Maxima of Entries of Haar Distributed Matrices

TIEFENG JIANG
Revised March 17, 2003

Abstract Let $\mathbf{\Gamma}_n = (\gamma_{ij})$ be an $n \times n$ random matrix such that its distribution is the normalized Haar measure on the orthogonal group $O(n)$. Let also $W_n := \max_{1 \leq i,j \leq n} |\gamma_{ij}|$. We obtain the limiting distribution and a strong limit theorem on $W_n$. A tool has been developed to prove these results. It says that up to $n/(\log n)^2$ columns of $\mathbf{\Gamma}_n$ can be approximated simultaneously by those of some $\mathbf{Y}_n = (y_{ij})$ in which $y_{ij}$ are independent standard normals. Similar results are derived also for the unitary group $U(n)$, the special orthogonal group $SO(n)$, and the special unitary group $SU(n)$.

1 Introduction

To study the generality of mutual incoherence of two orthogonal bases, Donoho and Huo [14] studied a behavior of the largest entry in a random orthogonal matrix. Their result is stated in italics as follows:

Let $\mathbf{\Gamma} = (\gamma_{ij})$ denote a real $n \times n$ orthogonal matrix, uniformly distributed on the orthogonal group $O(n)$. Let $W_n = \max_{1 \leq i,j \leq n} |\gamma_{ij}|$. Then,

$$P \left\{ W_n > 2\sqrt{\log(n)/n(1+\epsilon)} \right\} \to 0$$

as $n \to \infty$ for any $\epsilon > 0$.

This result says roughly that the order of $W_n$, measured in probability, is at most $2\sqrt{\log(n)/n}$. Their simulations also suggest that a normalized $W_n$ converges to some probability distribution.

In this paper, for a sequence of such $W_n$'s, we find out its almost sure behavior. Moreover, we prove that the distribution of $W_n$ converges weakly to an extreme distribution. Further, we show that the similar results also hold for unitary groups $U(n)$, special orthogonal groups $SO(n)$, and the special unitary group $SU(n)$.

1School of Statistics, University of Minnesota.

Key Words: Haar measure, maxima, Gram-Schmidt procedure, large deviations.

AMS (1991) subject classifications: 15A52, 60B10, 60B15, 60F10
First, let us review the definitions of the groups mentioned above. The orthogonal group $O(n)$ and the unitary group $U(n)$ are the sets of all $n \times n$ real orthogonal matrices and complex unitary matrices, respectively. The special orthogonal group $SO(n)$ and the special unitary group $SU(n)$ are the subgroups of $O(n)$ and $U(n)$ such that every matrix in these subgroups has determinant equal to 1. All of the above groups are equipped with the natural matrix product. For details, see [19] and [35].

For any compact group $G$ with multiplication "$\cdot$", for example, $O(n)$, $U(n)$, $SO(n)$ and $SU(n)$, there exists an unique Haar-invariant probability measure $\mu$, that is, $\mu(g_1 \cdot C \cdot g_2) = \mu(C)$ for any measurable subset $C \subset G$, $g_1 \in G$ and $g_2 \in G$. We call such $\mu$ a normalized Haar measure, or normalized Haar distribution, when $G$ is equal to one of the above four groups. To check these details readers are referred to [16], [17] or [30].

Now we state our results. For any $n \times n$ matrix $\Gamma_n = (\gamma_{ij})$, we define

$$W_n = \max_{1 \leq i, j \leq n} |\gamma_{ij}|.$$  \hspace{1cm} (1.2)

As usual, $\log x$ is the natural logarithm of a positive number $x$.

The result in (1.1) says roughly that the magnitude of $W_n$ is at most $2\sqrt{(\log n)/n}$. The following gives the property of $W_n$ in terms of convergence in probability.

**Proposition 1** (i) Suppose $\Gamma_n$ follows the normalized Haar measure on $O(n)$ or $SO(n)$. Then $\sqrt{n/\log n}W_n$ converges to 2 in probability as $n \to \infty$;

(ii) if $\Gamma_n$ follows the normalized Haar measure on $U(n)$ or $SU(n)$, then $\sqrt{n/\log n}W_n$ converges to $\sqrt{2}$ in probability as $n \to \infty$.

This result is stronger than (1.1) by the definition of convergence in probability.

We next look at the almost sure behavior for a sequence of such $W_n$'s. To obtain an almost behavior, a structure of the sequence $\{W_n; n \geq 1\}$ has to be assumed. Inspired by a common procedure for simulating a sequence of Haar distributed matrices in statistical programs, we assume that $\{W_n; n \geq 1\}$ is an independent sequence. A result is obtained as follows:

**Theorem 1** Let $\{\Gamma_n; n \geq 1\}$ be a sequence of independent random matrices. Let also $W_n$ be as in (1.2). If, for each $n \geq 1$, $\Gamma_n$ follows the normalized Haar distribution on the orthogonal group $O(n)$ or the special orthogonal group $SO(n)$, then

(i) $\liminf_{n \to \infty} \sqrt{n/\log n}W_n = 2$ a.s. and $\limsup_{n \to \infty} \sqrt{n/\log n}W_n = \sqrt{6}$ a.s.;

(ii) the sequence $\{\sqrt{n/\log n}W_n, n \geq 2\}$ is dense in $[2, \sqrt{6}]$ a.s.
In contrast to Proposition 1 the quantity $\frac{\sqrt{n}}{\log n} W_n$, under the independence assumption, does not concentrate on any particular value in the long run. Instead, the sequence $\{\sqrt{n}/\log n W_n, n \geq 2\}$ visits every neighborhood whose center is in $[2, \sqrt{6}]$ almost surely. This has an analogy with the classical Hartman-Wintner-Strassen's Law of Iterated Logarithm: let $\{\xi_i; i \geq 1\}$ be a sequence of i.i.d. random variables with mean zero, variance one and partial sums $S_n = \sum_{i=1}^{n} \xi_i$. Let also $d_n = \sqrt{2n \log(\log n)}$. Then $S_n/d_n$ converges to zero in probability as $n$ goes to infinity. However, $\limsup_{n} S_n/d_n = 1$ a.s., and $\liminf_{n} S_n/d_n = -1$ a.s. and $\{S_n/d_n; n \geq 3\}$ is dense in $[-1,1]$ almost surely. See, e.g., section 7.9 from [15] for the real case and section 8.2 from [28] for extensions to random variables taking values in Banach spaces.

The heuristic for deriving Theorem 1 comes from the maximum of independent standard normals. One can check easily that the above result also holds if $\sqrt{n} W_n$ is replaced by $W'_n = \max_{1 \leq i \leq n^2} |\xi_{n,i}|$, where $\{\xi_{n,i}; 1 \leq i \leq n^2, n \geq 1\}$ is a triangular array of i.i.d. standard normals.

The next result is about the unitary groups.

**THEOREM 2** Suppose that $\{\Gamma_n; n \geq 1\}$ is a sequence of independent random matrices. Let $W_n$ be as in (1.2). If, for each $n \geq 1$, $\Gamma_n$ follows the normalized Haar distribution on the unitary group $U(n)$ or the special unitary group $SU(n)$, then

(i) $\liminf_{n \to \infty} \sqrt{\frac{n}{\log n}} W_n = \sqrt{2}$ a.s. and $\limsup_{n \to \infty} \sqrt{\frac{n}{\log n}} W_n = \sqrt{3}$ a.s.;

(ii) the sequence $\{\sqrt{n}/\log n W_n, n \geq 2\}$ is dense in $[\sqrt{2}, \sqrt{3}]$ a.s.

The key difference of proving Theorems 1 and 2 is that normalized entries of $\Gamma_n$ in Theorem 2 asymptotically follow the exponential distribution with parameter one. But the counterparts in Theorem 1 follow asymptotically the standard normal distribution. This distinction is also reflected in the following two results on limiting distributions. We first consider the case that $\Gamma_n$ is an Haar orthogonal invariant matrix.

**THEOREM 3** Suppose $\Gamma_n$ has the normalized Haar distribution on the orthogonal group $O(n)$ or the special orthogonal group $SO(n)$. Then

$$\lim_{n \to \infty} P(n W_n^2 - 4 \log n + \log(\log n) \leq x) = \exp(-K e^{-x/2})$$

for any $x \in \mathbb{R}$, where $K = \sqrt{1/2\pi}$.

Again, the heuristic of figuring out the above result is thinking of $\sqrt{n} W_n$ as the maximum of the absolute values of $n^2$ i.i.d. standard normals. Then the above conclusion is drawn
quickly. Actually, there are similar results for i.i.d. normals in the literature. For instance, on p. 377 from [8] the following result is stated: Let \( W'_n = \max_{1 \leq i \leq n} \xi_i \), where \( \{\xi_i; 1 \leq i \leq n\} \) are i.i.d. standard normals. Then

\[
P(W'_n \leq (2 \log n - \log(\log n) - \log(4\pi) + 2x)^{1/2}) \to e^{-e^{-x}}
\]  

(1.3)
as \( n \to \infty \) for any \( x \in \mathbb{R} \). This can also be seen in Theorem 1.5.3 on p.14 from [27]. Our Theorem 3 can be rewritten in the following form:

\[
P(\sqrt{n} W_n \leq (4 \log n - \log(\log n) - \log(2\pi) + 2x)^{1/2}) \to e^{-e^{-x}}, \quad x \in \mathbb{R},
\]  

(1.4)
as \( n \to \infty \). One can see clearly that \( W''_n \) and \( \sqrt{n} W_n \) share the same scale and the same limiting distribution. The only difference is the normalized constant: \( -\log(8\pi) \) corresponding to \( W''_n \) in (1.3) and \( -\log(2\pi) \) in (1.4). This is because there is no absolute value sign in the definition of \( W''_n \). But there is such sign in that of \( W_n \).

For the unitary group and its subgroup \( SU(n) \), we have the following conclusion:

**Theorem 4** Suppose \( \Gamma_n \) follows the normalized Haar distribution on the unitary group \( U(n) \) or the special unitary group \( SU(n) \). Then,

\[
\lim_{n \to \infty} P(n W_n^2 - 2 \log n \leq x) = \exp(-e^{-x})
\]

for any \( x \in \mathbb{R} \).

