

\[
\left| |x_{i_0}|^2 + \cdots + |x_{i_{m-1}}|^2 - \frac{m}{n} \right| \leq \frac{\log^2 n}{n}.
\]

In other words,

\[
\sqrt{|x_{i_0}|^2 + \cdots + |x_{i_{m-1}}|^2} - \frac{\sqrt{m}}{n} \leq \frac{\log^2 n}{\sqrt{m}} - \frac{\log^2 n}{\sqrt{mn}}.
\]

Consequently,

\[
|\sqrt{m}y_{ij} - \sqrt{n}x_{ij}| = \left| \sqrt{m}x_{ij}\left( \frac{1}{\sqrt{|x_{i_0}|^2 + \cdots + |x_{i_{m-1}}|^2}} - \frac{1}{\sqrt{m}/n} \right) \right|
\leq \left| \sqrt{m}x_{ij} \right| \frac{\log^2 n}{\sqrt{mn}} \left| \frac{m}{\sqrt{n}} \right|
\leq \frac{\log^{5/2} n}{m}, \tag{35}
\]

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 $E \cap F_{i_1 \ldots i_{m-1}} \cap G_{i_1 \ldots i_{m-1}}$

$$P\left(\sqrt{n}|x_{i_0}| \leq t_0, \ldots, \sqrt{n}|x_{i_d}| \leq t_d\right) = P\left(|g_{F,0}| \leq t_0, \ldots, |g_{F,d}| \leq t_d\right) + O(n^{-c''}).$$ \hspace{1cm} (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 [30, Proposition 25]. Let $\alpha$ be a number growing slowly to infinity that will be specified later. For each component $x_i$ of $x$ we decompose

$$x_i = x_{i,1}\sqrt{n|\xi_i|} \leq \alpha + x_{i,2}|\sqrt{n|\xi_i|} > \alpha := x_{i,\leq} + x_{i,>}.$$

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

**Claim 7.1.** With an appropriate choice of $\alpha$ we have

1. $\sqrt{n|x|^T u} \xrightarrow{d} N(0, 1)$;
2. $\sqrt{n|x|^T u}$ converges to zero in probability. \hfill \Box

For (ii), we will estimate the second moment

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

Because $\xi$ is symmetric, $Ex_{i,>} x_{j,>} = 0$ if $i \neq j$, and hence

$$nE(x^T u)^2 = n \sum_{1 \leq i \leq n} u_i^2 Ex_{i,>}^2 = nEx_{i,>}^2.$$

Now, by the exchangeability, $nEx_{i,>}^2 = 1$. Also, by (8)

$$nEx_{1,1}^2 1_{|\sqrt{n}\xi_1| \leq \alpha} = EN(0, 1)^2 1_{|N(0, 1)| \leq \alpha} + O(n^{-c}).$$

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

$$nEx_{1,>}^2 = nEx_{1,1}^2 1_{|\sqrt{n}\xi_1| > \alpha} = o(1).$$ \hspace{1cm} (37)

By Markov’s bound, $|\sqrt{n|x|^T u}| \to 0$ in probability as claimed in (ii).
For (i), by Carleman’s criteria, it suffices to show that for every fixed positive integer $k$ the $k$-moment of $\sqrt{n}x_T^T u$ asymptotically matches with that of $g_R$. We have

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

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

$$n^{k/2}E x_{i_1, \leq} \cdots x_{i_k, \leq} = Eg_{R,i_1} \cdots g_{R,i_k} + o(1).$$

Thus

$$n^{k/2} \sum_{i_1, \ldots, i_k} u_{i_1} \cdots u_{i_k} E x_{i_1, \leq} \cdots x_{i_k, \leq} = \sum_{i_1, \ldots, i_k} u_{i_1} \cdots u_{i_k} Eg_{R,i_1} \cdots g_{R,i_k} + o(\sum_s |u_{i_1} \cdots u_{i_k}|)$$

$$= E(g_R)^k + o(\sum_s |u_{i_1} \cdots u_{i_k}|),$$

where the implied constant can depend on $k$, and $\sum_s$ 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

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

Funding

H.N. is supported by NSF grants DMS-1358648, DMS-1128155 and CCF-1412958. V.V. is supported by NSF grant DMS-1307797 and AFORS grant FA9550-12-1-0083.

Acknowledgments

The authors are thankful to K. Wang for helpful discussion. They are also grateful to A. Knowles with help of references.

References

[1] Adamczak, R., D. Chafai and P. Wolff. “Circular law for random matrices with exchangeable entries.” *Random Structures & Algorithms* 48, no. 3 (2016): 454–79.

