EUCLIDEAN DISTANCE BETWEEN HAAR ORTHOGONAL AND GAUSSIAN MATRICES

CARLOS E. GONZÁLEZ-GUILLÉN, CARLOS PALAZUELOS,
AND IGNACIO VILLANUEVA

Abstract. In this work we study a version of the general question of how well a Haar distributed orthogonal matrix can be approximated by a random gaussian matrix. Here, we consider a gaussian random matrix Y_n of order n and apply to it the Gram-Schmidt orthonormalization procedure by columns to obtain a Haar distributed orthogonal matrix U_n. If F_i^m denotes the vector formed by the first m-coordinates of the ith row of $Y_n - \sqrt{n}U_n$ and $\alpha = \frac{m}{n}$, our main result shows that the euclidean norm of F_i^m converges exponentially fast to $\sqrt{\left(2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}\right)m}$, up to negligible terms.

To show the extent of this result, we use it to study the convergence of the supremum norm $\epsilon_n(m) = \sup_{1\leq i\leq n, 1\leq j\leq m} |y_{i,j} - \sqrt{n}u_{i,j}|$ and we find a coupling that improves by a factor $\sqrt{2}$ the recently proved best known upper bound of $\epsilon_n(m)$. Applications of our results to Quantum Information Theory are also explained.

1. Introduction

One of the classical problems in random matrix theory is to compare a random gaussian matrix $Y_n = (y_{i,j})_{i,j=1}^n$ with a Haar distributed random matrix $U_n = (u_{i,j})_{i,j=1}^n$ in the orthogonal group $\mathcal{O}(n)$.

It has been well known for long [2] that the distribution of one single coordinate of U_n converges to the distribution of one single coordinate of Y_n, when properly normalized. That is, for a fixed pair (i, j) we have that $\sqrt{n}u_{i,j}$ converges in distribution to a standard normal. Since then, many authors have studied the problem of how many entries of $\sqrt{n}U_n$ can be simultaneously well approximated by the corresponding entries of Y_n; that is, by independent standard normal distributions.

A number of papers (for instance [7], [15], [17]) in the 1980's made further progress in this direction. Later, in [6] the authors proved that the joint distribution of the first l_n coordinates of the first column of $Y_n - \sqrt{n}U_n$ converges to 0 in variation distance as n grows to infinity, provided that $l_n = o(n)$. In [5] it was proven that the joint distribution of the upper left $l_n \times m_n$ block of $Y_n - \sqrt{n}U_n$

converges to 0 in variation distance provided that l_n, m_n are both $o(n^{\frac{1}{3}})$. Later, in [3] this order was improved to $O(n^{\frac{1}{3}})$.

The latest major achievement in this direction came from [11, 12]. In those papers the author shows that the joint distribution of the upper left $l_n \times m_n$ block of $Y_n - \sqrt{n}U_n$ converges to 0 in variation distance if and only if l_n, m_n are both $o(n^{\frac{1}{2}})$. This settles the long standing open problem of finding the best ratio in the variation distance case. In the same paper Jiang also shows [11, Theorem 3] the existence of a coupling between Y_n and U_n such that

$$\epsilon_n(m) = \sup_{1\leq i\leq n, 1\leq j\leq m} |y_{i,j} - \sqrt{n}u_{i,j}|$$

converges to 0 in probability if and only if $m = o(\frac{n}{\log n})$. Moreover, if $m = \frac{\beta n}{\log n}$ then the previous supremum converges in probability to $2\sqrt{\beta}$. These results have been applied in [16] to study the eigenvector distribution of a wide class of Wigner ensembles. For further history and applications of these results, see [11, 12, 16].

Given the relevance of the euclidean norm in many contexts, and motivated by these previous works, we study in this paper the behaviour of the euclidean norm of blocks of $Y_n - \sqrt{n}U_n$. We are interested not only in the order needed for convergence to 0, but in the general value of the norm. To show the extent of our main result, we show later how to recover from it one of the main results of [11]. We have also applied it to solve a question in Quantum Information Theory [8].

Let us fix the notation needed to state our main result. Our probability spaces will be $\mathbb{R}^{n^2}$ with the gaussian measure. For every $n \in \mathbb{N}$, $Y_n = (y_{i,j})_{i,j=1}^n$ will be a gaussian random matrix, that is, the variables $(y_{i,j})_{i,j=1}^n$ are independent standard normal variables. For every $1 \leq j \leq n$ we consider the column vector $\mathbf{y}_j = (y_{i,j})_{i=1}^n$. With probability 1, they form a basis of $\mathbb{R}^n$. Following [11], we apply the Gram-Schmidt orthonormalization procedure to $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ to obtain an orthonormal basis $(\boldsymbol{\nu}_j)_{j=1}^n$. We call U_n the matrix $(\nu_{i,j})_{i,j=1}^n$. We recall that U_n is Haar distributed.

For every $1 \leq m \leq n$ and for every $1 \leq i \leq n$, let F_i^m be the vector formed by the the first m-coordinates of the ith row of $Y_n - \sqrt{n}U_n$. We describe the asymptotic generic behavior of $\|F_i^m\|$, where $\|\cdot\|$ is the euclidean norm. Let $[x]$ denote the integer part of x.

Theorem 1.1. *Let $n \in \mathbb{N}$, let $0 < \alpha \leq 1$ be fixed and let $m = [\alpha n]$. Let Y, U, F_i^m be as above. Then, there exists $0 < \delta < \frac{1}{2}$ such that,*

$$\sup_i \|F_i^m\| \leq \sqrt{\left(2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}\right)m} + O(m^\delta)$$

and

$$\inf_i \|F_i^m\| \geq \sqrt{\left(2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}\right) m} - O(m^\delta),$$

both with probability exponentially close to 1 as n grows to infinity.

In case the ratio $\alpha_n = \frac{m}{n}$ is not constant but a funcion of n, it follows immediately from our result that $\sup_i \|F_i^m\| \to 0$ if and only if $m = o(n)$.

We have chosen this presentation of the main theorem for the sake of clarity. The actual proof shows further insight into the result. Specifically, we want to mention that there is a trade off between the rate of the concentration and the order δ appearing in Theorem 1.1. In our proof we show how to make $\delta = \frac{2}{5}$ keeping a very fast concentration rate. Nevertheless, the parameters can be changed easily to obtain a different value for δ, at the cost of modifying the rate of the exponential convergence of the probability.

For a clearer understanding of the bound we can use the Taylor expansion of $(1-\alpha)^{3/2}$ and we get that

$$2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha} = \frac{\alpha}{2} + \frac{\alpha^2}{12} + \frac{\alpha^3}{32} + \frac{\alpha^4}{64} + \frac{4}{3\alpha}r(\alpha),$$

where $r(\alpha)$ is the remainder of the 5-th order Taylor polynomial of $(1-\alpha)^{3/2}$.

We clarify next some aspects of our result. The coupling is given by the Gram-Schmidt procedure performed columnwise. With the same technics we use here it is not difficult to obtain exponential concentration results for the euclidean norm of each of the columns of $Y_n - \sqrt{n}U_n$. Since the Gram-Schmidt procedure is essentially not symmetric by columns, all of those euclidean norms will necessarily be different with probability exponentially close to 1. This makes an analogous statement to our main theorem impossible, since the euclidean norms of the columns will never concentrate around the same value. This "flatness" phenomenom is very relevant in our applications.

Therefore, our main contribution can be seen as a *delocalization* result. We show that the euclidean norm of the whole block $\left(\sum_{1\leq i\leq n, 1\leq j\leq m} |y_{i,j} - \sqrt{n}u_{i,j}|^2\right)^{\frac{1}{2}}$ is well delocalized among the euclidean norm of the rows. The lack of independence is the main difficulty in this case and we need to deal with different technicalities to overcome this and prove our result. Our main tools are standard versions of the concentration of measure phenomenom and the Gram-Schmidt procedure. Still, the proof is long and technical.

As a consequence of Theorem 1.1 we prove in Section 4 a result about the supremun norm $\epsilon_n(m)$. From this result, we can recover [11, Theorem 3], with a slight improvement in the bound at the cost of using a random coupling.

Theorem 1.2. *For each $n \geq 2$, there exist matrices $Y'_n = (y'_{ij})_{i,j=1}^n$ and $U'_n = (u'_{i,j})_{i,j=1}^n$ whose $2n^2$ entries are real random variables defined on the same probability space such that*

(i) *the law of U'_n is the normalized Haar measure on the orthogonal group $\mathcal{O}(n)$;*
(ii) *$\{y'_{i,j}; 1 \leq i,j \leq n\}$ are independent standard normals;*
(iii) *set*

$$\epsilon_n(m) = \max_{1\leq i\leq n,\, 1\leq j\leq m} |\sqrt{n}u'_{i,j} - y'_{i,j}|$$

for $m = 1, 2, \cdots, n$. Then, there exists $0 < \delta < \frac{1}{2}$ such that for any $\varepsilon > 0$ we have

$$\epsilon_n(m) \geq (1-\varepsilon)(\sqrt{\varphi(\alpha)} - O(m^{-\delta}))\sqrt{2\log n} \quad \text{and}$$
$$\epsilon_n(m) \leq (1+\varepsilon)(\sqrt{\varphi(\alpha)} + O(m^{-\delta}))\sqrt{2\log(nm)}$$

with probability $1 - o(1)$, where we consider $0 < \alpha \leq 1$ fixed, $m = [\alpha n]$ and $\varphi(\alpha) = 2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}$ is the function appearing in Theorem 1.1.

If we let α change with n in Theorem 1.2 so that $m_n = o(\frac{n}{\log n})$ we recover the convergence to 0 already obtained in the above mentioned [11, Theorem 3] (see Corollary 4.2). Furthermore, if we pick $m_n = \frac{\beta n}{\log n}$ we get that

$$\sqrt{\beta} \leq \epsilon_n(m) \leq \sqrt{2\beta}.$$

Note that in [11, Theorem 3] the author obtains for this case $\epsilon_n(m) \to 2\sqrt{\beta}$. Therefore, we improve the bound by a factor $\sqrt{2}$. The key point is that our Theorem 1.1 allows us to modify the coupling. The price we pay is that we do not obtain an explicit coupling, but a randomized one (with high probability). Details are given in Section 4.