Historically, Genedenko [18] studied the limiting behavior of \( U_n := \max_{1 \leq i \leq n} \xi_i \), where \( \{\xi_i; i = 1, 2, \ldots\} \) is a sequence of i.i.d. random variables. He actually obtained the sufficient and necessary conditions for different limiting distributions of \( U_n \). A recent treatment in this direction can be found in [33]. When the \( \xi_i \)'s are weakly dependent, a good way to study \( U_n \) is the Chen-Stein Poisson approximation method. See, e.g., Arratia, Goldstein, and Gordon [3] and Jiang [21] for details and applications in biology in Jiang [22].

The primary concern of Random Matrix Theory is the eigenvalues of different random matrices; see [29] for a book-length treatment. However, Diaconis, Eaton and Lauritzen [11], and D’Aristotile, Diaconis and Newman [7] studied the entries of random orthogonal matrices based on statistical problems. As mentioned earlier, this paper investigates the entries of Haar invariant matrices based on an image analysis problem initially studied by Donoho and Huo [14]. On the other hand, the maxima of the entries of sample correlation matrices is treated by Jiang [23] due to a statistical testing problem. It seems that the study of entries of random matrices is also interesting.
The proofs of the above theorems rely on the following approximation theorems. The first one is about Haar measures on the orthogonal groups. It describes how an Haar invariant orthogonal matrix is similar to a matrix with i.i.d. standard normals as entries. Such a relationship is characterized by measuring their component-wise differences. It is also the rigorous mathematical implementation of our heuristics of deriving Theorems 1 and 3 as mentioned earlier.

**Theorem 5** For each \( n \geq 2 \), there exists matrices \( \Gamma_n = (\gamma_{ij})_{1 \leq i,j \leq n} \) and \( \mathbf{Y}_n = (y_{ij})_{1 \leq i,j \leq n} \) whose \( 2n^2 \) elements are random variables defined on the same probability space such that

(i) the law of \( \Gamma_n \) is the normalized Haar measure on the orthogonal group \( O_n \);  
(ii) \( \{y_{ij}; 1 \leq i,j \leq n\} \) are i.i.d. random variables with the standard normal distribution;  
(iii) set \( \epsilon_n(m) = \max_{1 \leq i \leq n, 1 \leq j \leq m} |\sqrt{m} \gamma_{ij} - y_{ij}| \) for \( m = 1, 2, \ldots, n \). Then

\[
P(\epsilon_n(m) \geq rs + 2t) \leq 4me^{-mr^2/16} + 3mn \left( \frac{1}{s}e^{-s^2/2} + \frac{1}{t} \left( 1 + \frac{t^2}{3(m + \sqrt{m})} \right)^{-n/2} \right)
\]

for any \( r \in (0, 1/4), s > 0, \ t > 0, \) and \( m \leq (r/2)n \).

The idea behind the proof of the above theorem is as follows: Let \( \mathbf{Y}_n = (y_{ij}) \) be an \( n \times n \) matrix where the \( y_{ij} \)'s are independent standard normals. Then \( \mathbf{Y}_n/\sqrt{n} \) has roughly the same law as that of \( \Gamma_n \) as in Theorem 5. Why? First, it is orthogonal invariant. Second, it is almost orthogonal: the length of the first column of \( \mathbf{Y}_n/\sqrt{n} \) is \( (\sum_{i=1}^{n} y_{i1}^2/n)^{1/2} \) which goes to one rapidly (the convergence rate is governed by large deviations); the inner product of the first and second columns of \( \mathbf{Y}_n/\sqrt{n} \) is equal to \( (1/\sqrt{m}) \cdot (\sum_{i=1}^{n} y_{i1}y_{i2}/\sqrt{n}) \), which goes to zero in the order of \( o(1/\sqrt{n}) \) by the classical central limit theorem. Such heuristic is rigorously executed by using the Gram-Schmidt algorithm on \( \mathbf{Y}_n \), which generates an Haar invariant orthogonal matrix.

Recall \( i = \sqrt{-1} \). The next approximation theorem is about the unitary group.

**Theorem 6** For each \( n \geq 2 \), there exists two \( n \times n \) matrices \( \Gamma_n = (\gamma_{pq}) \) and \( \mathbf{Y}_n = ((x_{pq} + iy_{pq})/\sqrt{2}) \) such that \( \gamma_{pq} \)'s, \( x_{pq} \)'s and \( y_{pq} \)'s are random variables defined on the same probability space, and

(i) the law of \( \Gamma_n \) is the normalized Haar measure on the unitary group \( U(n) \);  
(ii) the \( 2n^2 \) random variables \( \{x_{pq}, y_{pq}; 1 \leq p, q \leq n\} \) are independent standard normals;  
(iii) set \( \epsilon_n(m) = \max_{1 \leq p \leq n, 1 \leq q \leq m} |\sqrt{n} \gamma_{pq} - (x_{pq} + iy_{pq})/\sqrt{2}| \) for \( m = 1, 2, \ldots, n \). Then

\[
P(\epsilon_n(m) \geq rs + 2t) \leq 4me^{-mr^2/8} + mne^{-s^2} \frac{6mn}{t} \left( 1 + \frac{t^2}{12(m + t\sqrt{n})} \right)^{-n}
\]

for any \( r \in (0, 1/4), s > 0, \ t > 0, \) and \( m \leq (r/2)n \).
One curious question is: what is the largest order of $m_n$ such that $\epsilon_n(m_n)$ goes to zero in probability? By choosing special values of $r, s, t$ and $m_n$, we have

**Corollary 1** Let $m_n = [n/(\log n)^2]$. Let also $\epsilon_n(m_n)$ be as in Theorem 5 or Theorem 6. Then $\epsilon_n(m_n) \to 0$ in probability as $n \to \infty$.

Recently, Jiang [25] proved that the maximum order of $m_n$ is $o(n/\log n)$ when $\Gamma_n$ is an orthogonal matrix generated by performing the Gram-Schmidt algorithm on the columns of $\mathbf{Y}_n$, where the entries of $\mathbf{Y}_n$ are independent standard normals.

We next make some remarks about Theorems 5 and 6.

Let $\Gamma_n = (\gamma_{ij})$ be a random orthogonal matrix which is uniformly distributed on $O(n)$. Borel [4] showed that

$$P(\sqrt{n} \gamma_{11} \leq x) \to \frac{1}{2\pi} \int_{-\infty}^{x} e^{-t^2/2} dt$$

as $n \to \infty$. He obtained this result in studying “Equivalence of Ensembles” in statistical mechanics. Later, D’Aristotle, Diaconis, Eaton, Freedman, Lauritzen and Newman have extended this result and applied it to some statistical problems; see [7], [11] and [13]. In particular, Diaconis, Lauritzen and Eaton [11] showed that the variation distance between the joint distribution of the entries of the upper-left $k_n \times k_n$ block of $\sqrt{n} \Gamma_n$ and that of $k_n^2$ independent standard normals converges to zero provided $k_n = o(n^{1/3})$ (the largest order of $k_n$ such that the variation distance goes to zero is an open problem, see section 6.3 from [10]; it has been solved recently by Jiang [24]: the largest order is $o(n^{1/2})$).

Our Theorems 5 and 6 study the relationship between the above $\Gamma_n$ and $\mathbf{Y}_n$, where $\mathbf{Y}_n$ is a matrix with independent standard normals as entries. Our results show that the largest difference between entries of the first $m$ columns of $\Gamma_n$ and the corresponding entries of $\mathbf{Y}_n$ converges to zero in probability when $m = m_n = O(n/(\log n)^2)$. This provides another way to characterize the relationship between $\Gamma_n$ and $\mathbf{Y}_n$.

There are some other studies on the entries of Haar invariant matrices. Pickrell [32], Olshansky and Vershik [31], and Borodin and Olshansky [5] have studied entries of matrices in terms of conjugation by random unitary matrices.

We now list some other recent results about Haar measures on some classical groups. Diaconis and Evans [12] proved a functional central limit theorem of eigenvalues of Haar distributed random matrices. Also, Johansson [26] obtained a result on the speed that the traces of Haar distributed random matrices converge in distribution to the standard normal distribution.
Finally, we give the outline of this paper. The proofs of Proposition 1, Theorems 1, 2, 3 and 4, and Corollary 1 are given in Section 2. Theorems 5 and 6 are given in Section 3. In Section 4, some known results are listed for proofs in previous sections.

2 Proofs of Theorems on Maxima of Entries

Let $\mathbb{C}$ be the set of all complex numbers. For $z = x + iy \in \mathbb{C}$, as usual, $|z| = \sqrt{x^2 + y^2}$. The notation $\|v\|$ is the Euclidian norm for a vector $v \in \mathbb{C}^n$. For a $p \times q$ matrix $M = (m_{ij})$, we use the following notation: $\|M\| := \max\{|m_{ij}|, 1 \leq i \leq p, 1 \leq j \leq q\}$. For a random vector $X$, its probability distribution is denoted by $\mathcal{L}(X)$. The standard normal distribution is denoted by $\mathcal{N}(0,1)$.

To prove the Theorems on maxima of entries, we accept, for now, Theorems 5 and 6. They will be proved later in Section 3. Theorems 1, 2, 3 and 4 are proved first. Then we prove Corollary 1 and Proposition 1. There is no circular reasoning in this process.

The following lemma tells us that we only need to work on the Haar measure on $O(n)$ or $U(n)$ in order to obtain conclusions for $SO(n)$ or $SU(n)$.

**Lemma 2.1** Let $\mu_1$, $\mu_2$, $\nu_1$ and $\nu_2$ be the normalized Haar measures on $O(n)$, $SO(n)$, $U(n)$ and $SU(n)$, respectively. We have

$$
(i) \quad \mu_1(\Gamma \in O(n); \|\Gamma\| \leq t) = \mu_2(\Gamma \in SO(n); \|\Gamma\| \leq t);
(ii) \quad \nu_1(\Gamma \in U(n); \|\Gamma\| \leq t) = \nu_2(\Gamma \in SU(n); \|\Gamma\| \leq t)
$$

for any $t > 0$.

