The Entries of Haar-invariant Matrices from the Classical Compact Groups

Tiefeng Jiang [1]

Abstract Let $\Gamma_n = (\gamma_{ij})_{n\times n}$ be a random matrix with the Haar probability measure on the orthogonal group $O(n)$, the unitary group $U(n)$ or the symplectic group $Sp(n)$. Given $1 \le m < n$, a probability inequality for a distance between $(\gamma_{ij})_{n\times m}$ and some mn independent F-valued normal random variables is obtained, where $F = \mathbb{R},\ \mathbb{C}$ or $\mathbb{H}$ (the set of real quaternions). The result is universal for the three cases. In particular, the inequality for $Sp(n)$ is new.

1 Introduction

To study a statistic related to an image analysis problem in Donoho and Huo [7], Jiang [12, 13] developed a method to approximate the entries of Haar-invariant orthogonal or unitary matrices by i.i.d. (independent and identically distributed) real or complex normal random variables. With the approximation result, the asymptotic distributions of the largest entries of Haar orthogonal or unitary matrices are obtained in [13]. Moreover, by the same tool, the Marchenko-Pastur law for the eigenvalues of the Jacobi matrices are derived in [11].

There are three classical compact groups: orthogonal group $O(n)$, unitary group $U(n)$ and symplectic group $Sp(n)$. The approximation results mentioned above are only for Haar-invariant orthogonal and unitary matrices, which generate the Haar probability measures on $O(n)$ and $U(n)$, respectively. Here, by saying "the Haar probability measure" on $O(n),\ U(n)$ or $Sp(n)$, we mean the unique Haar-invariant measure μ such that $\mu(G) = 1$ for $G = O(n),\ U(n)$ or $Sp(n)$. Sometimes we also call μ a normalized Haar measure. The purpose of this paper is to study the entries of random matrices generating the Haar probability measures on the three groups simultaneously. In particular, we obtain a new result on the symplectic groups.

Before stating our main results, we will review some background about the classical compact groups.

Let $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively. Write $\mathbb{H} = \{a = a_1 + a_2 i + a_3 j + a_4 k;\ a_1, a_2, a_3 \text{ and } a_4 \in \mathbb{R}\}$ for the set of real quaternions, where $i^2 = j^2 = k^2 = -1$ and $ij = k,\ jk = i,\ ki = j$. For $a = a_1 + a_2 i + a_3 j + a_4 k \in \mathbb{H}$, the conjugate of a is $a^* = a_1 - a_2 i - a_3 j - a_4 k$, and the absolute value of a is $|a| = \sqrt{a^* a} = \sqrt{a a^*} = (a_1^2 + a_2^2 + a_3^2 + a_4^2)^{1/2}$. Given matrix $\mathbf{A} = (a_{pq})$

[1] Supported in part by NSF#DMS-0449365, School of Statistics, University of Minnesota, 224 Church Street, MN55455, tjiang@stat.umn.edu.

Key Words: Random matrix, Haar measure, classical compact group, probability inequality, independence, Gaussian distribution.

AMS (2000) subject classifications: 15A52, 60B15, 62E17.

with $a_{pq} \in \mathbb{H}$, the transpose of $\mathbf{A}$ is $\mathbf{A}^T = (a_{qp})$; the conjugate of $\mathbf{A}$ is $\mathbf{A}^* = (a_{qp}^*)$. Recall

$$\begin{aligned}
O(n) &= \{A = (a_{pq})_{n\times n} : A^T A = I_n \text{ and } a_{pq} \in \mathbb{R} \text{ for all } 1 \le p, q \le n\}; \\
U(n) &= \{A = (a_{pq})_{n\times n} : A^* A = I_n \text{ and } a_{pq} \in \mathbb{C} \text{ for all } 1 \le p, q \le n\}; \\
Sp(n) &= \{A = (a_{pq})_{n\times n} : A^* A = I_n \text{ and } a_{pq} \in \mathbb{H} \text{ for all } 1 \le p, q \le n\}.
\end{aligned} \tag{1.1}$$

They are in order called the orthogonal, unitary and symplectic groups, respectively, see p.111 and p.113 from [14], or p.90 and p.92 from [17]. The symplectic group $Sp(n)$ sometimes is also called quaternionic unitary group. The three groups in (1.1) are all compact.

Now we recall three standard normal distributions that will be used later. Let ξ_1, ξ_2, ξ_3 and ξ_4 be i.i.d. random variables with distribution $N(0,1)$.

(a) The *standard real normal distribution* is simply $N(0,1)$. For convenience of notation, sometimes we also write $\mathbb{R}N(0,1)$ for $N(0,1)$.

(b) The *standard complex normal distribution*, denoted by $\mathbb{C}N(0,1)$, is the probability distribution of $(\xi_1 + i\xi_2)/\sqrt{2}$.

(c) The *standard quaternion normal distribution*, denoted by $\mathbb{H}N(0,1)$, is the probability distribution of $(\xi_1 + i\xi_2 + j\xi_3 + k\xi_4)/\sqrt{4}$.

The following approximation results are obtained in Jiang [13].

Theorem A.1 *(Theorem 4 in [13]) For each $n \ge 2$, there exist matrices $\mathbf{\Gamma}_n = (\gamma_{ij})_{1\le i,j\le n}$ and $\mathbf{Y}_n = (y_{ij})_{1\le i,j\le n}$ whose $2n^2$ elements are random variables defined on the same probability space such that*

(i) the law of $\mathbf{\Gamma}_n$ is the normalized Haar measure on the orthogonal group O_n;

(ii) $\{y_{ij}; 1 \le i, j \le n\}$ are i.i.d. random variables with the real standard normal distribution;

(iii) set $\epsilon_n(m) = \max_{1\le i\le n, 1\le j\le m} |\sqrt{n}\gamma_{ij} - y_{ij}|$ for $m = 1, 2, \cdots, n$. Then

$$P(\epsilon_n(m) \ge rs + 2t) \le 4me^{-nr^2/16} + 3mn\left(\frac{1}{s}e^{-s^2/2} + \frac{1}{t}\left(1 + \frac{t^2}{3(m+t\sqrt{n})}\right)^{-n/2}\right)$$

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

For the complex case, the following result holds.