The fact that Theorem 1.2 follows from Theorem 1.1 provides a better understanding of the order $\frac{n}{\log n}$ needed for the convergence of the supremum norm of the block. Roughly, each of the row vectors of the difference, when multiplied from the right by a random unitary, distributes uniformly on the unit sphere. Therefore, the distance between its supremum and euclidean norms is of the order $\log m$.

One of our original motivations to study this problem was to solve a question in Quantum Information Theory. The solution to this problem has an implication for random matrix theory in the form of a *non universality* result for certain statistic associated to a random matrix. This statistic separates Bernouilli random variables from gaussian random variables. We briefly describe the result next, but we refer the reader to Section 5 and [8] for more detailed definitions and further details.

Example 1.1 (Non-universality)**.** Given a square matrix γ_n of order n, we can consider it as an element of $\mathbb{R}^n \otimes \mathbb{R}^n$. Then, the statistic we will consider is the

projective tensor norm of γ_n as an element of $\ell_\infty^n \otimes \ell_\infty^n$. It is defined by

$$\|\gamma\|_{\ell_\infty^n \otimes_\pi \ell_\infty^n} = \inf\Big\{ \sum_{k=1}^N \|\mathbf{x}_k\|_\infty \|\mathbf{y}_k\|_\infty : \gamma = \sum_{k=1}^N \mathbf{x}_k \otimes \mathbf{y}_k \Big\},$$

where $\mathbf{x}_k, \mathbf{y}_k \in \mathbb{R}^n$ and $\|\mathbf{z}\|_\infty = \max\limits_{j=1,\cdots,n} |z_j|$ for any vector $\mathbf{z} = (z_1 \ldots, z_n) \in \mathbb{R}^n$.

Let $0 < \alpha < 1$ be fixed. For every $n \in \mathbb{N}$, let $m = \alpha n$. Let $X_n = (x_{i,j})_{i,j=1}^{n,m}$, $Y_n = (y_{i,j})_{i,j=1}^{n,m}$ be random matrices such that all of the random variables $x_{i,j}, y_{i,j}$ are independent identically distributed. Let $\gamma = \frac{1}{m} X Y^T$.

Suppose first that $x_{i,j}, y_{i,j}$ are Bernouilli variables taking the values ± 1 with probability $\frac{1}{2}$. Then, it follows easily from the definition of the projective tensor norm that $\|\gamma\|_{\ell_\infty^n \otimes_\pi \ell_\infty^n} \leq 1$ with probability 1. In [8] we use Theorem 1.1 as one of the main tools to prove that, if $x_{i,j}, y_{i,j}$ are standard normal variables, then there exists $\alpha_0 > 0$ such that for every $\alpha < \alpha_0$ there exists $\epsilon > 0$ such that $\|\gamma\|_{\ell_\infty^n \otimes_\pi \ell_\infty^n} > 1 + \epsilon$ with probability tending to 1 as n grows to infinity.

The rest of the paper is organized as follows. In Section 2 we fix our notation and we recall several known facts about the gaussian distribution that will be repeatedly used later on. Then, in Section 3 we prove our main result Theorem 1.1. In Section 4 we apply Theorem 1.1 to the study of the supremum norm of the $n \times m$ blocks of $Y_n - \sqrt{n} U_n$ and we prove Theorem 1.2. Next, we obtain as a corollary a slight improvement of [11, Theorem 3]. Finally, in Section 5 we sketch a proof of Example 1.1. The interested reader is referred to [8] for full details.

2. Preliminaries

In this section we fix our notation and for the sake of completeness we recall several known facts about the gaussian measure on $\mathbb{R}^n$ that will be used several times in the rest of the paper. We say that a real function $f(n)$ is $O(g(n))$ if there exist constants $C > 0$ and n_0 such that for all $n > n_0$ we have that $|f(n)| \leq C g(n)$. We say that $f(n)$ is $o(g(n))$ if $\lim\limits_{n\to\infty} \frac{f(n)}{g(n)} = 0$. We will use $O(g(x))$ and $o(g(x))$ to denote functions on these sets. We will say that a sequence of events E_n holds with probability exponentially small (respectively exponentially close to 1) if there exists $\alpha > 0$, independent of n such that $Pr(E_n) \leq O(e^{-n^\alpha})$ (respectively $Pr(E_n) \geq 1 - O(e^{-n^\alpha})$).

We recall the following well known bounds of the tail of a normal random variable.

Lemma 2.1. *Let Z be a standard normal random variable. Then, for every $t > 0$,*

$$\frac{t}{(1+t^2)\sqrt{2\pi}} e^{-\frac{t^2}{2}} \leq Pr(Z > t) \leq \frac{1}{t\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

Hence, for $t \geq 1$,

$$Pr(Z^2 > t^2) \leq e^{-\frac{t^2}{2}}.$$

We will later choose $t = m^{\frac{\epsilon}{2}}$ *to get*

$$Pr(Z^2 > m^{\epsilon}) \leq e^{-\frac{m^{\epsilon}}{2}}.$$

We will denote the standard gaussian probability measure (gaussian measure in short) in $\mathbb{R}^n$ by $\mathcal{G}_n$. We will refer to a gaussian vector (matrix) as a random vector whose coordinates are independent standard normal random variables in $\mathbb{R}$.

The following bound of the norm of a gaussian vector is well known. It can be easily deduced, for instance, from [13, Lemma 1].

Proposition 2.2. *For every* $0 < \epsilon < 1$,

$$\mathcal{G}_n\{\mathbf{x} \in \mathbb{R}^n \ : \ \|\mathbf{x}\|_2 \geq \frac{\sqrt{n}}{\sqrt{1-\epsilon}}\} \leq e^{\frac{-\epsilon^2 n}{4}}$$

and

$$\mathcal{G}_n\{\mathbf{x} \in \mathbb{R}^n \ : \ \|\mathbf{x}\|_2 \leq \sqrt{n}\sqrt{1-\epsilon}\} \leq e^{\frac{-\epsilon^2 n}{4}}.$$

We will use several times along the paper the well known fact that both the gaussian measure $\mathcal{G}_n$ in the space of vectors $\mathbb{R}^n$ and the gaussian measure $\mathcal{G}_{n^2}$ in the space of square matrices of order n are biunitarily invariant under the action of the orthogonal group $\mathcal{O}(n)$. Using this, it is very easy to see that the projection $P_L(\mathbf{x})$ of a random gaussian vector $\mathbf{x}$ onto a fixed subspace L of dimension k is a gaussian vector of this subspace.

One can see the rotationally invariant (uniform) measure μ_n in S^{n-1} as the pushforward measure of $\mathcal{G}_n$ given by the map $f(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$. That is, given a set $A \subset S^{n-1}$ we have that $\mu_n(A) = \mathcal{G}_n(f^{-1}(A))$.

Similarly, one can consider the pushforward measure of $\mathcal{G}_{n^2}$ induced by the map that takes the first k n-dimensional vectors $\mathbf{x}_1, ..., \mathbf{x}_k \in \mathbb{R}^n$ to the $span\{\mathbf{x}_1, ..., \mathbf{x}_k\}$, the linear subspace generated by them. This measure is the only one invariant under the action of $\mathcal{O}(n)$ and therefore we call it the Haar measure in the Grassmannian of the k-dimensional subspaces of $\mathbb{R}^n$.

The following proposition follows immediately from the previous explanation.

Proposition 2.3. *Let* $L \subset \mathbb{R}^n$ *be a Haar distributed* k*-dimensional subspace and let* $\mathbf{x} \in \mathbb{R}^n$ *be a gaussian vector independent from* L*. Then, for any* $0 < \epsilon < 1$,

$$Pr\left(\|P_L(\mathbf{x})\|_2 \geq \frac{\sqrt{k}}{\sqrt{1-\epsilon}}\right) \leq e^{\frac{-\epsilon^2 k}{4}}$$

and

$$Pr\left(\|P_L(\mathbf{x})\|_2 \leq \sqrt{k}\sqrt{1-\epsilon}\right) \leq e^{\frac{-\epsilon^2 k}{4}}.$$

If we replace the gaussian vector by a fixed unitary vector we obtain the following estimates.

Proposition 2.4. *Let $L \subset \mathbb{R}^n$ be a Haar distributed k-dimensional subspace and let $\mathbf{y} \in \mathbb{R}^n$ be a fixed unitary vector. Then, for any $0 < \rho < 1$ we have*

$$Pr\left(\|P_L(\mathbf{y})\| \geq \frac{1}{1-\rho}\sqrt{\frac{k}{n}}\right) \leq e^{-\frac{\rho^2 k}{4}},$$

and

$$Pr\left(\|P_L(\mathbf{y})\| \leq (1-\rho)\sqrt{\frac{k}{n}}\right) \leq e^{-\frac{\rho^2 k}{4}}.$$

For $t > 1$ we also have

$$Pr\left(\|P_L(\mathbf{y})\| \geq t\sqrt{\frac{k}{n}}\right) \leq e^{-\frac{k}{4}(t^2-2)}.$$

Proof. One can consider a Haar distributed k-dimensional subspace L as a Haar distributed orthogonal matrix U acting on a fixed k-dimensional subspace M. Hence, $P_L(\mathbf{y}) = P_M(U\mathbf{y})$. Now, the vector Uy is a random uniform vector on the unit sphere of $\mathbb{R}^n$ and, according to our explanation above, it is $\bar{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ for a gaussian vector $\mathbf{x}$. Thus, $P_L(\mathbf{y})$ is equally distributed as $P_M(\bar{\mathbf{x}})$. Then, the result can be easily deduced from the known estimates on $P_M(\bar{\mathbf{x}})$, for example, from [9, Lemma 2.2]. Also, note that a version of this proposition with slightly worse constants of the first two bounds can be easily deduced from Proposition 2.2 and Proposition 2.3. □

3. Proof of Theorem 1.1

We briefly recall our notation: $Y_n = (y_{i,j})_{i,j=1}^n$ will be a normal gaussian random matrix. We consider the column vectors $\mathbf{y}_j = (y_{i,j})_{i=1}^n$. With probability 1, they form a basis of $\mathbb{R}^n$ and, in that case, we can apply the Gram-Schmidt orthonormalization procedure to $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ and we obtain an orthonormal basis $(\boldsymbol{\nu}_j)_{j=1}^n$. We call U_n the matrix $(\nu_{i,j})_{i,j=1}^n$. For every $1 \leq m \leq n$ and for every $1 \leq i \leq n$, F_i^m is the vector formed by the the first m-coordinates of the ith row of $Y_n - \sqrt{n}U_n$.