**Proof.** (i) Let $K = O(n)$, $G = \{e, e'\}$ and $H = SO(n)$, where $e = \text{diag}(1,1,\cdots,1)$ and $e' = \text{diag}(-1,1,1,\cdots,1)$. It is easy to check that both $G$ and $H$ are closed subgroups of $K$, and $H$ is a normal subgroup of $K$. Further, $K = GH := \{gh; g \in G$ and $h \in H\}$, $G \cap H = \{e\}$ and $K \subset \mathbb{R}^{n^2}$. Let $\mu_3$ be the normalized Haar measure on $G$, that is, $\mu_3(e) = \mu_3(e') = 1/2$. Then by Corollary 7.6.2 on p. 144 from [36],

$$
\int_K f(k)\mu_1(\text{d}k) = \int_H \int_G f(gh)\mu_3(\text{d}g)\mu_2(\text{d}h)
$$

for any $\mu_K$-integrable real function $f(x)$ defined on $K$. Choose $f(x) = I\{\|x\| \leq t\}$, $x \in K$ and $t > 0$. Note that $\|gh\| = \|h\|$ for any $g \in G$ and $h \in H$. Then (i) follows.

(ii) Similarly, let $K = U(n)$, $G = \{\text{diag}(e^{i\theta},1,1,\cdots,1); \theta \in [0,2\pi]\}$ and $H = SU(n)$. Then all the remaining arguments in (i) are valid here. So (ii) is proved.  \[\square\]
Proof of Theorem 1. By Lemma 2.1 and the independence assumption about the \( W_n \)'s, we only need to prove the result about \( O(n) \).

We claim that

\[
\limsup_{n \to \infty} \sqrt{\frac{n}{\log n}} W_n \leq \sqrt{6} \quad \text{a.s. and} \quad \liminf_{n \to \infty} \sqrt{\frac{n}{\log n}} W_n \geq 2 \quad \text{a.s.} \tag{2.6}
\]

and

\[
P(\sqrt{n/\log n} W_n \in (a, b) \ i.o.) = 1 \tag{2.7}
\]

for any \((a, b) \subset (2, \sqrt{6})\).

Suppose (2.6) and (2.7) are true. Then (2.7) implies that \( \{\sqrt{n/\log n} W_n; \ n \geq 2\} \) is dense in \([2, \sqrt{6}]\) a.s. So (ii) is valid. It follows from (ii) that (2.6) still holds if the two inequality signs are reversed respectively. Then (i) follows. Now we prove the two claims.

The proof of the lower bound in (2.6). For any \( \alpha \in (0, 1) \), set \( b_n = 2(1 - \alpha) \sqrt{\log n} \). Then, by Theorem 5, there exist \( \Gamma_n = (\gamma_1, \gamma_2, \cdots, \gamma_n) \) and \( Y = (y_1, y_2, \cdots, y_n) = (y_{ij}) \) for which \( \{y_{ij}; 1 \leq i, j \leq n\} \) are i.i.d. random variables with the standard normal distribution and Theorem 5 holds. Recall \( W_n = \max_{1 \leq j \leq n} \|\gamma_j\| \) and \( \|\| \cdot \| \) is a norm. By the definition of \( \epsilon_n(m) \)

\[
\max_{1 \leq j \leq m} \left| \sqrt{\frac{n}{\log n}} \gamma_j - \max_{1 \leq j \leq m} \|y_j\| \right| \leq \max_{1 \leq j \leq m} \left| \sqrt{\frac{n}{\log n}} \gamma_j - \|y_j\| \right| \leq \epsilon_n(m). \tag{2.8}
\]

It follows that

\[
P(W_n \leq b_n/\sqrt{n}) \leq P\left( \max_{1 \leq j \leq m} \|\sqrt{\frac{n}{\log n}} \gamma_j\| \leq b_n \right) \leq P(\|Y_{nm}\| \leq b_n + \epsilon_n(m)) \tag{2.9}
\]

for any \( 1 \leq m \leq n \), where \( Y_{nm} := (y_1, y_2, \cdots, y_m) \). Now, in Theorem 5, choose \( m = m_n = \lfloor n/\log n \rfloor^2 \), \( r = (\log n)^{-1} \), \( s = (\log n)(\log_2 n)^{-1/2} \) and \( t = \sqrt{(\log n)/\log_2 n} \), where \( \log_2 n := \log(\log n) \). We have that

\[
P(\epsilon_n(m) > 3\sqrt{(\log n)/\log_2 n})
\]

\[
\leq 4n \cdot \exp\left(-\frac{n}{16(\log n)^2}\right) + 3n^2 \cdot \exp\left(-\frac{(\log n)^2}{2\log_2 n}\right) + 3n^2 \left(1 + \frac{(\log n)^3}{6n\log_2 n}\right)^{-n/2}
\]

for sufficiently large \( n \), where we also used the facts that \( 1/s \leq 1, 1/t \leq 1 \) and \( 3(m + \sqrt{n}) \leq 6n/(\log n)^2 \) for \( n \) large enough. The last term above is \( O(\exp(- (\log n)^2)) \). Therefore we have

\[
P(\epsilon_n(m) > 3\sqrt{(\log n)/\log_2 n}) = O(e^{-n(\log n)^{3/2}}) \tag{2.10}
\]
as \( n \to \infty \). Recall \( b_n = (2 - 2\alpha)\sqrt{\log n} \). Then by (2.9),

\[
P(W_n \leq b_n / \sqrt{n}) \\
\leq P(\|Y_{nm}\| \leq b_n + 3\sqrt{(\log n) / (\log_2 n)}) + P(\epsilon_n(m) > 3\sqrt{(\log n) / \log_2 n}) \\
\leq P(\|Y_{nm}\| \leq b'_n) + O(e^{-\log n^{3/2}}),
\]
as \( n \to \infty \), where \( b'_n = (2 - \alpha)\sqrt{\log n} \). By the second inequality of Lemma 4.1, we have that \( P(y_{1,1} \geq b'_n) \leq \exp(-(2 - \alpha)^2(\log n)/2) \) as \( n \) is sufficiently large. It follows from independence that

\[
P(\|Y_{nm}\| \leq b'_n) = (1 - 2P(y_{1,1} \geq b'_n))^{mn} \leq \exp(-2mnP(y_{1,1} \geq b'_n)) \leq \exp(-nC)
\]
for sufficiently large \( n \), where \( C \) is a positive constant depending on \( \alpha \) only. Also, the fact that \( 1 + x \leq e^x \) for any \( x \in \mathbb{R} \) is used in the first inequality. In conclusion,

\[
P(W_n \leq b_n / \sqrt{n}) = O(e^{-\log n^{3/2}})
\]
as \( n \to \infty \), This implies \( \sum_{n \geq 1} P(W_n \leq b_n / \sqrt{n}) < \infty \). By the Borel-Cantelli lemma,

\[
\lim_{n \to \infty} \inf \frac{n}{\log n} W_n \geq 2(1 - \alpha) \quad a.s.
\]
The lower bound in (2.6) is proved since \( \alpha \in (0, 1) \) in the above inequality is arbitrary.

The proof of the upper bound in (2.6). For any \( \epsilon \in (0, 1) \), set \( h_n = (\sqrt{6} + \epsilon)\sqrt{(\log n) / n} \). Recall \( \Gamma_n = (\gamma_{ij}) \) follows the normalized Haar distribution on \( O(n) \). We know that \( \mathcal{L}(\gamma_{ij}) = \mathcal{L}(\xi_1(\sum_{i=1}^n \xi_i^2)^{-1/2}) \) for any \( 1 \leq i, j \leq n \), where \( \xi_i, 1 \leq i \leq n \) are independent standard normal normals. Then

\[
P(W_n \geq h_n) \leq n^2P(\xi_1 \geq h_n(\sum_{i=1}^n \xi_i^2)^{1/2}) \\
\leq n^2P(\xi_1 \geq h_n\sqrt{n - n^{2/3}}) + n^2P(\sum_{i=1}^n \xi_i^2 \leq n - n^{2/3}). \quad (2.11)
\]

Take \( X_i = \xi_i^2 - 1, a_n = n^{-1/3}, A = (-1, 1)^c \) and \( t_0 = 1/4 \) in (ii) of Lemma 4.2. Then \( E \exp(t_0X_1) < \infty \) and

\[
P(\sum_{i=1}^n \xi_i^2 \leq n - n^{2/3}) \leq P\left(\left|\sum_{i=1}^n (\xi_i^2 - 1)\right| \geq n^{2/3}\right) \leq e^{-n^{1/4}} \quad (2.12)
\]
for \( n \) large enough. By the second inequality of Lemma 4.1, since \( h_n\sqrt{n - n^{2/3}} > 1 \) as \( n \) is sufficiently large,

\[
n^2P(\xi_1 \geq h_n\sqrt{n - n^{2/3}}) \leq n^2 \exp\left(-\frac{h_n^2(n - n^{2/3})}{2}\right) = O(n^{-(1+\epsilon)}) \quad (2.13)
\]
as \( n \to \infty \). Therefore, by (2.11), (2.12) and (2.13), \( \sum_{n \geq 1} P(W_n \geq h_n) < \infty \). By the Borel-Cantelli lemma, \( \limsup_{n \to \infty} \sqrt{n/\log n} W_n \leq \sqrt{6} + \epsilon \), a.s. This implies the upper bound in (2.6).

The proof of (2.7). Since the \( W_n \)'s are independent, by the Borel-Cantelli lemma, we only need to show that

\[
\sum_{n \geq 1} P(\sqrt{n/\log n} W_n \in (a, b)) = \infty
\]

for any \((a, b) \subset (2, \sqrt{6})\). Replace \( h_n \) in (2.11) by \( b \sqrt{\log n} / n \). Note that \( h_n (n - n^{2/3})^{1/2} \sim (\sqrt{6} + \epsilon) \sqrt{\log n} \) and \( b \sqrt{\log n} / n \cdot \sqrt{n - n^{2/3}} \sim b \sqrt{\log n} \). By the same argument as in (2.11), we obtain that

\[
P(\sqrt{n/\log n} W_n \geq b) \leq n^{(4-b^2)/2}
\]

as \( n \) is sufficiently large for fixed \( b_1 < b \). By (2.8), we use the fact \( W_n \geq \max_{1 \leq j \leq m} \| \gamma_j \| \) to obtain

\[
P(\sqrt{n/\log n} W_n > a) \geq P(\| Y_{nm} \| - \epsilon(m) > a \sqrt{\log n}) \geq P(\| Y_{nm} \| > f_n) - P(\epsilon(m) \geq 3 \sqrt{(\log n) / \log 2 n}),
\]

where \( f_n = a \sqrt{\log n} + 3 \sqrt{(\log n) / \log 2 n} \) and \( m = m_n \) as in (2.10). Now, by independence and the first inequality of Lemma 4.1,

\[
P(\| Y_{nm} \| \geq f_n) = 1 - (1 - P(\| y_{1,1} \| \geq f_n))^{mn} \geq 1 - \left( 1 - \frac{2f_n}{\sqrt{2\pi(1 + f_n^2)}} e^{-f_n^2/2} \right)^{mn}.
\]

Since \( 1 - x \leq e^{-x} \) for all \( x \in \mathbb{R} \), \( 1 - (1 - x)^n \geq 1 - e^{-nx} \sim nx \) if \( nx \to 0 \). Remember \( a > 2 \).