Theorem A.2 *(Theorem 5 in [13]) For each $n \ge 2$, there exist two $n \times n$ matrices $\mathbf{\Gamma}_n = (\gamma_{pq})$ and $\mathbf{Y}_n = ((x_{pq} + iy_{pq})/\sqrt{2})$ such that γ_{pq}'s, x_{pq}'s and y_{pq}'s are random variables defined on the same probability space, and*

(i) the law of $\mathbf{\Gamma}_n$ is the normalized Haar measure on the unitary group $U(n)$;

(ii) the $2n^2$ random variables $\{x_{pq}, y_{pq}; 1 \le p, q \le n\}$ are independent real standard normals;

(iii) set $\epsilon_n(m) = \max_{1\le p\le n, 1\le q\le m} |\sqrt{n}\gamma_{pq} - (x_{pq} + iy_{pq})/\sqrt{2}|$ for $m = 1, 2, \cdots, n$. Then

$$P(\epsilon_n(m) \ge rs + 2t) \le 4me^{-nr^2/8} + mne^{-s^2} + \frac{6mn}{t}\left(1 + \frac{t^2}{12(m+t\sqrt{n})}\right)^{-n}$$

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

The above results are on orthogonal and unitary groups. In this paper, we will prove an approximation result for the symplectic group $Sp(n)$. In fact, unlike the method used in [13], which deals with $O(n)$ and $U(n)$ case by case, here we are able to treat the three cases simultaneously. The symplectic case is simply a corollary. The following is our main result.

THEOREM 1 *For each $n \geq 2$, there exist matrices $\mathbf{\Gamma}_n = (\gamma_{ij})_{1\leq i,j\leq n}$ 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 $\mathbf{\Gamma}_n$ is the normalized Haar measure on $O(n)$, $U(n)$ or $Sp(n)$;

(ii) $\{y_{ij}; 1 \leq i, j \leq n\}$ are i.i.d. random variables with the standard real, complex or quaternion normal distribution;

(iii) set $\epsilon_n(m) = \max_{1\leq i\leq n, 1\leq j\leq m} |\sqrt{n}\gamma_{ij} - y_{ij}|$ for $1 \leq m \leq n$. Then

$$
\begin{aligned}
& P\left(\epsilon_n(m) \geq rs + t\right) \\
\leq\ & 4me^{-\beta nr^2/16} + mn \cdot \chi_\beta(n,\, s) \cdot e^{-\beta s^2/2} + 3mn \cdot \chi_\beta(n, t) \cdot \left(1 + \frac{t^2}{3(m + t\sqrt{n})}\right)^{-\beta n/2}
\end{aligned}
$$

for any $r \in (0, 1/4)$, $s > 0$, $t > 0$ and $1 \leq m \leq nr/2$, where $\chi_\beta(n,\, s) = 1/s, 1$ or $8ns^2 + 1$ for $\beta = 1, 2$ or 4 according to $\mathbf{\Gamma}_n \in O(n)$, $U(n)$ or $Sp(n)$.

By taking $\beta = 1$ and 2 respectively in the above result, we see that the upper bounds of $P\left(\epsilon_n(m) \geq rs + t\right)$ in Theorems A.1 and A.2 and those in Theorem 1 are almost the same except the coefficients, which do not affect the orders of the bounds. The result corresponding to $\beta = 4$ is new.

Theorem 1 says that the first m columns of the $n\times n$ Haar-invariant matrices can be approximated by i.i.d. normal random variables. How large $m = m_n$ can be such that the error $\epsilon_n(m)$ goes to zero in probability as $n \to \infty$?

COROLLARY 1.1 *Let $\epsilon_n(m)$ be as in Theorem 1. If $m_n = o(n/(\log n))$, then $\epsilon_n(m_n) \to 0$ in probability as $n \to \infty$ for all of the three cases.*

It is shown in [12] that, for the case of orthogonal group $O(n)$, if $m_n = o(n/\log n)$, then $\epsilon_n(m_n) \to 0$ in probability as $n \to \infty$. Further, for any $\alpha > 0$, take $m_n = [n\alpha/\log n]$, it is proved that $\epsilon_n(m_m) \to 2\sqrt{\alpha}$ in probability as $n \to \infty$. This concludes that $m_n = o(n/\log n)$ is the largest order to make $\epsilon_n(m_n) \to 0$ in probability. We conjecture that the same is also true for unitary and symplectic groups.

In this paper, for matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$, we use the maximum norm $\|\mathbf{A} - \mathbf{B}\|_{\max} = \max_{i,j} |a_{ij} - b_{ij}|$ to measure the distance between $\mathbf{A}$ and $\mathbf{B}$. The variation norm is another way to study approximations of the entries of Haar-invariant matrices. The variation norm, which is stronger than the maximum norm, is used to investigate similar problems in [1, 2, 4, 5, 6, 9, 11, 12, 18, 19]. In particular, it is shown in [12] that, with the variation norm, the $p \times q$ upper-left block of an $n \times n$ Haar-invariant orthogonal matrix can be approximated by i.i.d. real standard normals, where $p = o(\sqrt{n})$ and $q = o(\sqrt{n})$. It is also proved in [12] that the order $o(\sqrt{n})$ is optimal.

The proof of Theorem 1 relies on the Gram-Schmidt algorithm. In fact, there are several ways to generate Haar-orthogonal, unitary and symplectic matrices, see, e.g., Mezzadri [16]. We find that the Gram-Schmidt method is quite convenient to construct Haar-invariant matrices so that they can be approximated by independent normal random variables efficiently. The Gram-Schmidt algorithm transforms a matrix of i.i.d. real, complex or symplectic Gaussian entries to an Haar-invariant matrix from $O(n)$, $U(n)$ or $Sp(n)$. Theorems A.1 and A.2 were proved by the Gram-Schmidt algorithm in Jiang [13]. However, Theorem A.1 (real case) and Theorem A.2 (complex case) were treated separately. Here we are able to provide a universal inequality for the three cases.

We prove Theorem 1 in the next section; we provide some results on matrices with quaternion entries in the Appendix.

2 Proofs of Main Results

Throughout this section, given $n \geq 1$ and $F = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, we assume that $\{y_{pq};\ 1 \leq p,\, q \leq n\}$ are i.i.d. random variables with common distribution $FN(0,1)$. Write $\mathbf{Y} = (y_{pq})_{n\times n} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)$, where the $\{\mathbf{y}_r;\ 1 \leq r \leq n\}$ are $n \times 1$ column vectors.