We start the proof of Theorem 1.1 with some observations about the Gram-Schmidt orthonormalization process. Let us examine the situation in step j. The gaussian vectors $\mathbf{y}_1, \ldots, \mathbf{y}_{j-1}$ have been chosen independently. Associated to them we have the orthonormal vectors $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{j-1}$. Both sets of vectors span the same

$j-1$ dimensional subspace L_{j-1}. This subspace is distributed according to the Haar measure in the Grassmanian of the $j-1$ dimensional subspaces of $\mathbb{R}^n$.

We consider the column vectors

$$\Delta_j = \sum_{k=1}^{j-1} \langle \mathbf{y}_j, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k = P_{L_{j-1}}(\mathbf{y}_j),$$

where $P_{L_{j-1}}$ is the orthogonal projection onto L_{j-1}, and we write

$$\mathbf{y}_j - \sqrt{n}\boldsymbol{\nu}_j = \Delta_j + (\mathbf{y}_j - \Delta_j) - \sqrt{n}\boldsymbol{\nu}_j.$$

Let us call $\Delta'_j = (\mathbf{y}_j - \Delta_j) - \sqrt{n}\boldsymbol{\nu}_j$ and let us note that $(\mathbf{y_j} - \Delta_j)$ has the same direction as $\boldsymbol{\nu}_j$ (by definition of $\boldsymbol{\nu}_j$) so that

$$\Delta'_j = (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})\boldsymbol{\nu}_j = (\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n})\boldsymbol{\nu}_j,$$

where $P_{L_{j-1}^\perp}$ is the projection onto the orthogonal subspace of L_{j-1}. Note that Δ_j and Δ'_j are orthogonal to each other.

Associated to the Δ_j's and Δ'_j's, for every $1 \leq i \leq n$ and for every $1 \leq m \leq n$ we have the (truncated) row vectors

$$G_i^m = (\Delta_j(i))_{j=1}^m = \left(\sum_{k=1}^{j-1} \langle \mathbf{y}_j, \boldsymbol{\nu}_k \rangle \langle \boldsymbol{\nu}_k, \mathbf{e}_i \rangle \right)_{j=1}^m = \left(\left\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \right\rangle \right)_{j=1}^m$$

and

$$\begin{aligned} H_i^m = (\Delta'_j(i))_{j=1}^m &= \left((\|\mathbf{y}_j - \Delta_j\| - \sqrt{n}) \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle \right)_{j=1}^m \\ &= \left((\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n}) \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle \right)_{j=1}^m. \end{aligned}$$

Then,

$$\|F_i^m\|^2 = \langle G_i^m + H_i^m, G_i^m + H_i^m \rangle = \|G_i^m\|^2 + \|H_i^m\|^2 + 2\langle G_i^m, H_i^m \rangle.$$

We will upper and lower bound $\|G_i^m\|$ and $\|H_i^m\|$ outside of a set of exponentially small probability. Moreover, we show that the leading terms of $\sup_i \|G_i^m\|$ and $\inf_i \|G_i^m\|$ are equal (and the same happens for $\|H_i^m\|$) and thus the bounds are sharp. After that, we will see that $\langle G_i^m, H_i^m \rangle$ is negligible when compared with those bounds outside of a set of probability exponentially small. Finally we will get that $\|F_i^m\|$ is upper and lower bounded by the bounds of $\sqrt{\|G_i^m\|^2 + \|H_i^m\|^2}$.

First, we bound $\|G_i^m\|$.

Proposition 3.1. *With the notation of Theorem 1.1,*

$$\sup_i \|G_i^m\| \leq \sqrt{\frac{\alpha}{2} m} + O(m^\delta)$$

and

$$\inf_i \|G_i^m\| \geq \sqrt{\frac{\alpha}{2}m} - O(m^\delta)$$

with probability exponentially close to 1.

Proof. We note that

$$G_{i,j}^m = \sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \langle \boldsymbol{\nu}_k, \mathbf{y}_j \rangle = \left\langle \sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k, \mathbf{y}_j \right\rangle.$$

Therefore, to obtain the j-th coordinate of G_i^m we consider the Haar distributed $j-1$ dimensional subspace $L_{j-1} = span\{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{j-1}\} = span\{\mathbf{y}_1, \ldots, \mathbf{y}_{j-1}\}$. We project $\mathbf{e}_i$ onto it and we obtain the vector $\sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k$. *Independently*, we consider a random gaussian vector $\mathbf{y}_j$ and calculate the inner product

$$\left\langle \sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k, \mathbf{y}_j \right\rangle = \left\langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j \right\rangle.$$

The independence of $\mathbf{y}_j$ with respect to $\mathbf{y}_1, \ldots, \mathbf{y}_{j-1}$ guarantees that

(3.1)

$$G_{i,j}^m = \left\langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j \right\rangle \text{ is distributed like } \|P_{L_{j-1}}(\mathbf{e}_i)\| g_j = \left(\sum_{k=1}^{j-1} \langle \boldsymbol{\nu}_k, \mathbf{e}_i \rangle^2 \right)^{\frac{1}{2}} g_j,$$

where g_j is a standard normal variable, independent of $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{j-1}$, and, therefore, independent also of all of the previous $g_{j'}$, $j' < j$.

Hence, with the notation $\langle \boldsymbol{\nu}_k, \mathbf{e}_i \rangle = \nu_{k,i}$ we have that $\|G_i^m\|^2$ has the same distribution as $\sum_{j=2}^m \sum_{k=1}^{j-1} \nu_{k,i}^2 g_j^2$.

The fact that the factors $\sum_{k=1}^{j-1} \nu_{k,i}^2$ multipying each of the g_j's are not constant and depend on the previous $g_{j'}$'s makes it impossible to apply directly a concentration bound. We circumvent this difficulty by grouping the sum in blocks with a constant factor. This inceases the total sum by a very small amount.

We partition the set $\{2, \ldots, m\}$ in N blocks of size $h = \frac{m-1}{N}$. Then, we have

$$\sum_{j=2}^{m} \sum_{k=1}^{j-1} \nu_{k,i}^2 g_j^2 = \sum_{l=1}^{N} \sum_{j=(l-1)h+2}^{lh+1} \sum_{k=1}^{j-1} \nu_{k,i}^2 g_j^2 \leq \sum_{l=1}^{N} \left(\sum_{j=(l-1)h+2}^{lh+1} g_j^2 \right) \left(\sum_{k=1}^{lh} \nu_{k,i}^2 \right). \tag{3.2}$$

Note that $(\nu_{k,i})_{k=1}^{lh}$ can be seen as the projection of the vector $\mathbf{e_i}$ onto a random Haar distributed subspace of dimension lh. Now we can apply Proposition 2.2,

Proposition 2.4 and the union bound, and we get that, for every $0 < \rho < 1$,

$$\mathcal{G}_{n^2}\left\{\|G_i^m\|^2 \geq \sum_{l=1}^{N}\left(\frac{1}{(1-\rho)}h\right)\left(\frac{1}{(1-\rho)^2}\frac{lh}{n}\right)\right\} \leq 2Ne^{-\frac{\rho^2 h}{4}}.$$

Then, using the union bound on the i's and the definitions of α and N we have that, with probability greater than $1 - 2n\frac{m}{h}e^{-\frac{\rho^2 h}{4}}$,

$$\sup_i \|G_i^m\|^2 \leq \frac{1}{(1-\rho)^3}\frac{h^2}{n}\frac{N(N+1)}{2} \leq \frac{1}{(1-\rho)^3}\frac{\alpha}{2}(m+h). \tag{3.3}$$

Different choices of h, ρ yield now different versions of our result. For instance, we can choose $h = m^{1/2}$, $\rho = m^{-1/5}$ and we have $\|G_i^m\|^2 \leq \frac{\alpha}{2}m + O(m^{4/5})$ with probability $1 - 2n\sqrt{m}e^{-\frac{m^{1/10}}{4}}$. Easy computations show that $\|G_i^m\|^2 \leq \frac{\alpha}{2}m + O(m^{2/3+\epsilon})$ with probability tending to 0 exponentially in m^ϵ.

We can also choose $h = \frac{\epsilon}{2}m$ and $\rho = \frac{\epsilon}{8}$ and, using the Taylor expansion of $\frac{1}{(1-\rho)^3}$, we get that

$$\|G_i^m\|^2 \leq (1+\epsilon)\frac{\alpha}{2}m,$$

with probability greater than $1 - \frac{4n}{\epsilon}e^{-\frac{\epsilon^3 m}{2^9}}$.

For the sake of clarity, we have written Equation (3.2) as if N and h were always integers. If they were not, we can take $N' = [N] + 1$ and $h' = [h] + 1$. This adds at most $[N] + [h] + 1$ terms to the previous sum. It is very easy to see that this extra terms do not change the above estimates.

This upper bounds $\|G_i^m\|$.