It is easy to check that

\[
n^{(4-a^2)/2} \leq \frac{2mnf_n}{\sqrt{2\pi(1 + f_n^2)}} e^{-f_n^2/2} \to 0
\]

when \( n \to \infty \) for any \( a_1 > a \). In summary,

\[
P(\| Y_{nm} \| \geq f_n) \geq n^{(4-a^2)/2}
\]

as \( n \) is sufficiently large. Therefore by (2.10), (2.16) and the above inequality

\[
P(\sqrt{n/\log n} W_n > a) \geq n^{(4-a^2)/2}
\]

as \( n \) is sufficiently large for fixed \( a_1 > a \). Combining this with (2.15), we obtain

\[
P(\sqrt{n/\log n} W_n \in (a, b)) = P(\sqrt{n/\log n} W_n > a) - P(\sqrt{n/\log n} W_n \geq b) \geq n^{(4-a^2)/2} - n^{(4-b^2)/2} \sim n^{(4-a^2)/2}
\]
as \( n \to \infty \) for any interval \((a_1, b_1) \subset (a, b) \subset (2, \sqrt{6})\). Therefore, (2.14) follows. The entire proof is complete. \( \blacksquare \)

The essence of the proof of Theorem 1 is Theorem 5. Theorem 6 bears some analogy with Theorem 5. The following proof of Theorem 2, based on Theorem 6, is similar to that of Theorem 1. The difference is that instead of working on normal random variables we deal with the square of the norm of a random variable with the standard complex normal distribution. Such a norm follows the exponential distribution with parameter one.

**Proof of Theorem 2.** By Lemma 2.1 and the independence assumption about the \( W_n \)'s, we only need to prove the unitary group case.

As in the proof of Theorem 1, we only need to show that
\[
\sqrt{2} \leq \lim \inf_{n \to \infty} \sqrt{\frac{n}{\log n}} W_n \leq \lim \sup_{n \to \infty} \sqrt{\frac{n}{\log n}} W_n \leq \sqrt{3} \quad \text{a.s.} \quad (2.19)
\]
and
\[
P(\sqrt{n/\log n} W_n \in (a, b) \text{ i.o.}) = 1 \quad (2.20)
\]
for any \((a, b) \subset (\sqrt{2}, \sqrt{3})\). We claim that
\[
P(\sqrt{n/\log n} W_n \leq \sqrt{2}(1 - \alpha)) \leq O(e^{-(\log n)^{3/2}}) \quad (2.21)
\]
as \( n \to \infty \) for any \( \alpha \in (0, 1) \), and
\[
n^{2-b_2^2} \leq P(\sqrt{n/\log n} W_n \geq b) \leq n^{2-b_1^2} \quad (2.22)
\]
as \( n \) is sufficiently large for any \( b > \sqrt{2} \) and \( 0 < b_1 < b < b_2 \). If the claims are true, the lower bound in (2.19) follows from (2.21); the upper bound and (2.20) follow from (2.22). Now let's prove the claims.

Set \( b_n = (1 - \alpha)\sqrt{2\log n} \) for \( \alpha \in (0, 1) \). Then, as in (2.9), we obtain that
\[
P(W_n \leq b_n/\sqrt{n}) \leq P(\|Y_{nm}\| \leq b_n + \epsilon_n(m)) \quad (2.23)
\]
for any \( 1 \leq m \leq n \), where \( Y_{nm} \) is a matrix generated by the first \( m \) columns of the matrix \( Y_n = ((x_{pq} + iy_{pq})/\sqrt{2}) \), and the \( 2n^2 \) random variables \( \{x_{pq}, y_{pq}; 1 \leq p, q \leq n\} \) are i.i.d. with the standard normal distribution. Choosing \( m = m_n = \lfloor n/(\log n)^2 \rfloor \), \( r = (\log n)^{-1} \), \( s = (\log n)(\log_2 n)^{-1/2} \) and \( t = \sqrt{(\log n)/\log_2 n} \), we have by the same argument as in (2.10) that
\[
P(\epsilon_n(m) > 3\sqrt{(\log n)/\log_2 n}) = O(e^{-(\log n)^{3/2}})
\]
as $n \to \infty$. It follows from (2.23) that

$$P(W_n \leq b_n/\sqrt{n}) \leq P(\|Y_{nm}\| \leq b'_n) + O(e^{-(\log n)^{3/2}}) \text{ as } n \to \infty,$$  
(2.24)

where $b'_n = (\sqrt{2} - \alpha) \sqrt{\log n}$. Observe that $(x^2_{11} + y^2_{11})/2 \sim \text{Exp}(1)$, the exponential distribution with parameter 1. So $P((x^2_{11} + y^2_{11})/2 \geq t) = e^{-t}$ for $t > 0$. Using independence and the fact that $1 + x \leq e^x$ for all $x \in \mathbb{R}$, we have that

$$P(\|Y_{nm}\| \leq b'_n) = (1 - P((x^2_{11} + y^2_{11})/2 \geq b'_n)^{nm} \leq \exp(-nme^{-b'^2_n}) \leq e^{-n\alpha}$$

for sufficiently large $n$. This together with (2.24) yields (2.21).

Now we prove the second inequality in (2.22). Recall $\Gamma_n$ follows the normalized Haar distribution on the unitary group, then each element of $\Gamma_n$ has the same distribution as $(x_{11} + iy_{11})(\sum_{1 \leq p \leq n}(x^2_{p1} + y^2_{p1}))^{-1/2}$, where the $x_{pq}$'s and the $y_{pq}$'s are as in (2.23). Therefore,

$$P(\sqrt{n/\log n}W_n \geq b) \leq n^2P\left(\frac{x^2_{11} + y^2_{11}}{2} \geq \frac{b^2 \log n}{2n} \sum_{p=1}^{n}(x^2_{p1} + y^2_{p1})\right).$$

Again, $(x^2_{11} + y^2_{11})/2 \sim \text{Exp}(1)$, by using the same argument as in (2.11), we obtain the second inequality in (2.22). The first inequality in (2.22) can be shown by using the same spirit as deriving (2.18) and the fact that $(x^2_{11} + y^2_{11})/2 \sim \text{Exp}(1)$. We omit the details. $\blacksquare$

To prove Theorem 3, we need the following lemma.

**Lemma 2.2** Suppose that a random matrix $\Gamma_n$ follows the normalized Haar distribution on the orthogonal group $O(n)$. Let $\Gamma_n = (\gamma_1, \gamma_2, \cdots, \gamma_n)$. Define $A_j = \{\|\gamma_j\| \geq \sqrt{a_n + x}\}$ for $x > -a_n$ and $j = 1, 2, \cdots, n$, where $a_n = 4 \log n - \log(\log n)$. For any integer $m \geq 1$, we have that

$$\lim_{n \to \infty} n^m P(A_1 \cap A_2 \cap \cdots \cap A_m) = (\sqrt{1/2\pi e^{-x/2}})^m$$

for any $x \in \mathbb{R}$.

**Proof.** First, by Theorem 5, there exists an $n \times n$ random matrix $Y = (y_{ij}) = (y_1, y_2, \cdots, y_n)$ such that $\{y_{ij}; 1 \leq i, j \leq n\}$ are i.i.d. standard Gaussian random variables and Theorem 5 holds. Choose $r = n^{-1/4}$, $s = n^{1/8}$ and $t = n^{-1/8}$ in Theorem 5. Then we have that

$$P(\epsilon_n(m) \geq 3n^{-1/8}) \leq 4ne^{-n^{1/2}/16} + 3n^2e^{-n^{1/4}/2} + 3n^3\left(1 + \frac{n^{-1/4}}{3(m + \sqrt{n})}\right)^{-n/2}$$