[2] Benaych-Georges, F., and S. Péché. “Localization and delocalization for heavy tailed band matrices.” *Annales de l’Institut Henri Poincaré* 50, no. 4 (2014): 1385–403.

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

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

[7] Edelman, A. "Eigenvalues and condition numbers of random matrices." *SIAM Journal on Matrix Analysis and Applications* 9, no. 4 (1988): 543–60.

[8] Erdős, L., and A. Knowles. "Quantum diffusion and eigenfunction delocalization in a random band matrix model." *Communication in Mathematical Physics* 303 (2011): 509–54.

[9] Erdős, L., and A. Knowles. "Quantum diffusion and delocalization for band matrices with general distribution." *Annales de l’Institut Henri Poincaré* 12 (2011): 1227–319.

[10] Erdős, L., 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, no. 3 (2012): 587–640.

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

[12] Erdős, L., A. Knowles, H.-T. Yau, and J. Yin. "Spectral statistics of Erdős-Rényi graphs I: local semicircle law." *Annals of Probability* 41, no. 3B (2013): 2279–375.

[13] Erdős, L., B. Schlein, and H.-T. Yau. "Local semicircle law and complete delocalization for Wigner random matrices." *Communication in Mathematical Physics* 287 (2009): 641–55.

[14] Erdős, L., B. Schlein, and H.-T. Yau. "Wegner estimate and level repulsion for Wigner random matrices." *International Mathematics Research Notices* 2010, no. 3 (2010): 436–79.

[15] Goldstine H., and J. von Neumann. "Numerical inverting of matrices of high order." *Bulletin of the American Mathematical Society* 53 (1947): 1021–99.

[16] Garnaev A., and E. D. Gluskin. "The widths of a Euclidean ball." *Doklady Akademii Nauk SSSR* 277 (1984): 1048–52.

[17] Hanson D. L., and E. T. Wright. "A bound on tail probabilities for quadratic forms in independent random variables." *Annals of Mathematical Statistics* 42 (1971): 1079–83.

[18] Kashin, B. S. "Diameters of certain finite-dimensional sets in classes of smooth functions." *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya* 41 (1977): 334–51.

[19] Knowles A., and J. Yin. "Eigenvector distribution of Wigner matrices." *Probability Theory and Related Fields* 155, no. 3 (2013): 543–82.

[20] O’Rourke, S., V. Vu, and K. Wang. "Eigenvectors of random matrices: a survey." *Journal of Combinatorial Theory Series A* 144 (2016): 361–442.

[21] Rudelson, M., and R. Vershynin. "The Littlewood-Offord problem and invertibility of random matrices." *Advances in Mathematics* 218, no. 2 (2008): 600–33.

[22] Rudelson M., and R. Vershynin. "Smallest singular value of a random rectangular matrix." *Communications on Pure and Applied Mathematics* 62 (2009): 1707–39.

[23] Rudelson M., and R. Vershynin. "Hanson-Wright inequality and sub-gaussian concentration." *Electronic Communications in Probability* 18 (2013): 1–9.
[24] Rudelson M., and R. Vershynin. “Delocalization of eigenvectors of random matrices with independent entries.” *Duke Mathematical Journal* 164, no. 13 (2015): 2507–38.

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

[26] Tao T., and V. Vu. “Smooth analysis of the condition number and the least singular value.” *Mathematics of Computation* 79 (2010): 2333–52.

[27] Tao T., and V. Vu. “Random matrices: universality of ESDs and the circular law.” *Annals of Probability* 38, no. 5 (2010): 2023–65, with an appendix by M. Krishnapur.

[28] Tao T., and V. Vu. “Random matrices: the distribution of the smallest singular values.” *Geometric and Functional Analysis* 20, no. 1 (2010): 260–97.

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

[30] Tao T., and V. Vu. “Random matrices: universal properties of eigenvectors.” *Random Matrices Theory Application* 1, no. 1 (2012): 1150001, 27 p.

[31] Tao T., and V. Vu. “Random matrices: The Universality phenomenon for Wigner ensembles.” *Modern Aspects of Random Matrix Theory*. Proceedings of Symposia in Applied Mathematics, vol. 72. American Mathematical Society, 2014.

[32] Tran, L., V. Vu, and K. Wang, “Sparse random graphs: eigenvalues and eigenvectors.” *Random Structures & Algorithms* 42 (2013): 110–34.

[33] Vu V., and K. Wang. “Random weighted projections, random quadratic forms and random eigenvectors.” *Random Structures & Algorithms* 47 (2015): 792–821.