Similar reasonings prove the lower bound. To do this, one replaces Equation (3.2) by

$$\sum_{j=2}^{m}\sum_{k=1}^{j-1}\nu_{k,i}^2 g_j^2 \geq \sum_{l=2}^{N}\sum_{j=(l-1)h+2}^{lh+1}\sum_{k=1}^{(l-1)h+1}\nu_{k,i}^2 g_j^2,$$

and proceeds similarly as with the upper bound. □

We proceed now to bound $\|H_i^m\|$.

Proposition 3.2. *With the notation of Theorem 1.1,*

$$\sup_i \|H_i^m\| \leq \sqrt{\left(2 - \alpha/2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}\right)m} + O(m^\delta)$$

and

$$\inf_i \|H_i^m\| \geq \sqrt{\left(2 - \alpha/2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}\right)m} - O(m^\delta),$$

with probability exponentially close to 1.

Proof. Recall that

$$H_i^m = \left((\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})\nu_{j,i}\right)_{j=1}^m,$$

where $\mathbf{y}_j - \Delta_j$ is the projection of $\mathbf{y}_j$ onto the $n-(j-1)$ dimensional subspace orthogonal to the subspace $L_{j-1} = span\{\mathbf{y}_1, \dots, \mathbf{y}_{j-1}\}$. We will first bound the term $(\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 = \left(\|P_{L_{j-1}^\perp}\mathbf{y}_j\| - \sqrt{n}\right)^2$. For that, we need an auxiliar Lemma which we will also use later.

Lemma 3.3. *With the notation above, we have*

(i) *For every $0 < \rho < 1$ and for every $m < n$*

$$\left(\sqrt{n} - \|P_{L_{j-1}^\perp}\mathbf{y}_j\|\right)^2 \leq \left(\sqrt{n} - (1-\rho)^{\frac{1}{2}}\sqrt{n-j+1}\right)^2 \quad \text{for } 1 \leq j \leq m,$$

except for a set Z_1 with $\mathcal{G}_{n^2}(Z_1) \leq m\left(e^{-\frac{\rho^2(n-m+1)}{4}} + e^{-\frac{\rho^2(n-m+1)}{16}}\right)$.

(ii) *Let $0 < \rho_0 < 1$ and $j_0 \in \mathbb{N}$ such that $(1-\rho_0)^{-1}(n-j_0+1) \leq n$. Then, for every $m < n$*

$$\left(\sqrt{n} - \|P_{L_{j-1}^\perp}\mathbf{y}_j\|\right)^2 \geq \left(\sqrt{n} - (1-\rho_0)^{-\frac{1}{2}}\sqrt{n-j+1}\right)^2 \quad \text{for } j_0 < j \leq m,$$

except for a set Z_2 with $\mathcal{G}_{n^2}(Z_2) \leq (m-j_0)e^{-\frac{\rho_0^2(n-m+1)}{4}}$.

Proof. First we prove (i). We choose $\epsilon = \rho/2$ in Equation (2.3) and $\epsilon = \rho$ in Equation (2.3) and we get

$$(1-\rho)^{\frac{1}{2}}\sqrt{n-j+1} - \sqrt{n} \leq \|P_{L_{j-1}^\perp}\mathbf{y}_j\| - \sqrt{n} \leq \left(1-\frac{\rho}{2}\right)^{-\frac{1}{2}}\sqrt{n-j+1} - \sqrt{n}$$

except for a set of measure $e^{\frac{-\rho^2(n-j+1)}{4}} + e^{\frac{-\rho^2(n-j+1)}{16}}$. Using the fact that for $0 < \rho < 1$ $1-(1-\rho)^{\frac{1}{2}} \geq (1-\frac{\rho}{2})^{\frac{-1}{2}} - 1$ we have

$$\left|(1-\frac{\rho}{2})^{\frac{-1}{2}}\sqrt{n-j+1} - \sqrt{n}\right| \leq \left|(1-\rho)^{\frac{1}{2}}\sqrt{n-j+1} - \sqrt{n}\right|.$$

Then, taking squares and applying a union bound we get (i).

The proof of (ii) follows from Equation (2.3), the extra condition on ρ_0 and j_0 and the union bound. □

Now, in order to upper bound $\|H_i^m\|^2$ we first consider the case $\alpha < 1$. As in the proof of Theorem 3.1, we partition the set $\{1, \dots, m\}$ in N blocks of size $h = \frac{m}{N}$.

Then, using Lemma 3.3.(i), we write

$$\begin{aligned}\|H_i^m\|^2 &= \sum_{j=1}^{m} \left(\|\mathbf{y}_j - \Delta_j\| - \sqrt{n}\right)^2 \nu_{j,i}^2 \\ &\leq \sum_{l=1}^{N} \sum_{j=(l-1)h+1}^{lh} \left(\sqrt{n} - (1-\rho)^{\frac{1}{2}}\sqrt{n-lh+1}\right)^2 \nu_{j,i}^2 \\ &= \sum_{l=1}^{N} \left(\sqrt{n} - (1-\rho)^{\frac{1}{2}}\sqrt{n-lh}\right)^2 \sum_{j=(l-1)h+1}^{lh} \nu_{j,i}^2,\end{aligned} \tag{3.4}$$

outside of Z_1.

On the other hand, considering $(\nu_{k,i})_{i=(l-1)h+1}^{lh}$ as the projection of $\mathbf{e}_i$ onto a random subspace of dimension lh, Proposition 2.4 tells us that, for every $1 \leq i \leq n$ and $1 \leq l \leq N$,

$$\sum_{j=(l-1)h+1}^{lh} \nu_{j,i}^2 \leq \frac{1}{(1-\rho')^2}\frac{h}{n}$$

except for a set Z_1' with $\mathcal{G}_{n^2}(Z_1') \leq nNe^{-\frac{\rho^2 h}{4}}$.

So, we have that, outside of $Z_1 \cup Z_1'$,

$$\begin{aligned}\|H_i^m\|^2 &\leq \frac{1}{(1-\rho')^2}\frac{h}{n}\sum_{l=1}^{N}\left(\sqrt{n} - (1-\rho)^{\frac{1}{2}}\sqrt{n-lh}\right)^2 \\ &= \frac{1}{(1-\rho')^2}\frac{h}{n}\left[nN + (1-\rho)\left(nN - h\frac{N(N+1)}{2}\right)\right. \\ &\qquad \left. - 2(1-\rho)^{\frac{1}{2}}\sqrt{n}\sum_{l=1}^{N}\sqrt{n-lh}\right].\end{aligned}$$

We can bound

$$\sum_{l=1}^{N}\sqrt{n-lh} \geq \int_1^N \sqrt{n-xh}dx = \frac{2}{3h}\left((n-h)^{3/2} - (n-Nh)^{3/2}\right).$$

Then, putting this together with the definitions of α and N, we get that

$$\begin{aligned}\|H_i^m\|^2 \leq \frac{1}{(1-\rho')^2}m\left[1 + (1-\rho)\left(1 - \frac{\alpha}{2} - \frac{\alpha h}{2m}\right)\right. \\ \left. -(1-\rho)^{\frac{1}{2}}\frac{4}{3\alpha}\left(\left(1-\frac{\alpha h}{m}\right)^{3/2} - (1-\alpha)^{3/2}\right)\right],\end{aligned}$$

with probability greater than

$$1 - \frac{m^2}{\alpha h} e^{-\frac{\rho^2 h}{4}} - m\left(e^{-\frac{\rho^2(n-m)}{4}} + e^{-\frac{\rho^2(n-m)}{16}}\right). \tag{3.5}$$

Again, different choices of h, ρ, ρ' yield now different versions of our result. For instance, taking $h = m^{1/2}$, $\rho = \rho' = m^{-1/5}$ we get

$$\|H_i^m\|^2 \leq \left(2 - \alpha/2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}\right) m + O(m^{4/5}),$$

with probability tending to one exponentially in $m^{1/10}$. Easy computations also show that $\|G_i^m\|^2 \leq \frac{\alpha}{2} m + O(m^{2/3+\epsilon})$ with probability tending to 0 exponentially in m^ϵ.

The reasonings above do not apply directly to the case $\alpha = 1$, as the bound of the probability in Equation (3.5) becomes trivial in that case. To overcome this issue, in case $\alpha = 1$ we consider $h = \frac{n-\sqrt{n}}{N}$ and rewrite Equation (3.4) as

$$\begin{aligned}
\|H_i^m\|^2 &\leq \sum_{j=1}^{n} \left(\|\mathbf{y}_j - \Delta_j\| - \sqrt{n}\right)^2 \nu_{j,i}^2 \\
&= \sum_{j=1}^{n-\sqrt{n}} \left(\|\mathbf{y}_j - \Delta_j\| - \sqrt{n}\right)^2 \nu_{j,i}^2 + \sum_{j=n-\sqrt{n}+1}^{n} \left(\|\mathbf{y}_j - \Delta_j\| - \sqrt{n}\right)^2 \nu_{j,i}^2 \\
&\leq \sum_{l=1}^{N} \sum_{j=(l-1)h+1}^{lh} \left(\sqrt{n} - (1-\rho)^{\frac{-1}{2}} \sqrt{n - lh + 1}\right)^2 \nu_{j,i}^2 + \sum_{j=n-\sqrt{n}+1}^{n} n\nu_{j,i}^2,
\end{aligned}$$

outside of the set Z_1 defined in Lemma 3.3 in the case $m = n - \sqrt{n}$.

The first summand is treated as previously where now $m = n - \sqrt{n}$. We note that, using Proposition 2.4 and the union bound once again, the second summand verifies, with probability greater than $1 - ne^{-\frac{\rho^2\sqrt{n}}{4}}$,

$$n \sum_{j=n-\sqrt{n}+1}^{n} \nu_{j,i}^2 \leq n \frac{1}{(1-\rho^2)\sqrt{n}} = \frac{\sqrt{n}}{(1-\rho^2)}.$$

This only adds an $O(\sqrt{n})$ term which does not modify the result. This finishes the proof of the upper bound.