12
as \( n \) is sufficiently large, where we use the fact \( s^{-1} \leq 1 \) and \( t^{-1} \leq n \). It is easy to see that the last term above is bounded by \( \exp(-Cn^{1/4}) \) for each \( n \geq 1 \) and some positive constant \( C \) depending on \( m \) only. Therefore
\[
P(\epsilon_n(m) \geq 3n^{-1/8}) = o(e^{-n^{1/8}}) \quad (2.25)
\]
as \( n \to \infty \). Now, set
\[
d_n^+ = \sqrt{a_n + x + \epsilon_n(m)}, \quad d_n^- = \sqrt{a_n + x - \epsilon_n(m)},
\]
\[
B_j^+ = \{\|y_j\| \geq d_n^+\} \quad \text{and} \quad B_j^- = \{\|y_j\| \geq d_n^-\}
\]
for \( j = 1, 2, \cdots, m \). It follows from (2.8) that \( \|y_j\| - \epsilon_n(m) \leq \|y_j\| - \|\hat{y}_j\| \leq \|\hat{y}_j\| + \epsilon_n(m) \) for \( j = 1, 2, \cdots, m \). Thus, \( \{\|y_j\| \geq d_n^+\} \subset A_j \subset \{\|y_j\| \geq d_n^-\}, \ j = 1, 2, \cdots, m \) and
\[
P(B_1^+ \cap B_2^+ \cap \cdots \cap B_m^+) \leq P(A_1 \cap A_2 \cap \cdots \cap A_m)
\]
\[
\leq P(B_1^- \cap B_2^- \cap \cdots \cap B_m^-). \quad (2.26)
\]
We next calculate \( P(B_1^- \cap B_2^- \cap \cdots \cap B_m^-) \). Set \( h_n = \sqrt{a_n + x} - 3n^{-1/8} \). It is easy to see that
\[
B_1^- \cap B_2^- \cap \cdots \cap B_m^-
\]
\[
\subset \left( B_1^- \cap B_2^- \cap \cdots \cap B_m^- \cap \{\epsilon_n(m) < 3n^{-1/8}\} \right) \cup \{\epsilon_n(m) \geq 3n^{-1/8}\}
\]
\[
\subset \{ \min_{1 \leq j \leq m} \|y_j\| \geq h_n \} \cup \{\epsilon_n(m) \geq 3n^{-1/8}\}.
\]
Therefore
\[
P(B_1^- \cap B_2^- \cap \cdots \cap B_m^-) \leq P(\min_{1 \leq j \leq m} \|y_j\| \geq h_n) + P(\epsilon_n(m) \geq 3n^{-1/8})
\]
\[
= P(\max_{1 \leq i \leq n} |\eta_i| \geq h_n)^m + P(\epsilon_n(m) \geq 3n^{-1/8}), \quad (2.27)
\]
where \( \{\eta_i; 1 \leq i \leq n\} \) are i.i.d. random variables with the standard normal distribution.
We claim that
\[
\limsup_{n \to \infty} n^m P(B_1^- \cap B_2^- \cap \cdots \cap B_m^-) \leq ((2\pi)^{-1/2}e^{-x/2})^m. \quad (2.28)
\]
By (2.25) and (2.27), to prove the claim, it suffices to show that
\[
\limsup_{n \to \infty} (n P(\max_{1 \leq i \leq n} |\eta_i| \geq h_n))^m \leq ((2\pi)^{-1/2}e^{-x/2})^m. \quad (2.29)
\]
Indeed, by Lemma 4.1, we have that
\[
n^2 P(|\eta_1| \geq h_n) \sim n^2 \cdot \frac{2}{\sqrt{2\pi h_n}} e^{-h_n^2/2} \sim \frac{1}{\sqrt{2\pi}} e^{-x/2}
\]
as $n \to \infty$. Given $t \in (0, 1)$. By Taylor’s expansion, $(1-t)^n = 1 - nt + (n(n-1)/2)t^2(1-\delta)^{n-2}$ for some $\delta$ such that $0 < \delta < t < 1$. Therefore $|(1-t)^n - 1 + nt| \leq (nt)^2$ for $n \geq 2$. Now choose $t = P(\eta_1 \geq h_n)$. By (2.30)

$$1 - (1 - P(\eta_1 \geq h_n))^n \sim \frac{1}{n} \left(\frac{1}{2\pi}e^{-x/2}\right) + o\left(\frac{1}{n}\right)$$

as $n \to \infty$. So

$$nP(\max_{1 \leq i \leq n} |\eta_i| \geq h_n) = n \left(1 - (1 - P(\eta_1 \geq h_n))^n\right) \sim \frac{1}{2\pi}e^{-x/2}$$

as $n \to \infty$. Therefore (2.29) is validated since $m$ is fixed, and the claim (2.28) then follows. By the same arguments, we obtain

$$\lim \inf_{n \to \infty} \left\{ n^m P(B_1^+ \cap B_2^+ \cap \cdots \cap B_m^+) \right\} \geq (\sqrt{1/2\pi}e^{-x/2})^m.$$

This inequality together with (2.26) and (2.28) yields the desired result. \hfill \blacksquare

Now we are ready to prove Theorem 3.

**Proof of Theorem 3.** As in the proof of Theorem 1, we only need to deal with the orthogonal group case.

Let us continue the notation in Lemma 2.2. Recall (1.2) and $\Gamma = (\gamma_1, \gamma_2, \cdots, \gamma_n)$. The following equality is true:

$$W_n = \max_{1 \leq j \leq n} \|\gamma_j\|.$$ 

So, to prove the theorem, we only need to show that

$$P(\max_{1 \leq j \leq n} \sqrt{n} \|\gamma_j\| \geq \sqrt{a_n + x}) \to 1 - e^{-J}, \quad x \in \mathbb{R},$$

where $J = \sqrt{1/2\pi} \cdot e^{-x/2}$. Recall the definitions of $A_j$’s in Lemma 2.2. Fix integer $m \geq 1$.

For any $n \geq 2m$, by the Bonferroni inequality (see, e.g., p. 22 from [15]), the probability in (2.31) is bounded below and above respectively by

$$\sum_{j=1}^{n} P(\bigcup_{1 \leq i \leq j} A_j) - \sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{2m+1} \sum_{j_1<j_2<\cdots<j_{2m}} P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_{2m}})$$

(2.32)

and

$$\sum_{j=1}^{n} P(\bigcup_{1 \leq i \leq j} A_j) - \sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{2m+2} \sum_{j_1<j_2<\cdots<j_{2m+1}} P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_{2m+1}}).$$

(2.33)
By assumption, \( \mathbf{\Gamma}_n = (\gamma_1, \gamma_2, \cdots, \gamma_n) \) satisfies the normalized Haar distribution. By right multiplying \( \mathbf{\Gamma}_n \) with permutation matrices (see, e.g., p.25 from [20] for the definition), we know that the \( n \) random vectors \( \gamma_1, \gamma_2, \cdots, \gamma_n \) are exchangeable. Thus the term in (2.32) is equal to

\[
nP(A_1) \left[ - \binom{n}{2} P(A_1 \cap A_2) + \cdots + (-1)^{2m+1} \binom{n}{2m} P(A_1 \cap A_2 \cap \cdots \cap A_{2m}) \right].
\]

If \( i \geq 1 \) is fixed, we know that \( \binom{n}{i} / n^i \to 1 / i! \) as \( n \to \infty \). Letting \( n \to \infty \), by Lemma 2.2, we obtain

\[
\liminf_{n \to \infty} P(\max_{1 \leq j \leq n} \|\sqrt{n}\| \gamma_j\| \geq \sqrt{a_n + x}) \geq \frac{2m}{i!} \sum_{i=1}^{2m+1} (-1)^{i+1} \frac{J^i}{i!}.
\]

Applying the same argument to (2.33), we obtain

\[
\limsup_{n \to \infty} P(\max_{1 \leq j \leq n} \|\sqrt{n}\| \gamma_j\| \geq \sqrt{a_n + x}) \leq \frac{2m+1}{i!} \sum_{i=1}^{2m+1} (-1)^{i+1} \frac{J^i}{i!}.
\]

Pass the limit \( m \to +\infty \) for the above two inequalities. Remember that the left hand sides of the above two inequalities are irrelevant to \( m \). Also, \( e^{-J} = \sum_{i=0}^{\infty} (-1)^{i}J^i / i! \). Then (2.31) is concluded.  

Now we prove Theorem 4.

**Proof of Theorem 4.** Again, as in the proof of Theorem 2, it suffices to prove the theorem for the unitary group case.

Let \( a_n = 2 \log n \) and \( A_j = \left\{ \|\sqrt{n}\| \gamma_j \| \geq \sqrt{a_n + x} \right\}, \ j = 1, 2, \cdots, n \). We first claim that

\[
\lim_{n \to \infty} n^m P(A_1 \cap A_2 \cap \cdots \cap A_m) = (e^{-x})^m
\]

for any integer \( m \geq 1 \). If (2.35) is true, by following the proof of Theorem 3 completely, the proof of Theorem 4 is then terminated. Now we prove the claim.

Recall the proof of Lemma 2.2. Take a sequence of i.i.d. random variables \( \{\eta, \eta_1, \cdots, \eta_n\} \) with \( \eta \) following the standard complex normal distribution. The corresponding of (2.25) is still true by Theorem 6. Let \( h_n = \sqrt{a_n + x} - 3n^{-1/8} \). Since \( |\eta|^2 \) follows the exponential distribution with parameter one we have \( n^2 P(|\eta|^2 \geq h_n) = n^2 e^{-h_n^2} \sim e^{-x} \) as \( n \to \infty \). Following the rest arguments in the proof of Lemma 2.2, we obtain (2.35).

Now we prove Corollary 1.

**Proof of Corollary 1.** As usual, for a real number \( x \), the notation \( \lfloor x \rfloor \) stands for the
largest integer less than or equal to \( x \). We next only focus on orthogonal case. The unitary case can be done by the same arguments.

Now, choosing \( r = (\log n)^{-1} \), \( s = (\log n)^{3/4} \), \( t = (\log n)^{-1/4} \) and \( m = m_n = [n/(\log n)^2] \) in Theorem 5. Then by Theorem 5, we have that

\[
P(\epsilon_n(m_n) \geq 3(\log n)^{-1/4}) \\
\leq 4n \cdot \exp \left( -\frac{n}{16(\log n)^2} \right) + 3n^2(\log n)^{-3/4} \cdot \exp \left( -\frac{(\log n)^{3/2}}{2} \right) \\
+ 3n^2(\log n)^{1/4} \left( 1 + \frac{1}{3} \cdot \frac{(\log n)^{-1/2}}{[n(\log n)^{-2}] + \sqrt{n}} \right)^{-n/2}
\]

for sufficiently large \( n \). The first two terms on the right hand side go to zero. The last term is bounded by

\[
n^3 \left( 1 + \frac{(\log n)^{3/2}}{4n} \right)^{-n/2} \leq n^3 \cdot \exp( -(\log n)^{3/2}/9)\]

as \( n \) is sufficiently large. So the third term also goes to zero. It follows that \( \epsilon_n(m_n) \) goes to zero in probability. ■

We prove Proposition 1 to end this section.

**Proof of Proposition 1.** By Lemma 2.1, it suffices to show (i) and (ii) for the orthogonal case and the unitary case, respectively. We only deal with the \( O(n) \) case. The \( U(n) \) case is similar.