Review that $h^* = \alpha_1 - \alpha_2 i - \alpha_3 j - \alpha_4 k$ if $h = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{H}$. Moreover, if $\mathbf{a} = (a_1, \cdots, a_n)^T \in \mathbb{H}^n$ and $\mathbf{b} = (b_1, \cdots, b_n)^T \in \mathbb{H}^n$, then $\mathbf{a}^*\mathbf{b} = \sum_{r=1}^n a_r^* b_r$, and the Euclidean norm of $\mathbf{a}$ is $\|\mathbf{a}\| = (\sum_{r=1}^n |a_r|^2)^{1/2}$. We will use the Gram-Schmidt procedure to prove Theorem 1. Let us first review it. Set

$$\begin{aligned} &\mathbf{w}_1 = \mathbf{y}_1 \text{ and } \mathbf{e}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}; \\ &\mathbf{w}_p = \mathbf{y}_p - \sum_{q=1}^{p-1} \mathbf{e}_q \cdot (\mathbf{e}_q^* \mathbf{y}_p) \text{ and } \mathbf{e}_p = \frac{\mathbf{w}_p}{\|\mathbf{w}_p\|} \text{ for } 2 \leq p \leq n. \end{aligned} \tag{2.1}$$

In the real and complex cases, the positions of $\mathbf{e}_q$ and $\mathbf{e}_q^*\mathbf{y}_p$ in the product $\mathbf{e}_q \cdot (\mathbf{e}_q^*\mathbf{y}_p)$ are not vital. However, it indeed makes a difference for the quaternion case ($F = \mathbb{H}$) because the quaternion numbers are not commutative. The following lemma is useful.

Lemma 2.1 *(Lemma in 2.1 in [10]) Let the notation be as in (2.1). Then* $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ *is an Haar-invariant orthogonal, unitary or symplectic matrix according to* $F = \mathbb{R}$, $\mathbb{C}$ *or* $\mathbb{H}$*. Further, any row or column of* $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ *has the same distribution as that of*

(i) $\dfrac{1}{\|Z_n^1\|}(\xi_{11}, \cdots, \xi_{n1})^T$ *for* $F = \mathbb{R}$;

(ii) $\dfrac{1}{\|Z_n^2\|}(\xi_{11} + i\xi_{12}, \cdots, \xi_{n1} + i\xi_{n2})^T$ *for* $F = \mathbb{C}$;

(iii) $\dfrac{1}{\|Z_n^4\|}(\xi_{11} + i\xi_{12} + j\xi_{13} + k\xi_{14}, \cdots, \xi_{n1} + i\xi_{n2} + j\xi_{n3} + k\xi_{n4})^T$ *for* $F = \mathbb{H}$,

where $\xi_{pq};\ 1 \leq p \leq n,\ 1 \leq q \leq 4$ *are independent* $N(0,1)$*-distributed random variables and* $\|Z_n^\beta\|^2 = \sum_{1\leq p\leq n, 1\leq q\leq \beta} \xi_{pq}^2$ *for* $\beta = 1, 2, 4$.

Set

$$\mathbf{\Sigma}_0 = \mathbf{0},\ \mathbf{\Sigma}_k = (\mathbf{e}_1, \cdots, \mathbf{e}_k)(\mathbf{e}_1, \cdots, \mathbf{e}_k)^*, \quad 1 \leq k \leq n. \tag{2.2}$$

Then,

$$\begin{aligned}\sum_{i=1}^{k-1}\mathbf{e}_i\cdot(\mathbf{e}_i^*\mathbf{y}_k) &= \{(\mathbf{e}_1,\cdots,\mathbf{e}_{k-1})(\mathbf{e}_1,\cdots,\mathbf{e}_{k-1})^*\}\,\mathbf{y}_k\\ &= \mathbf{\Sigma}_{k-1}\mathbf{y}_k\end{aligned}$$

for $2\le k\le n$. It follows that

$$\mathbf{w}_k=\mathbf{y}_k-\sum_{i=1}^{k-1}\mathbf{e}_i\cdot(\mathbf{e}_i^*\mathbf{y}_k)=(\mathbf{I}-\mathbf{\Sigma}_{k-1})\,\mathbf{y}_k,\quad 1\le k\le n. \tag{2.3}$$

For $\mathbf{x}=(x_1,\cdots,x_n)^T\in F^n$, where $F=\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, set $\|\mathbf{x}\|=\max_{1\le i\le n}|x_i|$.

LEMMA 2.2 *Let $\mathbf{y}_k,\mathbf{e}_k,\mathbf{w}_k$ and $\mathbf{\Sigma}_k$ be as in (2.1)-(2.3). Set $a_k=\sqrt{n}/\|\mathbf{w}_k\|$. Then*

$$\begin{aligned}&P\left(\max_{1\le k\le m}\|\sqrt{n}\mathbf{e}_k-\mathbf{y}_k\|\ge rs+t\right)\\ \le\;& P\left(\max_{1\le k\le m}|a_k-1|\ge r\right)+P\left(\max_{1\le k\le m}\|\mathbf{w}_k\|\ge s\right)+P\left(\max_{1\le k\le m}\|\mathbf{\Sigma}_{k-1}\mathbf{y}_k\|\ge t\right)\end{aligned}$$

for any $n\ge m\ge 1$, $r>0$, $s>0$ and $t>0$.

Proof. It suffices to show that

$$\|\sqrt{n}\mathbf{e}_k-\mathbf{y}_k\|\le|a_k-1|\cdot\|\mathbf{w}_k\|+\|\mathbf{\Sigma}_{k-1}\mathbf{y}_k\| \tag{2.4}$$

for $1\le k\le n$. Actually, if this is true, and $|a_k-1|<r$, $\|\mathbf{w}_k\|<s$ and $\|\mathbf{\Sigma}_{k-1}\mathbf{y}_k\|<t$, then the left hand side of (2.4) is less than $rs+t$. Thus, the lemma follows by taking the complement events. Now we prove (2.4). Use $\mathbf{w}_k=(\mathbf{I}-\mathbf{\Sigma}_{k-1})\mathbf{y}_k$ to obtain

$$\begin{aligned}\sqrt{n}\mathbf{e}_k-\mathbf{y}_k=a_k\mathbf{w}_k-\mathbf{y}_k &= (a_k-1)\mathbf{w}_k+(\mathbf{w}_k-\mathbf{y}_k)\\ &= (a_k-1)\mathbf{w}_k-\mathbf{\Sigma}_{k-1}\mathbf{y}_k\end{aligned}$$

for $1\le k\le n$. Then (2.4) follows by the triangle inequality. ■

In what follows, for convenience, we will use $U\stackrel{d}{=}V$ or $U\sim V$ to denote that random variables U and V have the same probability distribution. The notation $\chi^2(k)$ stands for the χ-square distribution with degree of freedom k.