For the lower bound we reason similarly. Consider j_0, ρ_0 as in Lemma 3.3.(ii). Then, with probability $1 - n(m-j_0)e^{-\frac{\rho_0^2(n-m+1)}{4}}$, for every $1 \leq i \leq n$ we have that

$$\begin{aligned} \|H_i^m\|^2 &\geq \sum_{j=j_0+1}^{m} \left(\|\mathbf{y}_j - \Delta_j\| - \sqrt{n}\right)^2 \nu_{j,i}^2 \\ &\geq \sum_{j=j_0+1}^{m} \left(\sqrt{n} - (1-\rho_0)^{-1/2}\sqrt{n-j+1}\right)^2 \nu_{j,i}^2. \end{aligned} \tag{3.6}$$

Partitioning the set $\{j_0+1, \ldots, m\}$ in N blocks of size $h = \frac{m-j_0}{N}$ and using similar reasonings to those used for the upper bound in the case $\alpha < 1$ we obtain

$$\begin{aligned} \inf_i \|H_i^m\|^2 \geq (1-\rho')^2(m-j_0) \Bigg[1 + (1-\rho_0)\left(1 - \frac{\alpha}{2} + \frac{\alpha(j_0+h)}{2m}\right) \\ - \frac{4(1-\rho_0)^{\frac{1}{2}}}{3(\alpha - \frac{\alpha j_0}{m})}\left(\left(1 + \frac{\alpha h}{m}\right)^{3/2} - \left(1 - \alpha + \frac{\alpha j_0}{m}\right)^{3/2}\right)\Bigg], \end{aligned}$$

with probability higher than $1 - \frac{m}{\alpha}(m-j_0)e^{-\frac{\rho_0^2(n-m+1)}{4}} - \frac{h}{\alpha}e^{-\frac{\rho'^2 h}{4}}$. As in expressions (3.3) and (3.5) different values of j_0, ρ_0, ρ and h give different bounds.

The case $\alpha = 1$ can be treated as in the upper bound. The terms in (3.6), where $n - \sqrt{n} + 1 \leq j \leq n$ can only add up to $O(\sqrt{n})$ and the rest of the terms can be bounded as before.

Note that, as in the proof of Proposition 3.1, we are assuming that N, h and $\sqrt{n}$ are integers. If this is not the case we can consider $N' = [N] + 1$ and $h' = [h] + 1$ for the upper estimates ($N' = [N]$ and $h' = [h]$ for the lower estimates) and $[\sqrt{n}]$ for the case $\alpha = 1$. Adding or subtracting these extra terms will give negligible quantities compared with the sums. This finishes the proof. □

We now need to prove that $\langle G_i^m, H_i^m \rangle$ is negligible when compared with $\|G_i^m\|^2$ and $\|H_i^m\|^2$. More precisely, we will use similar techniques to show that $\langle G_i^m, H_i^m \rangle$ is, with probability exponentially close to 1, of smaller order in m than $\|G_i^m\|^2$ and $\|H_i^m\|^2$. As shown in Proposition 3.1 and Proposition 3.2 above, each of them is $\Theta(m)$. That is, there exist constants k_1, k_2 and m_0 such that for all $m > m_0$, we have that $k_1 m \leq \|G_i^m\|^2 \leq k_2 m$ and analogously for $\|H_i^m\|^2$.

Proposition 3.4. *With the notation of Theorem 1.1, given $\epsilon > 0$ we have*

$$\langle G_i^m, H_i^m \rangle = O(m^{\frac{1}{2}+\epsilon}),$$

with probability exponentially close to 1.

This proposition, together with Propositions 3.1 and 3.2, finishes the proof of Theorem 1.1.

For the sake of clarity we will first show the following technical lemma that will be used in the proof of Proposition 3.4.

Lemma 3.5. *Let $\mathbf{y}_j$ be a gaussian vector and L_{j-1} a Haar distributed subspace of dimension $j-1$, then*

$$Pr\left(\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle^2 > \frac{j-1}{n} m^\epsilon\right) \leq 2e^{-\frac{m^{\epsilon/2}-2}{4}}.$$

Proof. First of all note that $\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle = \langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j\rangle$. We have already shown in Equation (3.1) that $\langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j\rangle^2$ has the same distribution as the term $\|P_{L_{j-1}}(\mathbf{e}_i)\|^2 g_j^2$, where g_j is a standard normal variable. Putting together Lemma 2.1, Proposition 2.4 and the union bound, we get

$$Pr\left(\langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j\rangle^2 > \frac{j-1}{n} m^\epsilon\right) \leq e^{-\frac{j-1}{4}(m^{\epsilon/2}-2)} + e^{-\frac{m^{\epsilon/2}}{2}} \leq 2e^{-\frac{m^{\epsilon/2}-2}{4}}.$$

□

We will also need Hoeffding's inequality [10].

Proposition 3.6 (Hoeffding's inequality)**.** *Let $(X_i)_{i=1}^n$ be a family of independent random variables such that $a_i \leq X_i \leq b_i$ for $i = 1, ..., n$. Let $S = \sum_{i=1}^n X_i$. Then, for every $a > 0$,*

$$Pr(|S - \mathbb{E}(S)| > a) \leq 2e^{-\frac{2a^2}{\sum_i (b_i - a_i)^2}}.$$

Proof of Proposition 3.4. Recall that we have

$$G_i^m = \left(\sum_{k=1}^{j-1} \langle \mathbf{y}_j, \boldsymbol{\nu}_k\rangle \langle \boldsymbol{\nu}_k, \mathbf{e}_i\rangle\right)_{j=1}^m = \left(\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle\right)_{j=1}^m$$

and

$$H_i^m = \left((\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})\langle \boldsymbol{\nu}_j, \mathbf{e}_i\rangle\right)_{j=1}^m = \left((\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n})\langle \boldsymbol{\nu}_j, \mathbf{e}_i\rangle\right)_{j=1}^m.$$

Therefore,

(3.7)

$$\begin{aligned}\langle G_i^m, H_i^m\rangle &= \sum_{j=1}^m \langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle \left(\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n}\right)\langle \boldsymbol{\nu}_j, \mathbf{e}_i\rangle \\ &= \sum_{j=1}^m \left|\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle\right| \left(\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n}\right)\langle \boldsymbol{\nu}_j, \mathbf{e}_i\rangle \operatorname{sign}\left(\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle\right).\end{aligned}$$

We claim that the probability distribution of the previous expression is the same as the probability distribution of

$$\sum_{j=1}^{m} \left|\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle\right| \left(\|P_{\tilde{L}_{j-1}^{\perp}}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}\right) \langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle \epsilon_j, \tag{3.8}$$

where $\tilde{\mathbf{y}}_1, \cdots, \tilde{\mathbf{y}}_n$ are independent gaussian vectors, $\tilde{\boldsymbol{\nu}}_1, \cdots, \tilde{\boldsymbol{\nu}}_n$ are the corresponding vectors obtained from the Gram-Schmidt orthonormalization procedure and $\epsilon_1, \cdots, \epsilon_n$ are independent and identically distributed Bernouilli variables taking values in an independent probability space.

In order to see this, let us consider the space $\mathbb{R}^n \times \overset{(n)}{\cdots} \times \mathbb{R}^n$ with the Gaussian measure $\mathcal{G}_n$ on each $\mathbb{R}^n$. For each j we denote by $\mathbf{z}_j = (z_{k,j})_{k=1}^n$ the gaussian vector in the corresponding copy of $\mathbb{R}^n$. For each $j \geq 2$ we consider in $\mathbb{R}^n$ the equivalence relation $\mathbf{z} \sim_j \mathbf{z}'$ if and only if $(z_1, \ldots, z_{j-1}) = \pm(z'_1, \ldots, z'_{j-1})$ and $(z_j, \ldots, z_n) = (z'_j, \ldots, z'_n)$. Then, $\mathbb{R}^n = (\mathbb{R}^n / \sim_j) \times \{-1, 1\}$, with the identification $\mathbf{z}_j = ([\mathbf{z}_j], \sigma_j)$, with $\sigma_j \in \{-1, 1\}$. We define the probability measure $\mathcal{G}'_n$ on $\mathbb{R}^n / \sim_j$ by the density $f'([\mathbf{z}_j]) = 2f(\mathbf{z}_j)$, where f is the density of $\mathcal{G}_n$, and we call μ the uniform probability on $\{-1, 1\}$. We clearly have $\mathcal{G}_n = \mathcal{G}'_n \otimes \mu$.

Let us now consider a family of independent gaussian vectors $(\mathbf{z}_1, \cdots, \mathbf{z}_n)$ in the previous probability space. For $j = 1$ there is no equivalence relation, and we define $\tilde{\mathbf{y}}_1 = \mathbf{z}_1$, which is clearly a gaussian vector. Consequently, we define $\tilde{\boldsymbol{\nu}}_1 = \frac{1}{\|\tilde{\mathbf{y}}_1\|} \tilde{\mathbf{y}}_1$.

Now, for each $2 \leq j \leq m$, we consider $\tilde{L}_{j-1}$, the random $(j-1)$-dimensional subspace spanned by $\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{j-1}$. The vectors $\tilde{\boldsymbol{\nu}}_1, \ldots, \tilde{\boldsymbol{\nu}}_{j-1}$ form an orthonormal basis of $\tilde{L}_{j-1}$. Hence, we can complete this set to obtain a basis of $\mathbb{R}^n$, $\{\tilde{\boldsymbol{\nu}}_1, \ldots, \tilde{\boldsymbol{\nu}}_{j-1}, \boldsymbol{\nu}^*_j, \ldots, \boldsymbol{\nu}^*_n\}$. The added vectors needed to complete the orthonormal basis can be chosen at will. In general they will change as j changes.