Let \( \Upsilon_n \) have the normalized Haar measure on \( O(n) \). Then

\[
P \left( \left| \frac{1}{\sqrt{\log n}} W_n - 2 \right| \geq \epsilon \right) \leq P \left( nW_n^2 \geq (2 + \epsilon)^2 \log n \right) + P \left( nW_n^2 \leq (2 - \epsilon)^2 \log n \right) \quad (2.36)
\]

for any \( \epsilon \in (0, 1) \). By Theorem 3,

\[
\limsup_{n \to \infty} P \left( nW_n^2 \geq (2 + \epsilon)^2 \log n \right) \leq \limsup_{n \to \infty} P(nW_n^2 - 4 \log n + \log(\log n) > x) \\
= 1 - \exp(-Ke^{-x/2})
\]

for any \( x > 0 \). Letting \( x \uparrow +\infty \), we have that the middle probability in (2.36) goes to zero as \( n \to \infty \). By the same argument, the last probability also goes to zero. Therefore, \( \sqrt{n/\log n} W_n \) goes to 2 in probability. ■
3 Proofs of Theorems 5 and 6

There are a lot of methods to generate random matrices with the normalized Haar distribution on the orthogonal and the unitary groups. For example, let \( \mathbf{Y} = (y_{ij}) = (y_1, y_2, \cdots, y_n) \) be an \( n \times n \) random matrix whose \( n^2 \) elements are i.i.d. random variables with the standard normal distribution. Performing the Gram-Schmidt procedure on the columns of \( \mathbf{Y} \), we then obtain an orthogonal random matrix with the normalized Haar distribution on the orthogonal group \( O(n) \); see Proposition 7.2 (take \( p = n \)) on page 234-235 from [16]. Also, the matrix \( \mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1/2} \) follows the normalized Haar distribution on \( O(n) \); see Proposition 7.1 in [17]. For the unitary counterparts of the above two procedures, one only needs to replace \( \mathbf{Y} \) by \( (\mathbf{Y} + i\mathbf{Z})/\sqrt{2} \), where \( \mathbf{Z} \) is an independent copy of \( \mathbf{Y} \). Then change the operation “\( T \)” to “\( * \)”, where \( \mathbf{Y}^* = (\overline{y_{ij}})^T \). In this section, we prove Theorems 5 and 6 via the Gram-Schmidt algorithm. Let us review it first.

For the sequence of \( n \times 1 \) complex vectors \( \{y_1, y_2, \cdots, y_n\} \), define \( \mathbf{w}_1 = y_1 \), and

\[
\mathbf{w}_i = y_i - \sum_{j=1}^{i-1} \frac{y_i^* \mathbf{w}_j}{||\mathbf{w}_j||^2} \mathbf{w}_j, \quad i = 2, 3, \cdots, n, \tag{3.37}
\]

where \( ||\mathbf{w}_j||^2 = \mathbf{w}_j^* \mathbf{w}_j \) \( j = 1, 2, \cdots, n \). Then, \( \{\mathbf{w}_i, 1 \leq i \leq n\} \) are orthogonal, i.e., \( \mathbf{w}_i^* \mathbf{w}_j = 0 \) for any \( 1 \leq i < j \leq n \). Let \( \gamma_i = (1/||\mathbf{w}_i||) \mathbf{w}_i, \quad i = 1, 2, \cdots, n \). We then obtain an unitary matrix \( \mathbf{U}_n = (\gamma_1, \gamma_2, \cdots, \gamma_n) \). So (3.37) can be rewritten as follows:

\[
\mathbf{w}_i = y_i - \sum_{j=1}^{i-1} (y_i^* \gamma_j) \gamma_j, \quad i = 2, 3, \cdots, n, \tag{3.38}
\]

For further reference see e.g. p.15 from [20] and Section A.5 on page 603 from [2].

Define

\[
\Delta_1 = 0, \quad \Delta_i = \sum_{j=1}^{i-1} (y_i^* \gamma_j) \gamma_j \quad \text{and} \quad L_i = \sqrt{\frac{n}{||\mathbf{w}_i||^2}} - 1, \quad i = 1, 2, \cdots, n. \tag{3.39}
\]

We need some preparations to prove Theorems 5 and 6.

**Lemma 3.1** For any \( m \) such that \( 1 \leq m \leq n \), we have that

\[
\varepsilon_n(m) := \max_{1 \leq i \leq m} \|\sqrt{n}\gamma_i - y_i\| \\
\leq \max_{2 \leq i \leq m} \|\Delta_i\| + \left( \max_{1 \leq i \leq m} L_i \right) \left( \max_{1 \leq i \leq m, 1 \leq j \leq n} |y_{ij}| + \max_{2 \leq i \leq m} ||\Delta_i|| \right).
\]

**Proof.** Note that \( \mathbf{w}_i = y_i - \Delta_i \). We have that

\[
\sqrt{\frac{n}{||\mathbf{w}_i||^2}} \mathbf{w}_i - y_i = -\Delta_i + (y_i - \Delta_i)\left( \sqrt{\frac{n}{||\mathbf{w}_i||^2}} - 1 \right).
\]
Then, the desired inequality follows from the triangle inequality of $\| \cdot \|$.

The following properties of $I(x)$ will be used later. It is called a rate function in the theory of large deviations. The proof is standard and is omitted. The reader is referred to p. 35 from [9].

**Lemma 3.2** Let $\xi \sim N(0, 1)$ and $I(x) = \sup_{\theta \in \mathbb{R}} \{ \theta x - \log(E \exp(\theta \xi^2)) \}$ for $x \in \mathbb{R}$. Then
\[(i) \ E \exp(\theta \xi^2) = (1 - 2\theta)^{-1/2} \text{ for } \theta < 1/2;\]
\[(ii) \ I(x) = \begin{cases} (x - 1 - \log x)/2 & \text{if } x > 0; \\ +\infty & \text{otherwise.} \end{cases}\]
\[(iii) \text{ Define } J(x) = I(x)/x \text{ for } x > 0. \text{ Then both } I(x) \text{ and } J(x) \text{ are increasing on } (1, \infty) \text{ and decreasing on } (0, 1).\]

We collect the following elementary facts. The proof is omitted.

**Lemma 3.3** The following holds:
\[(i) \ x - 1 - \log x \geq (x - 1)^2/2 \text{ for } x \in (0, 1];\]
\[(ii) \ 2x - \log(1 + 2x) \geq x^2 \text{ for } x \in (0, 1/4];\]
\[(iii) \ (1 - x)^{-2} \geq 1 + 2x \text{ and } (1 + x)^{-2} \leq 1 - x \text{ for } x \in (0, 1/4].\]

The next lemma gives a probability bound about the tail of a product of a normal and a random variable with $F$-distribution. The proof is based mainly on the property of the standard normal distribution.

**Lemma 3.4** Let $\{\xi, \xi_i, i = 1, 2, \cdots, n\}$ be a sequence of i.i.d. random variables with $\xi \sim N(0, 1)$. Define $\eta^2 = (\sum_{k=1}^m \xi_k^2)/(\sum_{k=1}^n \xi_k^2)$ for some $1 \leq m < n$. Then
\[P(|\xi| \geq t/\eta) \leq \frac{6}{\sqrt{2\pi t}} \left( 1 + \frac{i^2}{3(m + t\sqrt{n})} \right)^{-n/2}\]
for any $t > 0$.

**Proof.** Note that $\xi$ and $\eta$ are independent and $\eta \leq 1$. By the upper bound from Lemma 4.1 we have that
\[P(|\xi| \geq t/\eta) = E\{P(|\xi| \geq t\eta^{-1} | \eta)\} \leq E \left( \frac{2}{\sqrt{2\pi (t\eta^{-1})}} e^{-t^2\eta^{-2}/2} \right) \leq \frac{2}{\sqrt{2\pi t}} E e^{-t^2\eta^{-2}/2}. \quad (3.40)\]
Now, \( \eta^{-2} = 1 + \left( \sum_{k=m+1}^{n} \xi_k^2 \right) \left( \sum_{k=1}^{m} \xi_k^2 \right)^{-1} \), and \( \{\xi_{m+1}, \xi_{m+2}, \ldots, \xi_n\} \) and \( \sum_{k=1}^{m} \xi_k^2 \) are independent. Thus \( E e^{-t^2 \eta^{-2}/2} = e^{-t^2/2} E(M^{n-m}) \), where

\[
M = E \left\{ \exp \left( -\frac{t^2 \xi_n^2}{2 \sum_{k=1}^{m} \xi_k^2} \right) \mid \xi_1, \xi_2, \ldots, \xi_m \right\}.
\]

By (i) of Lemma 3.2, \( E \exp(-\beta \xi_n^2) = (1 + 2\beta)^{-1/2} \) for \( \beta > -1/2 \). Then \( M = (1 + t^2 (\sum_{k=1}^{m} \xi_k^2)^{-1})^{-1/2} \). In summary,

\[
P(|\xi| \geq t/\eta) \leq \frac{2e^{-t^2/2}}{\sqrt{2\pi} t} E \left\{ \left( 1 + \frac{t^2}{\sum_{k=1}^{m} \xi_k^2} \right)^{-\frac{n-m}{2}} \right\}.
\]

By (i) of Lemma 4.2,

\[
P \left( \sum_{k=1}^{m} \xi_k^2 \geq x \right) \leq 2e^{-mI(A)}, x > 0,
\]

where \( I(x) \) is as in (ii) of Lemma 3.2 and \( A = [x/m, \infty) \). By (iii) of Lemma 3.2, \( I(A) = I(x/m) \) if \( x \geq m \). Given \( t > 0 \). Choose \( x_0 = 2m + t\sqrt{8(n-m)} \). By (3.42) and (iii) of Lemma 3.2 on \( J(x) \),

\[
P \left( \sum_{k=1}^{m} \xi_k^2 \geq x_0 \right) \leq 2e^{-mI(x_0/m)} = 2e^{-x_0 J(x_0/m)} \leq 2e^{-x_0 J(2)} \leq 2e^{-x_0/16}
\]
since \( x_0/m > 2 \) and \( J(2) = I(2)/2 = (1 - \log 2)/4 > 1/16 \). Considering \( \sum_{k=1}^{m} \xi_k^2 > x_0 \) or not, we have from above that

\[
E \left\{ \left( 1 + \frac{t^2}{\sum_{k=1}^{m} \xi_k^2} \right)^{-\frac{n-m}{2}} \right\} \leq \left( 1 + \frac{t^2}{x_0} \right)^{-\frac{n-m}{2}} + 2e^{-x_0/16}.
\]

Since \( 1 + x \leq e^x \) for any \( x \in \mathbb{R} \), \( e^{-x_0/16} \leq \left( 1 + (t^2/x_0) \right)^{-x_0/16} \). Also, \( x_0^2/(16t^2) > (n-m)/2 \).

The above says that

\[
E \left\{ \left( 1 + \frac{t^2}{\sum_{k=1}^{m} \xi_k^2} \right)^{-\frac{n-m}{2}} \right\} \leq 3 \left( 1 + \frac{t^2}{x_0} \right)^{-\frac{n-m}{2}} \leq 3e^{t^2/2} \left( 1 + \frac{t^2}{3(m + t\sqrt{n})} \right)^{-n/2},
\]

where we use the facts \( (1 + t^2x_0^{-1})^{m/2} \leq \exp(t^2x_0^{-1}m/2) \leq e^{t^2/2} \) and \( x_0 < 3(m + t\sqrt{n}) \) in the last step. This and (3.41) yields the desired inequality.