LEMMA 2.3 *Given $1\le k\le n$ and $F=\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. Recall (2.1)-(2.3). We have that $\mathbf{A}(\mathbf{\Sigma}_{k-1}\mathbf{y}_k)\stackrel{d}{=}\mathbf{\Sigma}_{k-1}\mathbf{y}_k$ and $\mathbf{A}\mathbf{w}_k\stackrel{d}{=}\mathbf{w}_k$ for any $\mathbf{A}\in O(n)$, $U(n)$ or $Sp(n)$. In particular, all the entries of $\mathbf{\Sigma}_{k-1}\mathbf{y}_k$ have the same distribution. The same conclusion also holds for $\mathbf{w}_k$.*

Proof. Since $(\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_n)$ is Haar-invariant, $\mathbf{A}(\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_n)\stackrel{d}{=}(\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_n)$ for any $A\in O(n)$, $U(n)$ or $Sp(n)$ according to that $F=\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. Then, $\mathbf{A}(\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_k)\stackrel{d}{=}(\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_k)$ for any $1\le k\le n$. Now, recall (2.2) and (2.3),

$$\mathbf{A}(\mathbf{\Sigma}_{k-1}\mathbf{y}_k)=\mathbf{A}(\mathbf{e}_1,\cdots,\mathbf{e}_{k-1})\,(\mathbf{A}(\mathbf{e}_1,\cdots,\mathbf{e}_{k-1}))^*\,\mathbf{A}\mathbf{y}_k;$$
$$\mathbf{A}\mathbf{w}_k=(\mathbf{I}-\mathbf{A}(\mathbf{e}_1,\cdots,\mathbf{e}_{k-1})(\mathbf{A}(\mathbf{e}_1,\cdots,\mathbf{e}_{k-1}))^*)\mathbf{A}\mathbf{y}_k.$$

Observe that $(\mathbf{e}_1, \cdots, \mathbf{e}_{k-1})$ is a function of $\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}$, thus $\mathbf{y}_k$ and $(\mathbf{e}_1, \cdots, \mathbf{e}_{k-1})$ are independent, consequently $\mathbf{A}\mathbf{y}_k$ and $\mathbf{A}(\mathbf{e}_1, \cdots, \mathbf{e}_{k-1})$ are independent. Also, $\mathbf{A}\mathbf{y}_k \stackrel{d}{=} \mathbf{y}_k$ by the invariance of normal distribution. Then, $\mathbf{A}(\mathbf{\Sigma}_{k-1}\mathbf{y}_k) \stackrel{d}{=} (\mathbf{e}_1, \cdots, \mathbf{e}_{k-1})(\mathbf{e}_1, \cdots, \mathbf{e}_{k-1})^*\mathbf{y}_k = \mathbf{\Sigma}_{k-1}\mathbf{y}_k$. Similarly, $\mathbf{A}\mathbf{w}_k \stackrel{d}{=} \mathbf{w}_k$.

Finally, take $\mathbf{A}$ to be a permutation matrix, we see that the elements in the column vector $\mathbf{\Sigma}_{k-1}\mathbf{y}_k$ are exchangeable. This is also true for $\mathbf{w}_k$. ■

Let $\mathbf{A} = (a_{ij})_{n\times n}$ with $a_{ij} \in \mathbb{H}$ for all $1 \le i, j \le n$. Set $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$.

Lemma 2.4 *Let $F = \mathbb{R}\,(\beta = 1)$, $\mathbb{C}\,(\beta = 2)$ or $\mathbb{H}\,(\beta = 4)$. Let $\mathbf{A} = \mathbf{A}_{n\times n}$ be random, independent of $\mathbf{y}_1$ and satisfy that $\mathbf{A}^* = \mathbf{A}$, $\mathbf{A}^2 = \mathbf{A}$ and $tr(\mathbf{A}) = k$ for some constant $1 \le k \le n$. If $\mathbf{O}(\mathbf{A}\mathbf{y}_1) \sim \mathbf{A}\mathbf{y}_1$ for any $\mathbf{O} \in O(n)$, $U(n)$ or $Sp(n)$, then each entry of $\mathbf{A}\mathbf{y}_1 \sim \eta \cdot \left(\sum_{i=1}^{\beta k} \xi_i^2 / \sum_{i=1}^{\beta n} \xi_i^2\right)^{1/2}$, where $\eta \sim FN(0,1)$ and $\{\xi_i;\, i \ge 1\}$ are i.i.d. $N(0,1)$-distributed r.v.'s.*

Proof. If $k = n$, then $\mathbf{A} = \mathbf{I}$ by the assumption $\mathbf{A}^2 = \mathbf{A}$. The conclusion obviously holds. So, without loss of generality, assume $1 \le k < n$. We prove the lemma by three steps.

Step 1. By Lemmas 3.2 and 3.3, there exists $\mathbf{U} \in O(n)$, $U(n)$ or $Sp(n)$ according to $F = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$ such that $\mathbf{A} = \mathbf{U}^* \operatorname{diag}(\mathbf{I}_k, \mathbf{0})\mathbf{U}$. Since $\mathbf{U}$ is a function of $\mathbf{A}$, $\mathbf{U}$ is independent of $\mathbf{y}_1$. We claim that $\mathbf{U}$ and $\mathbf{U}\mathbf{y}_1$ are independent. In fact, take bounded, measurable and real-valued functions $f(\cdot)$ defined on the set of $n \times n$ matrices, and $g(\cdot)$ defined on F^n. By the given condition and the independence between $\mathbf{U}$ and $\mathbf{y}_1$, we have $\mathbf{U}\mathbf{y}_1 \sim \mathbf{y}_1$ conditioning on $\mathbf{U}$ or not, it follows that $Eg(\mathbf{U}\mathbf{y}_1) = Eg(\mathbf{y}_1)$, hence

$$\begin{aligned} E\{f(\mathbf{U})g(\mathbf{U}\mathbf{y}_1)\} &= E\{f(\mathbf{U}) \cdot E(g(\mathbf{U}\mathbf{y}_1)|\mathbf{U})\} \\ &= E\{f(\mathbf{U}) \cdot Eg(\mathbf{y}_1)\} = Ef(\mathbf{U}) \cdot Eg(\mathbf{U}\mathbf{y}_1). \end{aligned} \tag{2.5}$$

This proves the claim.