Let us denote by U_j the orthogonal matrix whose columns are the vectors of the previous basis. Then, given $\mathbf{z}_j$, we define $\tilde{\mathbf{y}}_j = U_j \mathbf{z}_j$. Since the orthogonal matrix U_j is independent of the Gaussian vector $\mathbf{z}_j$, we immediately deduce that $\tilde{\mathbf{y}}_j$ is a Gaussian vector independent of $\tilde{\mathbf{y}}_1, \cdots, \tilde{\mathbf{y}}_{j-1}$.

From the Gram-Schmidt orthogonalization procedure we have that

$$\tilde{\boldsymbol{\nu}}_j = \frac{\tilde{\mathbf{y}}_j - \sum_{k=1}^{j-1} \langle \tilde{\mathbf{y}}_j, \tilde{\boldsymbol{\nu}}_k \rangle \tilde{\boldsymbol{\nu}}_k}{\|\tilde{\mathbf{y}}_j - \sum_{k=1}^{j-1} \langle \tilde{\mathbf{y}}_j, \tilde{\boldsymbol{\nu}}_k \rangle \tilde{\boldsymbol{\nu}}_k\|}.$$

It follows immediately that

$$P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j) = \sum_{k=1}^{j-1} z_{k,j} \tilde{\boldsymbol{\nu}}_k \quad \text{and} \quad P_{\tilde{L}_{j-1}^{\perp}}(\tilde{\mathbf{y}}_j) = \sum_{k=j}^{n} z_{k,j} \boldsymbol{\nu}^*_k,$$

where we recall that $\mathbf{z}_j = (z_{k,j})_{k=1}^n$.

With the identification $\mathbf{z}_j = ([\mathbf{z}_j], \sigma_j)$, it is easy to see that $\big|\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle\big|$ and $\big(\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}\big)\langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle$ do not depend on $\sigma_2, \ldots, \sigma_n$ (or, equivalently, they only depend on the variables $[\mathbf{z}_j]$). Indeed, to see this we notice first that it follows from the definitions that the vectors $\tilde{\boldsymbol{\nu}}_j$ are independent of $\sigma_2, \cdots, \sigma_n$. Next, we notice that the dependence of $\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle$ with respect to $\sigma_2, \ldots, \sigma_n$ is cancelled out by the absolute value.

Hence, expression (3.7) applied to the independent Gaussian vectors $\tilde{\mathbf{y}}_1, \cdots, \tilde{\mathbf{y}}_n$ has the form

$$\sum_{j=1}^m \big|\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle\big|\big(\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}\big)\langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle \epsilon_j,$$

where $\tilde{\mathbf{y}}_1, \cdots, \tilde{\mathbf{y}}_n$ are Gaussian vectors independent of $\sigma_2, \ldots, \sigma_n$ and $\epsilon_j = \epsilon_j(\sigma_j)$ are independent identically distributed Bernouilli variables.

Equation (3.8) will allow us to apply Proposition 3.6: For fixed $(\mathbf{z}_1, [\mathbf{z_2}], \ldots, [\mathbf{z}_n])$ we can consider the independent random variables (function of $(\sigma_2, \cdots, \sigma_n)$)

$$\Big(\big|\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle\big|\big(\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}\big)\langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle \epsilon_j(\sigma_j)\Big)_{j=1}^m,$$

Then, Proposition 3.6 gives us that, for fixed $(\mathbf{z}_1, [\mathbf{z_2}], \ldots, [\mathbf{z}_n])$.

$$\mu^{\otimes m}\Big(|\langle G_i^m, H_i^m\rangle| \geq a\Big) \leq 2e^{-\frac{a^2}{2\sum_{j=1}^m \langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle^2 (\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n})^2 \langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle^2}}. \tag{3.9}$$

We consider first the case $\alpha < 1$.

It follows from Lemma 3.5, Proposition 2.4, Lemma 3.3 and a union bound argument that

$$\begin{aligned} 2\sum_{j=1}^m \langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle^2 \big(\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}\big)^2 \langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle^2 \leq \\ \leq 2\sum_{j=1}^m m^\epsilon \frac{j-1}{n}\big(\sqrt{n} - (1-\rho)^{\frac{1}{2}}\sqrt{n-j+1}\big)^2 \frac{m^\epsilon}{n} \leq m^{2\epsilon+1} \end{aligned} \tag{3.10}$$

with probability larger than

$$1 - C := 1 - 2me^{-(\frac{m^{\frac{\epsilon}{2}}-2}{4})} + m\left(e^{-\frac{\rho^2((1/\alpha-1)m+1)}{4}} + e^{-\frac{\rho^2((1/\alpha-1)m+1)}{16}}\right) + \sqrt{e}me^{\frac{1}{4}m^{2\epsilon}}.$$

(The second inequality in the second line of (3.10) follows from easy calculations).

We notice now that $Pr\Big(|\langle G_i^m, H_i^m\rangle| \geq a\Big)$ is upper bounded by

$$Pr\Big(|\langle G_i^m, H_i^m\rangle| \geq a \,\Big|\, 2\sum_{j=1}^{m}\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle^2\big(\|P_{\tilde{L}_{j-1}^{\perp}}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}\big)^2\langle\tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle^2 \leq m^{2\epsilon+1}\Big)$$
$$+Pr\Big(2\sum_{j=1}^{m}\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i\rangle^2\big(\|P_{\tilde{L}_{j-1}^{\perp}}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}\big)^2\langle\tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i\rangle^2 > m^{2\epsilon+1}\Big),$$

where we denote by $Pr(A|B)$ the probability of the event A conditioned to B.

We pick $\epsilon' > \epsilon$ and we fix $a = m^{\frac{1}{2}+\epsilon'}$. Then, Equations (3.9) and (3.10) imply that

$$Pr\Big(|\langle G_i^m, H_i^m\rangle| \geq m^{\frac{1}{2}+\epsilon'}\Big) \leq 2e^{-\frac{m^{1+2\epsilon'}}{m^{2\epsilon+1}}} + C = 2e^{m^{-2(\epsilon'-\epsilon)}} + C,$$

which tends exponentially fast to zero as n grows to infinity.

The case $\alpha = 1$ has to be considered separately as the bound in the concentration of Lemma 3.3 becomes trivial. In order to overcome this issue we reason as in the proof of Proposition 3.2. We can divide the sum in (3.10) in two terms

$$2\sum_{j=1}^{n-\sqrt{n}}\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle^2\left(\|P_{L_{j-1}^{\perp}}(\mathbf{y}_j)\| - \sqrt{n}\right)^2\langle\boldsymbol{\nu}_j, \mathbf{e}_i\rangle^2$$
$$+2\sum_{j=n-\sqrt{n}+1}^{n}\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i\rangle^2\left(\|P_{L_{j-1}^{\perp}}(\mathbf{y}_j)\| - \sqrt{n}\right)^2\langle\boldsymbol{\nu}_j, \mathbf{e}_i\rangle^2,$$

where the first summand is treated as previously giving an upper bound of $(n - \sqrt{n})^{2\epsilon+1}$ and the second is $O(n^{1/2+2\epsilon})$, which is negligible compared with the first. Proceeding as in the case $\alpha < 1$ the result follows. □

4. The supremum norm

In this section we use Theorem 1.1 to prove Theorem 1.2, that describes the asymptotic probabilistic behaviour of $\epsilon_n(m) = \sup_{1\leq i\leq n, 1\leq j\leq m}|y_{i,j} - \sqrt{n}u_{i,j}|$.

If we choose $m_n = \frac{\beta n}{\log n}$ or $m_n = o\left(\frac{n}{\log n}\right)$ we immediately obtain Corollary 4.2. This result should be compared with [11, Theorem 3]: In it, Jiang showed that if Y_n is a gaussian random matrix and U_n its Gram-Schmidt orthogonalization, then $\epsilon_n(m)$ converges to 0 in probability if and only if $m_n = o\left(\frac{n}{\log n}\right)$, and he also showed that if $m_n = \frac{\beta n}{\log n}$ then $\epsilon_n(m)$ converges in probability to $2\sqrt{\beta}$. Our Corollary 4.2 shows the existence of couplings between a gaussian matrix Y_n and a Haar distributed orthogonal matrix U_n such that $\epsilon_n(m)$ also converges to 0 in

probability if and only if $m_n = o\left(\frac{n}{\log n}\right)$ but now, when $m_n = \frac{\beta n}{\log n}$, the upper bound for $\epsilon_n(m)$ converges in probability to $\sqrt{2\beta}$.

Before we start our reasonings, we state and prove for completeness a lemma which is well known, but for which we have not found an explicit reference.

Lemma 4.1. *Let $\{\mathbf{w_j} = (w_{ij})_{i=1}^n\}_{j=1}^m$ be m unitary vectors each of them randomly uniformly distributed in the sphere of $\mathbb{R}^n$. Then, for any $\epsilon > 0$ we have*

$$Pr\left(\sup_{i,j}|w_{i,j}| > (1+\epsilon)\frac{\sqrt{2\log(nm)}}{\sqrt{n}}\right) \xrightarrow{n} 0.$$

If now $\mathbf{w} = (w_i)_{i=1}^n$ is an unitary vector randomly uniformly distributed in the sphere of $\mathbb{R}^n$,

$$Pr\left(\sup_{i=1,\dots,n}|w_i| < (1-\epsilon)\frac{\sqrt{2\log n}}{\sqrt{n}}\right) \xrightarrow{n} 0.$$

Proof. In order to prove the first expression, we consider the function that projects a unitary vector in $\mathbb{R}^n$ onto its ith coordinate. This function has Lipschitz constant 1 and its median is 0. Thus, a straightforward consequence of Levy's lemma [14] shows that, for $1 \le i \le n$ and $1 \le j \le m$,

$$Pr\left(|w_{i,j}| > t\right) \le \sqrt{\frac{\pi}{2}}e^{-(n-1)t^2/2}.$$

Taking $t = (1+\epsilon)\frac{\sqrt{2\log(nm)}}{\sqrt{n}}$ and applying a union bound we get

$$Pr\left(\sup_{i,j}|w_{i,j}| > (1+\epsilon)\frac{\sqrt{2\log(nm)}}{\sqrt{n}}\right) \le \sqrt{\frac{\pi}{2}}(nm)^{-\epsilon\frac{n-1}{n}+\frac{1}{n}} \xrightarrow{n} 0.$$