The following is a key result in analyzing the tail of \( \epsilon_n(m) \) as stated in Theorem 5. Its proof relies on Lemma 3.4.

**Lemma 3.5** Let \( \{y_1, y_2, \ldots, y_n\} \) be a sequence of i.i.d. \( \mathbb{R}^n \)-valued random vectors with \( y_1 \sim N(0, I_n) \), where \( I_n \) is the \( n \times n \) identity matrix. Let also \( \Delta_i \) be as in (3.39). For any \( t > 0 \) and \( m \) such that \( 1 \leq m < n \), we have that

\[
P(\max_{1 \leq i \leq m} \|\Delta_i\| \geq t) \leq \frac{6mn}{\sqrt{2\pi t}} \left( 1 + \frac{t^2}{3(m + t\sqrt{n})} \right)^{-n/2}.
\]
Proof. Remember $\Delta_1 = 0$. We only need to deal with the case $m \geq 2$. Review $\Delta_i$ as in (3.39). Using $(y_j^T \gamma_j) \gamma_j = (\gamma_j y_j^T) y_i$, we obtain $\Delta_i = (\sum_{j=1}^{i-1} \gamma_j y_j^T) y_i$. Observe that $(\sum_{j=1}^{i-1} \gamma_j y_j^T) y_i = \sum_{j=1}^{i-1} \gamma_j y_j^T$ by orthogonality. Also, $\gamma_j$ is a function of $y_1, y_2, \ldots, y_j$. Hence $\{\gamma_1, \gamma_2, \ldots, \gamma_{i-1}\}$ and $y_i$ are independent. Consequently, conditionally on $y_1, y_2, \ldots, y_{i-1}$,

$$\Delta_i \sim N(0, \Sigma_i), \text{ where } \Sigma_i = \sum_{j=1}^{i-1} \gamma_j \gamma_j^T = (\sum_{j=1}^{i-1} \gamma_j y_j)_{1 \leq p, q \leq n}. \quad (3.43)$$

By (3.43), the $p$-th element of $\Delta_i$, say, $z_{pi}$, follows $N(0, \sum_{j=1}^{i-1} \gamma_j y_j)$ conditionally on $\{y_1, y_2, \ldots, y_{i-1}\}$. Let $\{\xi, \xi_i, i = 1, 2, \ldots, n\}$ be a sequence of independent standard normals which are also independent of $\{y_1, y_2, \ldots, y_n\}$. Then $\mathcal{L}(z_{pi}) = \mathcal{L}(\xi \cdot (\sum_{j=1}^{i-1} \gamma_j y_j)^{1/2})$.

Since $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$ has the normalized Haar distribution on $O(n)$, the $p$-th row of $\Gamma$ is uniformly distributed on the $n$-dimensional sphere $S^{n-1}$. Thus, the law of $\sum_{j=1}^{i-1} \gamma_j y_j$ is the same as that of $\eta_i^2 := (\sum_{k=1}^{i-1} \xi_k^2)/(\sum_{k=1}^{n} \xi_k^2)$. In summary, $\mathcal{L}(z_{pi}) = \mathcal{L}(\eta_i)$. It follows that

$$P\left( \max_{2 \leq i \leq m} \|\Delta_i\| \geq t \right) \leq mn \cdot \max_{1 \leq p, q \leq n, 2 \leq i \leq m} P(|z_{pi}| \geq t) = mn \cdot \max_{2 \leq i \leq m} P(|\xi| \geq t/\eta_i) \leq mnP(|\xi| \geq t/\eta_{m+1}).$$

The desired conclusion follows from Lemma 3.4. \hfill \blacksquare

The next lemma, proved by large deviations, is also a key step as Lemma 3.5 to analyze the tail of $\epsilon_n(m)$.

**Lemma 3.6** Let $\{y_1, y_2, \ldots, y_n\}$ be a sequence of i.i.d. $\mathbb{R}^n$-valued random vectors with $\mathcal{L}(y_1) = \mathcal{N}(0, I_n)$. We have that

$$P\left( \max_{1 \leq i \leq m} L_i \geq r \right) \leq 4me^{-nr^2/16}$$

for all $r \in (0, 1/4)$ and $m \leq nr/2$, where $L_i$ is defined in (3.39).

**Proof.** Obviously, for any $i$,

$$P(L_i \geq r) \leq P(\sqrt{n}/\|w_i\|^2 \leq 1 - r) + P(\sqrt{n}/\|w_i\|^2 \geq 1 + r). \quad (3.44)$$

By (3.38) and the orthogonality, $\|w_i\|^2 = \|y_i\|^2 - \sum_{j=1}^{i-1} (y_j^T \gamma_j)^2 \leq \|y_i\|^2$. Also, $\mathcal{L}(y_i) = \mathcal{L}(y_1)$. Then, by the first inequality of (iii) of Lemma 3.3 and (i) of Lemma 4.2,

$$\max_{1 \leq i \leq n} P(\sqrt{n}/\|w_i\|^2 \leq 1 - r) \leq P\left( \frac{\|y_i\|^2}{n} \geq 1 + 2r \right) \leq 2e^{-n\lambda} \quad (3.45)$$

20
for $r \in (0, 1/4)$ where $\lambda := \inf_{x \geq 1 + 2r} I(x)$ and $I(x)$ is given in (ii) of Lemma 3.2. Since $I(x)$ is increasing on $[1, \infty)$, $\lambda = I(1 + 2r) = (2r - \log(1 + 2r))/2 \geq r^2/2$ for $r \in (0, 1/4)$ by (ii) of Lemma 3.3. So

$$
\max_{1 \leq i \leq m} P(\sqrt{n/\|w_i\|^2} \leq 1 - r) \leq 2e^{-nr^2/2}
$$

(3.46)

for any $r \in (0, 1/4)$.

Now we estimate the last term in (3.44). By the second inequality of (iii) of Lemma 3.3, $(1 + r)^{-2} \leq 1 - r$ for $r \in (0, 1/4)$. It follows that

$$
P(\sqrt{n/\|w_i\|^2} \geq 1 + r) \leq P \left( \frac{\|w_i\|^2}{n} \leq 1 - r \right).
$$

(3.47)

Recall the definition of $w_i$ in (3.38), by the fact that $(y_i^T \gamma_j) \gamma_j = \gamma_j y_j^T$, we can rewrite $w_i = By_i$, where $B = I_n - \sum_{j=1}^{i-1} \gamma_j \gamma_j^T$. Observe that $\gamma_j$ is a function of $y_1, y_2, \ldots, y_j$. Thus, $\gamma_1, \gamma_2, \ldots, \gamma_{i-1}$ and $y_i$ are independent. By orthogonality, $B^2 = B$. It follows that

$$
w_i \sim N(0, I_n - \sum_{j=1}^{i-1} \gamma_j \gamma_j^T)
$$

conditionally on $y_1, y_2, \ldots, y_{i-1}$. Since $B^2 = B$, the rank of $B$ is equal to $\text{tr}(B) = \text{tr}(I_n) - \sum_{j=1}^{i-1} \text{tr}(\gamma_j \gamma_j^T) = n - i + 1$. By Lemma 4.3, there exists a sequence of independent standard normals $\{\xi_i, \xi_i, i = 1, 2, \ldots, n\}$ which are independent of $\{y_1, y_1, \ldots, y_n\}$ such that $\mathcal{L}(\|w_i\|^2) = \mathcal{L}(\sum_{j=1}^{n-i+1} \xi_j^2)$ conditionally on $y_1, y_2, \ldots, y_{i-1}$. This implies that $\mathcal{L}(\|w_i\|^2) = \mathcal{L}(\sum_{j=1}^{n-i+1} \xi_j^2)$ unconditionally. Note that $\sum_{j=1}^{n-i+1} \xi_j^2 \geq \sum_{j=1}^{n-m} \xi_j^2$ for $1 \leq i \leq m$.

By using (i) of Lemma 4.2, we have that

$$
\max_{1 \leq i \leq m} P \left( \frac{\|w_i\|^2}{n} \leq 1 - r \right) \leq P \left( \frac{1}{n - m} \sum_{j=1}^{n-m} \xi_j^2 \leq a \right) \leq 2e^{-(n-m)I(A)} = 2e^{-(n-m)I(a)},
$$

(3.48)

where $a := n(1 - r)/(n - m)$, $A = (-\infty, a]$ and $I(x)$ is as in Lemma 3.2. We use the fact that $I(x)$ is decreasing on $(0, 1)$ in the equality, and $a < 1$ since $m \leq nr/2$. By (i) of Lemma 3.3,

$$
(n - m)I(a) \geq (n - m) \cdot \frac{(1 - a)^2}{4} \geq \frac{(nr - m)^2}{4(n - m)} \geq \frac{nr^2}{16}
$$

(3.49)

as $m \leq nr/2$. Now, combining (3.47), (3.48) and (3.49), we obtain

$$
\max_{1 \leq i \leq m} P(\sqrt{n/\|w_i\|^2} \geq 1 + r) \leq 2e^{-nr^2/16}.
$$

(3.50)
This together with (3.44) and (3.46) implies that
\[ P(\max_{1 \leq i \leq m} L_i \geq r) \leq m \cdot \max_{1 \leq i \leq m} P(L_i \geq r) \leq 4me^{-nr^2/16}. \]

We now are ready to prove Theorems 5 and 6.

**Proof of Theorem 5.** Let \( \{y_1, y_2, \ldots, y_n\} \) be a sequence of real-valued i.i.d. \( n \)-dimensional random vectors with \( \mathcal{L}(y_1) = N(0, I_n) \). Let also \( y_{ij} \) be the \( i \)-th element of \( y_j \). We prove the theorem by performing the Gram-Schmidt procedure on \( y_i \)'s. If \( \max_{1 \leq i \leq m} \|\Delta_i\| \leq t \), \( \max_{1 \leq i \leq m} L_i \leq r \) and \( \max_{1 \leq i \leq m, 1 \leq j \leq n} |y_{ij}| \leq s \), then \( \epsilon_n(m) \leq rs + 2t \) for \( r \in (0, 1/4) \) by Lemma 3.1. Then

\[
P(\epsilon_n(m) > rs + 2t) \leq P(\max_{1 \leq i \leq m} \|\Delta_i\| > t) + P(\max_{1 \leq i \leq m} L_i > r) + P(\max_{1 \leq i \leq m, 1 \leq j \leq n} |y_{ij}| > s).
\]

By Lemma 4.1, it is easy to see that

\[
P(\max_{1 \leq i \leq m, 1 \leq j \leq n} |y_{ij}| \geq s) \leq \frac{mn}{\sqrt{2\pi s}}e^{-s^2/2}
\]

for any \( s > 0 \). This together with Lemmas 3.5 and 3.6 yields the desired inequality.