Step 2. Take an Haar-invariant random matrix $\mathbf{O} \in O(n)$, $U(n)$ or $Sp(n)$ according to $F = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$ for which $\mathbf{O}$ is independent of $\mathbf{U}$ and $\mathbf{y}_1$. By the given condition, $\mathbf{O}(\mathbf{A}\mathbf{y}_1) \sim \mathbf{A}\mathbf{y}_1$ for any $\mathbf{O}$, this is evidently still true when $\mathbf{O}$ is a random variable independent of $\mathbf{A}\mathbf{y}_1$. Since $\mathbf{U}$ and $\mathbf{U}\mathbf{y}_1$ are independent by the claim, we know $\mathbf{O}\mathbf{U}^*$ is independent of $\mathbf{U}\mathbf{y}_1$. Also, $\mathbf{O}\mathbf{U}^* \sim \mathbf{O}$ by the right-Haar invariance. Therefore,

$$\mathbf{A}\mathbf{y}_1 \stackrel{d}{=} \mathbf{O}\mathbf{A}\mathbf{y}_1 = (\mathbf{O}\mathbf{U}^*)\begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}(\mathbf{U}\mathbf{y}_1) \stackrel{d}{=} \mathbf{O}\begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{y}_1 = \mathbf{O}_{n\times k}\begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix},$$

where $\mathbf{O}_{n\times k}$ is the first k columns of $\mathbf{O}$, and $(y_1, \cdots, y_k)^T$ is the first k entries of $\mathbf{y}_1$. Thus, the first entry of $\mathbf{A}\mathbf{y}_1 \sim \sum_{i=1}^k e_{1i}y_i$, where $(e_{11}, \cdots, e_{1n})$ is the first row of $\mathbf{O}$. Since $\mathbf{O}$ and $\{y_i;\ 1 \le i \le n\}$ are independent, by Lemma 2.1, the first entry of $\mathbf{A}\mathbf{y}_1$ has the same distribution as that of

$$\frac{\sum_{i=1}^k \eta_i y_i}{(\sum_{i=1}^n |\eta_i|^2)^{1/2}} = \left(\frac{\sum_{i=1}^k |\eta_i|^2}{\sum_{i=1}^n |\eta_i|^2}\right)^{1/2} \cdot \sum_{i=1}^k \lambda_i y_i$$

where $\{\eta_i;\ 1 \le i \le n\}$ are i.i.d. $FN(0,1)$-distributed r.v.'s, and $\lambda_i = \eta_i(\sum_{i=1}^k |\eta_i|^2)^{-1/2}$. Obviously $\sum_{i=1}^k \lambda_i y_i \sim FN(0,1)$. Moreover, since η_i's are independent and $|\eta_i|^2 \sim \chi^2(\beta)/\beta$ for each $1 \le i \le n$, to finish the proof, it remains to show that $\sum_{i=1}^k \lambda_i y_i$ and $\sum_{i=1}^k |\eta_i|^2 / \sum_{i=1}^n |\eta_i|^2$ are independent.

Step 3. Since $\{y_i,\ \eta_i;\ 1 \le i \le n\}$ are i.i.d., and $\sum_{i=1}^n |\eta_i|^2 = \sum_{i=1}^k |\eta_i|^2 + \sum_{i=k+1}^n |\eta_i|^2$, it is enough to show $\sum_{i=1}^k \lambda_i y_i$ and $\sum_{i=1}^k |\eta_i|^2$ are independent. Noticing λ_i's and y_i's are independent, and $\sum_{i=1}^k \lambda_i y_i \sim FN(0,1)$ conditioning on η_i's or not. Take bounded, measurable functions $f_1(\cdot)$ defined on F and $g_1(\cdot)$ defined on $\mathbb{R}$, by the same argument as in (2.5), we obtain that

$$E\left\{f_1(\sum_{i=1}^k \lambda_i y_i) \cdot g_1(\sum_{i=1}^k |\eta_i|^2)\right\} = Ef_1(\sum_{i=1}^k \lambda_i y_i) \cdot Eg_1(\sum_{i=1}^k |\eta_i|^2).$$

This implies that $\sum_{i=1}^k \lambda_i y_i$ and $\sum_{i=1}^k |\eta_i|^2$ are independent. ■

Lemma 2.5 *Recall (2.1)-(2.3). For $F = \mathbb{R}\,(\beta = 1)$, $\mathbb{C}\,(\beta = 2)$ or $\mathbb{H}\,(\beta = 3)$, let η and $\{\xi_1, \xi_2, \cdots\}$ be as in Lemma 2.4. Then,*

(i) each entry of $\mathbf{\Sigma}_{k-1}\mathbf{y}_k \stackrel{d}{=} \eta \cdot \left(\sum_{i=1}^{\beta(k-1)} \xi_i^2 / \sum_{i=1}^{\beta n} \xi_i^2\right)^{1/2}$ for $2 \le k \le n$;

(ii) each entry of $\mathbf{w}_k \stackrel{d}{=} \eta \cdot \left(\sum_{i=1}^{\beta(n-k+1)} \xi_i^2 / \sum_{i=1}^{\beta n} \xi_i^2\right)^{1/2}$ for $1 \le k \le n$.

Proof. Review $\mathbf{\Sigma}_{k-1} = (\mathbf{e}_1, \cdots, \mathbf{e}_{k-1})(\mathbf{e}_1, \cdots, \mathbf{e}_{k-1})^*$ for $2 \le k \le n$, and $\mathbf{w}_k = (\mathbf{I} - \mathbf{\Sigma}_{k-1})\mathbf{y}_k$ for $1 \le k \le n$. It is not difficult to see from the orthogonality of $\mathbf{e}_i$'s that $\mathbf{A}^* = \mathbf{A}$, $\mathbf{A}^2 = \mathbf{A}$ for $\mathbf{A} = \mathbf{\Sigma}_{k-1}$ or $\mathbf{I} - \mathbf{\Sigma}_{k-1},\ 1 \le k \le n$. By Lemmas 2.3 and 2.4, to prove the lemma, it suffices to check that $\mathrm{tr}(\mathbf{\Sigma}_k) = k-1$ and $\mathrm{tr}(\mathbf{I} - \mathbf{\Sigma}_k) = n-k+1$ for $1 \le k \le n$.