For the proof of the second expression we first consider a gaussian vector $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. It follows from the independence of the coordinates of $\mathbf{x}$ and the lower bound on Lemma 2.1 that, for any $t > 0$,

$$Pr\left(\sup_{i=1,\dots,n}|x_i| < t\right) = \left(Pr\left(|x_i| < t\right)\right)^n \le \left(1 - \frac{2}{\sqrt{2\pi}}\frac{t}{1+t^2}e^{-t^2/2}\right)^n.$$

Now we fix $0 < \rho < 1$. If we choose $t = (1-\rho)\sqrt{2\log n}$ in the equation above and we take limits as n grows to infinity, we have that

$$Pr\left(\sup_{i=1,\dots,n}|x_i| < (1-\rho)\sqrt{2\log n}\right) \xrightarrow{n} 0. \tag{4.1}$$

We now use the fact that a normalized gaussian vector $\mathbf{x}$ distributes like a uniform unitary vector $\mathbf{w}$. Therefore, for every $t, s > 0$ we have that

$$\begin{aligned} &Pr\left(\sup_i |w_i| < t\right) = Pr\left(\frac{\sup_i |x_i|}{\|\mathbf{x}\|} < t\right) \\ &\leq Pr\left(\left\{\sup_i |x_i| < ts\right\} \cap \left\{\|\mathbf{x}\| \leq s\right\}\right) + Pr\left(\|\mathbf{x}\| > s\right) \\ &\leq Pr\left(\sup_i |x_i| < ts\right) + Pr\left(\|\mathbf{x}\| > s\right). \end{aligned}$$

We fix $t = (1-\epsilon)\frac{\sqrt{2\log n}}{\sqrt{n}}$, $s = \frac{\sqrt{n}}{\sqrt{1-\delta}}$ and we apply Proposition 2.2 to get that, for $0 < \delta < 1$, $0 < \epsilon < 1$, with $\frac{(1-\epsilon)}{\sqrt{1-\delta}} < 1$,

$$\begin{aligned} &Pr\left(\sup_i |w_i| < (1-\epsilon)\frac{\sqrt{2\log n}}{\sqrt{n}}\right) \\ &\leq Pr\left(\sup_i |x_i| < \frac{(1-\epsilon)}{\sqrt{1-\delta}}\sqrt{2\log n}\right) + e^{-\delta^2 n/4}. \end{aligned}$$

According to Equation (4.1) this last expression tends to zero when n grows to infinity, so this concludes the proof. □

Theorem 1.1 gives us control over the euclidean norm of the rows F_i^m of $Y_n - \sqrt{n}U_n$. We will use these estimates to obtain information about the supremum

$$\epsilon_n(m) = \sup_{1\leq i\leq n, 1\leq j\leq m} |y_{i,j} - \sqrt{n}u_{i,j}|.$$

First of all we notice that U_n is the Gram-Schmidt orthogonalization of Y_n. Therefore, the columns of $Y_n - \sqrt{n}U_n$ are not equally distributed. For instance, it is very easy to see that, with very high probability, their euclidean norm is strictly increasing. In turn, this implies that the coordinates of each of the F_i^m are not equally distributed. To avoid this problem, we will randomly choose a slightly better coupling that the one given by the Gram-Schmidt orthogonalization procedure.

Proof of Theorem 1.2: Let Y_n, U_n be as in Theorem 1.1. We consider a Haar distributed orthogonal random matrix $V_m \in \mathcal{O}(m)$. We define the orthogonal matrix $V = (v_{i,j})_{i,j=1}^n \in \mathcal{O}(n)$ by

$$V = \begin{pmatrix} V_m & 0 \\ 0 & I_{m-n} \end{pmatrix}.$$

We now define $Y' = YV$, $U' = UV$. Due to the orthogonal invariance of both the gaussian distribution and the Haar distribution, we have that Y' is a random

gaussian matrix and U' is Haar distributed in the orthogonal group. Note that U' is *not* the Gram-Schmidt orthogonalization of Y'.

We have now that $Y'_n - \sqrt{n}U'_n = (Y_n - \sqrt{n}U_n)V$. Call $F_{i,j}$ to the j^{th} coordinate of the vector F_i^m defined as in Theorem 1.1. Then the first m coordinates of the i^{th} row of $Y'_n - \sqrt{n}U'_n$ form the vector $\mathbf{x}_i = (x_{i,j})_{j=1}^m$, where

$$x_{i,j} = \sum_{k=1}^{m} F_{i,k} v_{k,j}.$$

Therefore $\mathbf{x}_i \in \mathbb{R}^m$ is a vector whose direction is uniformly random and it verifies $\|\mathbf{x}_i\| = \|F_i^m\|$. That is, for every $1 \leq i \leq n$, $\frac{\mathbf{x}_i}{\|F_i^m\|}$ is a unitary vector uniformly distributed.

We will first prove the upper bound of $\epsilon_n(m)$. It follows from the first part of Lemma 4.1 that, for every $t > 0$,

$$Pr\left(\sup_{i,j} \frac{|x_{i,j}|}{\|F_i^m\|} > (1+\varepsilon)\frac{\sqrt{2\log(nm)}}{\sqrt{n}} \right) \xrightarrow{n} 0.$$

We have that

$$Pr\left(\epsilon_n(m) > (1+\varepsilon) \sup_i \|F_i^m\| \frac{\sqrt{2\log(nm)}}{\sqrt{n}} \right) \xrightarrow{n} 0. \tag{4.2}$$

We recall that, according to Theorem 1.1, there exists $0 < \delta < \frac{1}{2}$ such that

$$\sup_i \|F_i^m\| \leq \sqrt{\left(2 - \frac{4}{3}\frac{(1-(1-\alpha)^{3/2})}{\alpha}\right) m} + O(m^{\frac{1}{2}-\delta}) := \sqrt{\varphi(\alpha)m} + O(m^{\frac{1}{2}-\delta})$$

with probability exponentially close to 1. Putting this together with Equation (4.2) we get the upper bound.

For the lower bound we consider the columns of the matrix $(x_{i,j})_{i=1,j=1}^{n,m}$. This matrix is invariant under the action of the orthogonal group on the left, hence $\tilde{\mathbf{x}}_j = (x_{i,j})_{i=1}^n$ is a vector whose direction is uniformly random. We will bound the probability of $\epsilon_n(m)$ being small by the probability of the coordinates of $\tilde{\mathbf{x}}_j$ being small. It is clear that $\epsilon_n(m) = \sup_{i,j} |x_{ij}| \geq \sup_i |x_{ij_0}|$, where j_0 is the column with the largest norm, that is, $\|\tilde{\mathbf{x}}_{j_0}\| = \sup_j \|\tilde{\mathbf{x}}_j\|$. Thus, for any $\epsilon > 0$,

$$Pr\left(\epsilon_n(m_n) \geq (1-\varepsilon) \sup_j \|\tilde{\mathbf{x}}_j\| \sqrt{\frac{2\log n}{n}} \right) \geq Pr\left(\sup_i |x_{ij_0}| \geq \|\tilde{\mathbf{x}}_{j_0}\|(1-\varepsilon)\sqrt{\frac{2\log n}{n}} \right),$$

which tends to one as n grows to infinity. Here, the last inequality follows from considering the vector $\frac{\tilde{\mathbf{x}}_{j_0}}{\|\tilde{\mathbf{x}}_{j_0}\|}$ in Lemma 4.1.

Moreover, we have that

$$\|\tilde{\mathbf{x}}_{j_0}\| = \sup_j \|\tilde{\mathbf{x}}_j\|^2 \geq \frac{1}{m}\sum_{j=1}^{m}\|\tilde{\mathbf{x}}_j\|^2 = \frac{1}{m}\sum_{i=1}^{n}\|\mathbf{x}_i\|^2 = \frac{1}{m}\sum_{i=1}^{n}\|F_i^m\|^2 \geq \frac{n}{m}\inf_i \|F_i^m\|^2.$$

Putting this together with the lower bound of $\inf_i \|F_i^m\|$ from Theorem 1.1 (which happens with probability exponentially close to 1) we get

$$Pr\left(\epsilon_n(m) \geq (1-\varepsilon)\left(\sqrt{\varphi(\alpha)} - O(m^{-\delta})\right)\sqrt{2\log n}\right) \xrightarrow{n} 1,$$

which finishes the proof. □

Remark 4.1. We expect the lower bound of Theorem 1.2 to be $\sqrt{\varphi(\alpha)}\sqrt{2\log(nm)}$. To prove that, one needs to overcome the lack of independence of the rows F_i^m.

In our previous results, $\alpha = \frac{m}{n}$ was a constant number. For our next result we need to apply Theorem 1.2 in the case of a non constant ratio α_n convergent to 0. It is very easy to see that this can be done:

If nm_n grows to infinity, for example if $m_n \geq \frac{1}{\sqrt{n}}$, it is easy to check that Theorem 1.2 remains valid. In case $m_n = \frac{1}{\sqrt{n}}$ we have that $\epsilon_n(m)$ in Theorem 1.2 converges to 0. This makes Theorem 1.2 also true if $m_n = O\left(\frac{1}{\sqrt{n}}\right)$.

Now we can state and prove the announced improvement of [11, Theorem 3].