Finally, we prove Theorem 6 by using the same argument as that of Theorem 5. The major difference is that the squared norm of a standard complex normal follows the exponential distribution with parameter one.

**Proof of Theorem 6.** For a complex vector \( \mu \) and a positive semidefinite complex matrix \( H \), denote by \( \mathbb{C}N_n(\mu, H) \) the \( n \)-dimensional complex normal distribution with mean \( \mu \) and covariance matrix \( H \). The complex normal distribution is uniquely determined by its mean and covariance matrix, e.g., Theorem 2.7 from [1] or p.374 from [16]. Let \( \{y_1, y_2, \ldots, y_n\} \) be a sequence of complex-valued i.i.d. \( n \)-dimensional random vectors with \( \mathcal{L}(y_1) = \mathbb{C}N_n(0, I_n) \). Then there exist two independent sequences of i.i.d. real-valued random variables \( \{\xi, \xi_j, j = 1, 2, \ldots\} \) and \( \{\eta, \eta_j, j = 1, 2, \ldots\} \) with the law \( N(0, 1) \) such that they are independent of \( \{y_1, y_2, \ldots, y_n\} \) and the distribution of \( y_1 \) is equal to that of \( (1/\sqrt{2})(\xi_1 + i\eta_1, \xi_2 + i\eta_2, \ldots, \xi_n + i\eta_n)^T \). We prove the theorem next by performing the Gram-Schmidt procedure for \( \{y_1, y_2, \ldots, y_n\} \) as at the beginning of this section. Then \( \Gamma_n = (\gamma_1, \gamma_2, \cdots, \gamma_n) \) is an unitary invariant matrix.

By Lemma 3.1 and the same argument as in the proof of Theorem 5, we only need to estimate the tail probabilities of maximum of random variables \( \Delta_i, L_i \) and \( |y_{ij}| \) over certain indices, respectively, where \( y_{ij} \) is the \( i \)-th element of \( y_j \).
First, $\mathcal{L}(|y_{ij}|^2) = \mathcal{L}((\xi^2 + \eta^2)/2)$. Note that $(\xi^2 + \eta^2)/2$ follows the exponential distribution $\text{Exp}(1)$. We then have that

$$P\left( \max_{1 \leq i \leq m, 1 \leq j \leq n} |y_{ij}| \geq s \right) \leq mnP((\xi^2 + \eta^2)/2 \geq s^2) = mne^{-s^2}. \quad (3.51)$$

Re-examining the proof of Lemma 3.5, in the current complex normal case, conditionally on $y_1, y_2, \cdots, y_{i-1}$, we see that

$$\Delta_i \sim \mathbb{C}N_n(0, U_i U_i^*) \quad \text{and} \quad w_i \sim \mathbb{C}N_n(0, I_n - U_i U_i^*), \quad (3.52)$$

where $U_i = (\gamma_1, \gamma_2, \cdots, \gamma_{i-1}) = (\gamma_{pq})$. Clearly, the $p$-th element of $\Delta_i$, say, $z_{pi}$, has the same distribution as $\mathcal{L}(\lambda_i (\xi + \eta \sqrt{-1}))$, where $\lambda_i := (\sum_{j=1}^{i-1} |\gamma_{pq}|^2/2)^{1/2}$. By the Haar invariance of $U_n$, $\mathcal{L}(\gamma_{pq}, \gamma_{pq}, \cdots, \gamma_{pq}) = \mathcal{L}(\gamma_1) = \mathcal{L}(y_i//|y_i|)$. So, $\mathcal{L}(\lambda_i^2) = \mathcal{L}((\sum_{k=1}^{2(i-1)} \xi_k^2)/(2 \sum_{k=1}^{2n} \xi_k^2))$. Consequently,

$$P\left( \max_{1 \leq i \leq m} \|\Delta_i\| \geq t \right) \leq mn \max_{1 \leq i \leq m, 1 \leq p \leq n} P(|z_{pi}| \geq t) \leq 2mn \max_{1 \leq i \leq m, 1 \leq p \leq n} P(|\xi| \geq t/(2\lambda_i)) \leq 2mnP(|\xi| \geq t/(2\lambda_{m+1})).$$

where $\lambda_{m+1} = (\sum_{k=1}^{2n} \xi_k^2 / \sum_{k=1}^{2n} \xi_k^2)^{1/2}$ and the fact that $|\lambda_i (\xi + \eta \sqrt{-1})| \leq 2\lambda_i \max\{|\xi|, |\eta|\}$ is used in the last inequality. By Lemma 3.4, we have that

$$P\left( \max_{1 \leq i \leq m} \|\Delta_i\| \geq t \right) \leq \frac{12mn}{\sqrt{2\pi}} \left( 1 + \frac{t^2}{12m + t\sqrt{n}} \right)^{-n}. \quad (3.53)$$

Recall the proof of Lemma 3.6, two key steps to estimate the tail probability of $L_i$ are (3.45) and (3.50). In our current case, $\|y_1\|^2$ has the same law as that of $(1/2) \sum_{i=1}^{2n} \xi_i^2$. Thus, by following the proof of (3.46), we have

$$\max_{1 \leq i \leq m} P\left( \sqrt{n/\|w_i\|^2} \leq 1 - r \right) \leq 2e^{-n r^2} \quad (3.54)$$

for any $r \in (0, 1/4)$.

Now we turn to the estimate of $P\left( \sqrt{n/\|w_i\|^2} \geq 1 + r \right)$. Note that

$$(I_n - U_i U_i^*)^* = I_n - U_i U_i^* \quad \text{and} \quad (I_n - U_i U_i^*)^2 = I_n - U_i U_i^*.$$ 

By the spirit of the proof of Lemma 4.3, we obtain that $\mathcal{L}(\|w_i\|^2) = \mathcal{L}((1/2) \sum_{j=1}^{2n+i+1} \xi_j^2)$. Now repeating the corresponding calculations in the proof of Lemma 3.6, it follows that

$$\max_{1 \leq i \leq m} P\left( \sqrt{n/\|w_i\|^2} \geq 1 + r \right) \leq 2e^{-n r^2/8}. \quad (3.55)$$
From (3.54) and (3.55) we obtain that
\[
P(\max_{1 \leq i \leq m} L_i \geq r) \leq 4 m e^{-nr^2/8}. \tag{3.56}
\]
The proof is completed by adding up the three probabilities respectively in (3.51), (3.53) and (3.56). \hfill \blacksquare

4 Appendix

In this section we list some known results used in the previous sections.

The following is Lemma 3 on page 49 from [6].

**Lemma 4.1** Suppose \( X \sim N(0,1) \). Then
\[
\frac{1}{\sqrt{2\pi}} \cdot \frac{x}{1 + x^2} e^{-x^2/2} \leq P(X > x) \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{x} e^{-x^2/2}
\]
for all \( x > 0 \).

For \( A \subset \mathbb{R} \), the notation \( A^o \) and \( \bar{A} \) stand for the interior and the closure of \( A \) in \( \mathbb{R} \), respectively. The first part of next lemma gives sharp estimates of rare events induced by partial sums of independent random variables (e.g., (c) of Remarks on page 27 from [9]). Taking \( d = 1 \) and \( C = \sigma^2 \) from Theorem 3.7.1 on page 109 from [9], we obtain the second part of next lemma, which is called moderate deviations.

**Lemma 4.2** Let \( \{X, X_i; i = 1, 2, \cdots\} \) be a sequence of i.i.d. random variables. Let \( S_n = \sum_{i=1}^{n} X_i, n \geq 1 \). Then
(i) For any \( A \subset \mathbb{R} \) and \( n \geq 1 \),
\[
P(S_n/n \in A) \leq 2 e^{-nI(A)},
\]
where \( I(x) = \sup_{t \in \mathbb{R}} \{tx - \log E(e^{tX})\} \) and \( I(A) = \inf_{x \in A} I(x) \).

(ii) Assume further that \( EX = 0 \), \( \var(X) = \sigma^2 > 0 \) and \( Ee^{t_0 X} < \infty \) for some \( t_0 > 0 \). Let \( \{a_n; n = 1, 2, \cdots\} \) be a sequence of positive numbers such that \( a_n \to 0 \) and \( na_n \to \infty \) as \( n \to \infty \). Then
\[
\lim_{n \to \infty} a_n \log P \left( \sqrt{n} S_n \in A \right) = - \inf_{x \in A} \left\{ \frac{x^2}{2\sigma^2} \right\}
\]
for any subset \( A \subset \mathbb{R} \) such that \( \inf \{ |x|; x \in A^o \} = \inf \{ |x|; x \in \bar{A} \} \).
The following is (ii) on p.186 from [34].

**Lemma 4.3** Suppose \( y \) is a \( \mathbb{R}^n \)-valued random vector with multi-normal distribution with mean 0 and covariance matrix \( \Sigma \) of rank \( r \). If \( \Sigma^2 = \Sigma \), then there exists a sequence of i.i.d. random variables \( \{ \xi_j; j = 1, 2, \cdots , n \} \) with the standard normal distribution such that \( \|y\|^2 \) has the same distribution as that of \( \sum_{j=1}^{r} \xi_j^2 \), that is, \( \|y\|^2 \sim \chi^2(r) \).

**Acknowledgment.** I am greatly indebted to my colleague Morris Eaton for teaching me the theory of Haar integral as well as discussing this paper with me. The author also very much thanks an anonymous referee for his very careful reading and pointing out some relevant papers. The comments by the referee help the author improve the exposition of this paper significantly.

**References**


email: tjiang@stat.umn.edu.
Mailing Address: School of Statistics, 313 Ford Hall, 224 Church Street S.E., Minneapolis, MN55455.