In fact, $\mathrm{tr}(\mathbf{\Sigma}_k) = \mathrm{tr}(\sum_{i=1}^{k-1} \mathbf{e}_i\mathbf{e}_i^*) = \sum_{i=1}^{k-1} \mathrm{tr}(\mathbf{e}_i\mathbf{e}_i^*) = k-1$ by (3.3) in the Appendix. Further, $\mathrm{tr}(\mathbf{I} - \mathbf{\Sigma}_k) = \mathrm{tr}(\mathbf{I}) - \mathrm{tr}(\mathbf{\Sigma}_k) = n-k+1$. ■

Lemma 2.6 *Let $\mathbf{w}_k$ be as in (2.3). Let also $\beta = 1,\ 2$ or 4 according to $F = \mathbb{R}, \mathbb{C}$ or $\mathbb{H}$. Then $\|\mathbf{w}_k\|^2 \sim \beta^{-1} \cdot \chi^2(\beta(n-k+1))$ for any $1 \le k \le n$.*

Proof. Recall $\mathbf{w}_k = (\mathbf{I} - \mathbf{\Sigma}_{k-1})\mathbf{y}_k$. As in the proof of (ii) in Lemma 2.5, $(\mathbf{I} - \mathbf{\Sigma}_{k-1})^* = \mathbf{I} - \mathbf{\Sigma}_{k-1}$, $(\mathbf{I} - \mathbf{\Sigma}_{k-1})^2 = (\mathbf{I} - \mathbf{\Sigma}_{k-1})$ and $\mathrm{tr}(\mathbf{I} - \mathbf{\Sigma}_{k-1}) = n-k+1$ for any $1 \le k \le n$. So by Lemma 3.3, $\mathbf{I} - \mathbf{\Sigma}_{k-1} = \mathbf{O}\,\mathrm{diag}\,(\mathbf{I}_{n-k+1}, \mathbf{0})\mathbf{O}^*$ for some $\mathbf{O} \in O(n),\ U(n)$ or $Sp(n)$ according to $F = \mathbb{R},\ \mathbb{C}$ or $\mathbb{H}$.

By the invariance of normal distributions, we know $\mathbf{O}^*\mathbf{y}_k \sim \mathbf{y}_k$. Therefore, $\|\mathbf{w}_k\|^2 = \|\mathrm{diag}\,(\mathbf{I}_{n-k+1}, \mathbf{0})\mathbf{O}^*\mathbf{y}_k\|^2 \stackrel{d}{=} \|\mathrm{diag}\,(\mathbf{I}_{n-k+1}, \mathbf{0})\mathbf{y}_k\|^2 \stackrel{d}{=} \sum_{i=1}^{n-k+1} |\eta_i|^2 \stackrel{d}{=} (1/\beta)\chi^2(\beta(n-k+1))$, where $\beta = 1, 2$ and 4 and $\eta_1, \eta_2, \cdots$ are i.i.d. r.v.'s with distribution $FN(0,1)$. ■

With the above characterization of the distributions of various random variables, we now are ready to derive some probability inequalities.

Lemma 2.7 *(Lemma 3.3 in [13]) The following holds:*
(i) $x - 1 - \log x \ge (x-1)^2/2$ for $x \in (0,1]$;
(ii) $2x - \log(1+2x) \ge x^2$ for $x \in (0, 1/4]$;
(iii) $(1-x)^{-2} \ge 1 + 2x$ and $(1+x)^{-2} \le 1 - x$ for $x \in (0, 1/4]$.

Lemma 2.8 *(Lemma 3.2 in [13]) 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)* $E\exp(\theta\xi^2) = (1-2\theta)^{-1/2}$ *for* $\theta < 1/2$;
(ii)

$$I(x) = \begin{cases} (x - 1 - \log x)/2 & \text{if } x > 0; \\ +\infty & \text{otherwise.} \end{cases}$$

(iii) Define $J(x) = I(x)/x$ *for* $x > 0$. *Then both* $I(x)$ *and* $J(x)$ *are increasing on* $(1,\infty)$ *and decreasing on* $(0,1]$.

Lemma 2.9 *For* $F = \mathbb{R}\,(\beta = 1)$, $\mathbb{C}\,(\beta = 2)$ *or* $\mathbb{H}\,(\beta = 4)$, *let* $Z_\beta \sim FN(0,1)$. *Then*

$$P(|Z_\beta| \geq x) \leq \varphi_\beta(x)\cdot e^{-\beta x^2/2}$$

for any $x > 0$, *where* $\varphi_\beta(x) = 1/x$, 1 *or* $2x^2+1$ *for* $\beta = 1, 2$ *or* 4.

Proof. First, in view of (a), (b) and (c) in the Introduction,

$$P(|FN(0,1)| \geq x) = P\left(\frac{\sum_{i=1}^{\beta}\xi_i^2}{\beta} \geq x^2\right) \tag{2.6}$$

for $1 \leq k \leq n$, where ξ_1, ξ_2, ξ_3, ξ_4 are i.i.d. $N(0,1)$-distributed random variables. If $\beta = 1$, then, by the well known inequality, we have

$$P(|Z_1| \geq x) \leq \frac{2}{\sqrt{2\pi}\,x}e^{-x^2/2} \leq \frac{1}{x}e^{-\beta x^2/2}$$

for any $x > 0$. If $\beta = 2$, since $(\xi_1^2 + \xi_2^2)/2 \sim Exp(1)$, we obtain

$$P(|Z_2| \geq x) = e^{-\beta x^2/2}$$

for any $x > 0$. If $\beta = 4$, we know that $\xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2 \sim \chi^2(4)$, which has probability density function $xe^{-x/2}I(x \geq 0)/4$. Thus, use $((y+1)e^{-y})' = -ye^{-y}$ to obtain that

$$P(|Z_4| \geq x) = \int_{4x^2}^{\infty}\frac{t}{4}e^{-t/2}\,dt = \int_{2x^2}^{\infty}ye^{-y}\,dy = (2x^2+1)e^{-\beta x^2/2}$$

for any $x > 0$. The desired inequality follows by combining the three cases together. ■

Lemma 2.10 *Let* $\{\xi_i;\, i \geq 1\}$ *be a sequence of i.i.d.* $N(0,1)$*-distributed r.v.'s. For* $n \geq m \geq 3$, *set* $W = (\sum_{i=1}^m \xi_i^2/\sum_{i=1}^n \xi_i^2)^{1/2}$. *Then* $E(W^{-2}) \leq n$.

Proof. The conclusion obviously holds for $m = n$. Now, assume $3 \leq m < n$. Write $W^{-2} = 1 + \sum_{i=m+1}^n \xi_i^2/\sum_{i=1}^m \xi_i^2$. Then, by independence,

$$\begin{aligned} E(W^{-2}) &= 1 + (E\sum_{i=m+1}^{n}\xi_i^2)\cdot E\frac{1}{\sum_{i=1}^m \xi_i^2} \\ &\leq 1 + (n-m)E\frac{1}{\xi_1^2+\xi_2^2+\xi_3^2}. \end{aligned}$$