Corollary 4.2. *For each $n \geq 2$, there exists matrices $Y_n' = (y_{ij}')_{i,j=1}^n$ and $U_n' = (u_{i,j}')_{i,j=1}^n$ whose $2n^2$ entries are real random variables defined on the same probability space such that*

(i) *the law of U_n' is the normalized Haar measure on the orthogonal group $\mathcal{O}(n)$;*
(ii) *$\{y_{i,j}'; 1 \leq i,j \leq n\}$ are independent standard normals;*
(iii) *set*

$$\epsilon_n(m) = \max_{1\leq i\leq n,\, 1\leq j\leq m} |\sqrt{n}u_{i,j}' - y_{i,j}'|$$

for $m = 1, 2, \cdots, n$. Then $\epsilon_n(m) \to 0$ in probability as $n \to \infty$ provided $m_n = o(\frac{n}{\ln n})$ as $n \to \infty$. Moreover, if we make $m_n = \frac{\beta n}{\log n}$ then we have that, for every $\varepsilon > 0$, $\epsilon_n(m)$ belongs to the interval

$$(\sqrt{\beta} - \varepsilon, \sqrt{2\beta} + \varepsilon)$$

with probability $1 - o(1)$.

Proof. Let Y_n', U_n' be as in Theorem 1.2. Suppose first that $\alpha_n = \frac{m_n}{n} = \frac{c_n}{\ln n}$, where $(c_n)_{n\in\mathbb{N}}$ is sequence converging to 0. We consider the Taylor expansion of

$\varphi(\alpha_n) = \left(2 - \frac{4}{3}\frac{(1-(1-\alpha_n)^{3/2})}{\alpha_n}\right)$ in Theorem 1.2. Then, there exists $0 < \delta < \frac{1}{2}$ such that for any $\varepsilon', \varepsilon'' > 0$ and for n large enough we have

$$\epsilon_n(m) \leq (1+\varepsilon')\left(\sqrt{\frac{\alpha_n}{2} + \varepsilon''} + O(m_n^{-\delta})\right)\sqrt{2\log(nm_n)} \leq (1+\varepsilon')\sqrt{2c_n + 2\varepsilon''}$$

with probability tending to 1. This proves the $o(\frac{n}{\log n})$ statement of part (iii). Choosing $c_n = \beta$ for every n we get the upper bound of the $\frac{\beta n}{\log n}$ statement. For the lower bound, we reason similarly using the lower bound of Theorem 1.2. □

5. A non-universality result

In [8] we used Theorem 1.1 as the main technical tool to solve a question in Quantum Information Theory related to the probability of finding *classical correlations* among *quantum correlations*. Part (a) of [8, Theorem 0.1] can be interpreted as a non-universality result distinguishing gaussian and Bernouilli matrices. The precise statement is Example 1.1. In this section we briefly sketch its proof. For a detailed exposition, the reader is referred to [8].

We consider a gaussian random matrix G of order n. We consider also the orthogonal matrices U and V of its left and right singular values respectively. It follows from the bi-orthogonally invariance of the gaussian distribution that U and V are independent from each other and Haar distributed. The singular values of G are distributed according to the Marcenko-Pastur law. Following [1] we consider the m biggest singular values of G, and the $n \times m$ matrices U', V', submatrices of U, V respectively, formed by the right and left singular vectors corresponding to those biggest m singular values. It follows from [1, Theorem 1] that

$$\frac{n}{m}\sum_{i,j=1}^{n} G_{i,j}\left(U'V'^T\right)_{i,j} \geq (2 - \epsilon + o(1))n^{\frac{3}{2}} \tag{5.1}$$

with probability $1 - o(1)$, where m is the number of singular values of G which are bigger than $(2-\epsilon)\sqrt{n}$. For a fixed $0 < \epsilon < 2$ the Marcenko-Pastur law states that the quotient $\frac{m}{n}$ converges to the fixed number $\frac{1}{2\pi}\int_{(2-\epsilon)^2}^{4}\sqrt{\frac{4}{x} - 1}dx$.

On the other hand, it follows from [1, Theorem 4] that

$$\sup_{a_i=\pm 1, b_j=\pm 1}\sum_{i,j=1}^{n} G_{i,j}a_i b_j \leq 1.6652 n^{\frac{3}{2}}, \tag{5.2}$$

with probability $1 - o(1)$. Equations (5.1) and (5.2) together imply that, with probability $1 - o(1)$, for two independent Haar distributed random matrices U, V,

$$\left\|\frac{n}{m}U'V'^T\right\|_{\ell_\infty^n \otimes_\pi \ell_\infty^n} \geq \frac{2-\epsilon}{1.6652} + o(1). \tag{5.3}$$

Now we consider two independent random gaussian matrices X, Y. Their Gram-Schmidt orthonormalizations U, V are Haar distributed and, therefore, they verify Equation (5.3) with probability $1 - o(1)$.

Let us consider now the $n \times m$ submatrices X', Y' corresponding to X, Y. To finish the proof we use our main result, Theorem 1.1, to lower bound $\left\| \frac{1}{m} X' Y'^T \right\|_{\ell_\infty^n \otimes_\pi \ell_\infty^n}$. Let us be more precise. We can write

$$\frac{x_{i,j}}{\sqrt{m}} = \frac{\sqrt{n} u_{i,j}}{\sqrt{m}} + \frac{1}{\sqrt{m}}(x_{i,j} - \sqrt{n} u_{i,j}) \quad \text{and} \quad \frac{y_{i,j}}{\sqrt{m}} = \frac{\sqrt{n} v_{i,j}}{\sqrt{m}} + \frac{1}{\sqrt{m}}(y_{i,j} - \sqrt{n} v_{i,j}).$$

Then, we can use Theorem 1.1 to bound the euclidean norm of the vectors $\left(\frac{1}{\sqrt{m}}(x_{i,j} - \sqrt{n} u_{i,j}) \right)_{j=1}^m$ and $\left(\frac{1}{\sqrt{m}}(y_{i,j} - \sqrt{n} v_{i,j}) \right)_{j=1}^m$. Next, Grothendieck's inequality (see for instance [4, Page 172]) allows us to conclude that, with probability $1 - o(1)$, we have

$$\left\| \frac{1}{m} X' Y'^T \right\|_{\ell_\infty^n \otimes_\pi \ell_\infty^n} \geq \frac{2 - \epsilon - o(1)}{1.6652} - \Big(2\big(\varphi(\alpha)\big) + \big(\varphi(\alpha)\big)^2 \Big) K_G,$$

where $\varphi(\alpha) = \sqrt{2 - \frac{4}{3} \frac{(1-(1-\alpha)^{3/2})}{\alpha}}$ and K_G is Grothendieck's constant. Our result follows now easily.

Acknowledgments

We would like to thank Carlos H. Jiménez for many helpful discussions on previous versions.

Author's research was supported by Spanish research projects MTM2011-26912 and "Ramón y Cajal" program.

References

[1] A. Ambainis, A. Backurs, K. Balodis, D. Kravcenko, R. Ozols, J. Smotrovs, M. Virza, *Quantum strategies are better than classical in almost any XOR game*, Automata, Languages, and Programming Lecture Notes in Computer Science Volume 7391, 2012, 25-37. arXiv:1112.3330.

[2] E. Borel, *Introduction géometrique à quelques théories physiques*, Gauthier-Villars, Paris, 1906.

[3] B. Collins, *Intégrales matricielles et probabilités non-commutatives*, Thése de doctorat, Université Pierre et Marie Curie-Paris VI (2003).

[4] A. Defant and K. Floret, *Tensor Norms and Operator Ideals*, North-Holland, (1993).

[5] P. W. Diaconis, M. L. Eaton, S. L. Lauritzen, *Finite de Finetti theorems in linear models and multivariate analysis*, Scand. J. Statist. 19 289-315 (1992). MR1211786

[6] P. W. Diaconis, D. Freedman, *A dozen de Finetti-style results in search of a theory*, Ann. Inst. H. Poincaré Probab. Statist. 23 397-423 (1987). MR0898502

[7] L. Gallardo, *Au sujet du contenu probabiliste d'un lemma d'Henri Poincar é*, Ann. Univ. Clemont 69 192-197. 1983

[8] C. E. González-Guillén, C. H. Jiménez, C. Palazuelos, I. Villanueva, *Sampling quantum nonlocal correlations with high probability*. Preprint.
[9] S. Dasgupta, A. Gupta, *An elementary proof of a theorem of Johnson and Lindenstrauss*, Random Struct. Alg. 22(1), 60-65 (2003).
[10] W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, Journal of the American Statistical Association 58 (301), 13-30 (1963).
[11] T. Jiang, *How Many Entries of A Typical Orthogonal Matrix Can Be Approximated By Independent Normals?* Ann. Probab. 34(4), 1497-1529 (2006).
[12] T. Jiang, *Maxima of entries of Haar distributed matrices*, Probability Theory and Related Fields, 131, 121-144 (2005).
[13] B. Laurent, P. Massart, *Adaptive estimation of a quadratic functional by model selection*, Ann. Stat., 28(5), 1302-1338 (2000).
[14] V. D. Milman, G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed spaces*, Number 1200 in Lectures Notes in Mathematics, Springer Verlag, New York, (1986).
[15] A. J. Stam, *Limit theorems for uniform distributions on spheres in high- dimensional Euclidean spaces* J. Appl. Probab. 19 221-228 (1982) MR0644435
[16] T. Tao, V. Vu *Random matrices: Universal properties of eigenvectors*, Random Matrices: Theory Appl. 01, 1150001 (2012) and arXiv:1103.2801.
[17] M. Yor, (1985). *Inégalités de martingales continus arretes a un temps quelconques I.* Lecture Notes in Math. 1118. Springer, Berlin.

Carlos E. González-Guillén
Instituto de Matemática Interdisciplinar, IMI
Departamento de Matemáticas del
Área Industrial, E.T.S.I. Industriales, UPM
28006 Madrid, Spain
carlos.gguillen@upm.es

Carlos Palazuelos
Instituto de Ciencias Matemáticas, ICMAT
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
Plaza de Ciencias s/n. 28040, Madrid. Spain
carlospalazuelos@ucm.es

Ignacio Villanueva
Instituto de Matemática Interdisciplinar, IMI
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
Plaza de Ciencias s/n. 28040, Madrid. Spain
ignaciov@ucm.es