Now,

$$\begin{aligned} E\frac{1}{\xi_1^2+\xi_2^2+\xi_3^2} &= (\sqrt{2\pi})^{-3}\int_{\mathbb{R}^3}\frac{e^{-(x^2+y^2+z^2)/2}}{x^2+y^2+z^2}\,dx\,dy\,dz \\ &= (\sqrt{2\pi})^{-3}\int_0^\infty\int_0^\pi\int_0^{2\pi}\frac{1}{r^2}e^{-r^2/2}\cdot(r^2\sin\theta_1)\,dr\,d\theta_1\,d\theta_2 \\ &= (\sqrt{2\pi})^{-3}\cdot\frac{1}{2}\int_{-\infty}^\infty e^{-r^2/2}\,dr\cdot\int_0^\pi \sin\theta_1\,d\theta_1\cdot(2\pi)=1 \end{aligned}$$

where the spherical coordinate transform $x = r\sin\theta_1\cos\theta_2,\ y = r\sin\theta_1\sin\theta_2,\ z = r\cos\theta_1$ for $(r,\theta_1,\theta_2)\in[0,\infty)\times[0,\pi]\times[0,2\pi]$ is used above. Combining the above two assertions, the conclusion follows. ■

Lemma 2.11 *Let $\{\xi_i;\, i\ge 1\}$ be a sequence of i.i.d. $N(0,1)$-distributed r.v.'s. For $n>m\ge 1$, set $W=(\sum_{i=1}^m\xi_i^2/\sum_{i=1}^n\xi_i^2)^{1/2}$. Then*

$$Ee^{-a/W^2}\le 3\left(1+\frac{2a}{3(m+\sqrt{2an}\,)}\right)^{-n/2}$$

for any $a>0$.

Proof. Write $W^{-2}=1+(\sum_{k=m+1}^n\xi_k^2)(\sum_{k=1}^m\xi_k^2)^{-1}$. Since $\{\xi_{m+1},\xi_{m+2},\cdots,\xi_n\}$ and $\sum_{k=1}^m\xi_k^2$ are independent, we have $Ee^{-aW^{-2}}=e^{-a}E(M^{n-m})$, where

$$M=E\left\{\exp\left(-\frac{a\xi_n^2}{\sum_{k=1}^m\xi_k^2}\right)\mid \xi_1,\xi_2,\cdots,\xi_m\right\}.$$

By (i) of Lemma 2.8, $E\exp(-\beta\xi_n^2)=(1+2\beta)^{-1/2}$ for $\beta>-1/2$. It follows that $M=(1+2a(\sum_{k=1}^m\xi_k^2)^{-1})^{-1/2}$. Hence,

$$Ee^{-a/W^2}=e^{-a}E\left\{\left(1+\frac{2a}{\sum_{k=1}^m\xi_k^2}\right)^{-(n-m)/2}\right\}. \tag{2.7}$$

By (ii) of Lemma 2.8 and the Chernoff bound, see, e.g., Remark (c) on p.27 from [3], we get

$$P\left(\frac{1}{m}\sum_{k=1}^m\xi_k^2\in A\right)\le 2e^{-mI(A)} \tag{2.8}$$

where $I(A)=\inf_{x\in A}I(x)$ and $I(x)=\sup_{\theta\in\mathbb{R}}\{\theta x-\log Ee^{\theta\xi_1^2}\}=(x-1-\log x)/2$ for $x>0$; $I(x)=+\infty$ for $x\le 0$. Thus

$$P\left(\sum_{k=1}^m\xi_k^2\ge x\right)\le 2e^{-mI(x/m)},\ x\ge m, \tag{2.9}$$

since $I(x)$ is increasing on $[1,\infty)$. By (iii) of Lemma 2.8, $J(x):=I(x)/x$ is increasing on $(1,+\infty)$ and decreasing on $(0,1]$. Set $x_0=2m+4\sqrt{a(n-m)}$. Then

$$P\left(\sum_{k=1}^m\xi_k^2\ge x_0\right)\le 2e^{-mI(x_0/m)}=2e^{-x_0J(x_0/m)}\le 2e^{-x_0J(2)}\le 2e^{-x_0/16}$$

since $x_0/m > 2$ and $J(2) = I(2)/2 = (1 - \log 2)/4 > 1/16$. Considering $\{\sum_{k=1}^m \xi_k^2 \le x_0\}$ and its complement event, and noticing $(1 + 2a/(\sum_{k=1}^m \xi_k^2))^{-(n-m)/2} \le 1$, we have from above that

$$E\left\{\left(1 + \frac{2a}{\sum_{k=1}^m \xi_k^2}\right)^{-(n-m)/2}\right\} \le \left(1 + \frac{2a}{x_0}\right)^{-(n-m)/2} + 2e^{-x_0/16}.$$

Easily, $1 + x \le e^x$ for any $x \in \mathbb{R}$, then $e^{-x_0/16} \le (1 + (2a/x_0))^{-x_0^2/(32a)}$. Also, $x_0^2/(32a) > (n-m)/2$. The above implies that

$$E\left\{\left(1 + \frac{2a}{\sum_{k=1}^m \xi_k^2}\right)^{-(n-m)/2}\right\} \le 3\left(1 + \frac{2a}{x_0}\right)^{-(n-m)/2} \le 3e^a\left(1 + \frac{2a}{3(m + \sqrt{2an})}\right)^{-n/2},$$

where the facts $(1 + 2ax_0^{-1})^{m/2} \le \exp(ax_0^{-1}m) \le e^a$ and $x_0 < 3(m + \sqrt{2an})$ are used in the last step. This and (2.7) conclude the inequality stated in the lemma. ■

Lemma 2.12 *Let $\mathbf{y}_k, \mathbf{e}_k, \mathbf{w}_k$ and $\mathbf{\Sigma}_k$ be as in (2.1)-(2.3), and let $\beta = 1, 2$ or 4 according to $F = \mathbb{R}, \mathbb{C}$ or $\mathbb{H}$. Then*

$$P(\max_{1\le i\le m} \|\mathbf{\Sigma}_{i-1}\mathbf{y}_i\| \ge t) \le 3mn \cdot \chi_\beta(n,t) \cdot \left(1 + \frac{t^2}{3(m + t\sqrt{n}\,)}\right)^{-\beta n/2}$$

for any $1 \le m < n$ and $t > 0$, where

$$\chi_\beta(n,t) = \begin{cases} t^{-1} & \text{if } \beta = 1; \\ 1 & \text{if } \beta = 2; \\ 8nt^2 + 1 & \text{if } \beta = 4. \end{cases} \tag{2.10}$$

Proof. Since $\mathbf{\Sigma}_0 = \mathbf{0}$, without loss of generality, we now assume $1 < m < n$. By (i) of Lemma 2.5,

$$\begin{aligned} P(\max_{1\le i\le m} \|\mathbf{\Sigma}_{i-1}\mathbf{y}_i\| \ge t) &\le mn \cdot \max_{2\le i\le m} P(|\eta| \cdot W_i \ge t) \\ &\le mn \cdot P(|\eta| \cdot W_{m+1} \ge t) \end{aligned} \tag{2.11}$$

where $\eta \sim FN(0,1)$ and $W_i = \left(\sum_{k=1}^{\beta(i-1)} \xi_k^2 / \sum_{k=1}^{\beta n} \xi_k^2\right)^{1/2} \le 1$ are independent, $2 \le i \le n+1$. By Lemma 2.9, $P(|FN(0,1)| \ge x) \le \varphi_\beta(x) \cdot e^{-\beta x^2/2}$ where $\varphi_\beta(x) = 1/x$, 1 or $2x^2 + 1$ for $\beta = 1, 2$ or 4. Thus,

$$P(|\eta| \cdot W_{m+1} \ge t) \le E\left\{\varphi_\beta(tW_{m+1}^{-1}) \cdot \exp\left(-\frac{\beta t^2}{2} W_{m+1}^{-2}\right)\right\}. \tag{2.12}$$

Therefore,

(i) If $\beta = 1$, then $\varphi_\beta(tW_{m+1}^{-1}) \le 1/t$. Taking $a = \beta t^2/2$, replacing m by βm, and n by βn in Lemma 2.11, we obtain

$$P(|\eta| \cdot W_{m+1} \ge t) \le \frac{3}{t}\left(1 + \frac{\beta t^2}{3(\beta m + t\beta\sqrt{n}\,)}\right)^{-\beta n/2} \le \frac{3}{t}\left(1 + \frac{t^2}{3(m + t\sqrt{n}\,)}\right)^{-\beta n/2}.$$

(ii) If $\beta=2$, then $\varphi_\beta(tW_{m+1}^{-1})=1$. By (2.12) and Lemma 2.11, choose $a=\beta t^2/2$, replace m with βm, and n with βn in Lemma 2.11 to get

$$P(|\eta|\cdot W_{m+1}\geq t)\leq E\exp\left\{-\frac{\beta t^2}{2}W_{m+1}^{-2}\right\}\leq 3\left(1+\frac{t^2}{3(m+t\sqrt{n}\,)}\right)^{-\beta n/2}.$$

(iii) If $\beta=4$, then $\varphi_\beta(x)=2x^2+1$. Notice $\varphi_\beta(x)$ is increasing and $\psi(x):=e^{-\beta x^2/2}$ is decreasing over $[0,\infty)$, respectively. By the well known inequality, $E\left(\varphi(Z)\psi(Z)\right)\leq E\varphi(Z)\cdot E\psi(Z)$ for any random variable $Z\geq 0$. Since $1<m<n$, we know $\beta m\geq 8$. By (2.12) and Lemmas 2.10 and 2.11,

$$\begin{aligned}P(|\eta|\cdot W_{m+1}\geq t) &\leq E\left\{\varphi_\beta(tW_{m+1}^{-1})\right\}\cdot E\exp\left(-\frac{\beta t^2}{2}W_{m+1}^{-2}\right)\\ &\leq 3(8nt^2+1)\left(1+\frac{t^2}{3(m+t\sqrt{n}\,)}\right)^{-\beta n/2}.\end{aligned}$$

Collecting (2.11), (2.12) and the inequalities in (i), (ii) and (iii), we eventually conclude that

$$P(\max_{1\leq i\leq m}\|\mathbf{\Sigma}_{i-1}\mathbf{y}_i\|\geq t)\leq 3mn\cdot\chi_\beta(n,t)\cdot\left(1+\frac{t^2}{3(m+t\sqrt{n}\,)}\right)^{-\beta n/2}$$

for any $t>0$, where

$$\chi_\beta(n,t)=\begin{cases}t^{-1} & \text{if } \beta=1;\\ 1 & \text{if } \beta=2;\\ 8nt^2+1 & \text{if } \beta=4.\end{cases}\qquad\blacksquare \tag{2.13}$$

LEMMA 2.13 *Let $\mathbf{w}_k$ be as in (2.3), and $a_k=\sqrt{n}/\|\mathbf{w}_k\|$ be as in Lemma 2.2. Let also $\beta=1,2$ or 4 according to $F=\mathbb{R},\mathbb{C}$ or $\mathbb{H}$. Then*

$$P(\max_{1\leq i\leq m}|a_i-1|\geq r)\leq 4me^{-\beta nr^2/16}$$

for all $r\in(0,1/4)$ and $m\leq nr/2$.

Proof. First, $P(\max_{1\leq i\leq m}|a_i-1|\geq r)\leq m\cdot\max_{1\leq i\leq m}P(|a_i-1|\geq r)$. By (iii) of Lemma 2.7,

$$\begin{aligned}P(|a_i-1|\geq r) &\leq P\left(\frac{\sqrt{n}}{\|\mathbf{w}_i\|}\leq 1-r\right)+P\left(\frac{\sqrt{n}}{\|\mathbf{w}_i\|}\geq 1+r\right)\\ &\leq P\left(\frac{\|\mathbf{w}_i\|^2}{n}\geq(1-r)^{-2}\right)+P\left(\frac{\|\mathbf{w}_i\|^2}{n}\leq(1+r)^{-2}\right)\\ &\leq P\left(\frac{\|\mathbf{w}_i\|^2}{n}\geq 1+2r\right)+P\left(\frac{\|\mathbf{w}_i\|^2}{n}\leq 1-r\right)\end{aligned}\tag{2.14}$$

for any $1\leq i<n$ since $r\in(0,1/4)$. From Lemma 2.6, we know $\|\mathbf{w}_i\|^2\sim\sum_{j=1}^{\beta(n-i+1)}\xi_j^2/\beta\leq\sum_{j=1}^{\beta n}\xi_j^2/\beta$ for all $1\leq i\leq m$, where $\{\xi_j;\,j\geq 1\}$ are i.i.d. $N(0,1)$-distributed r.v.'s. Therefore,

$$\max_{1\leq i\leq m}P\left(\frac{\|\mathbf{w}_i\|^2}{n}\geq 1+2r\right)\leq P\left(\frac{\sum_{j=1}^{\beta n}\xi_j^2}{\beta n}\geq 1+2r\right)\leq 2e^{-\beta n\lambda}$$

for $r \in (0, 1/4)$ where $\lambda := \inf_{x \ge 1+2r} I(x)$ and $I(x)$ is given in (ii) of Lemma 2.8. Since $I(x)$ is increasing on $[1, \infty)$, $\lambda = I(1+2r) = (2r - \log(1+2r))/2 \ge r^2/2$ for $r \in (0, 1/4)$ by (ii) of Lemma 2.7. Thus

$$\max_{1\le i\le n} P\left(\frac{\|\mathbf{w}_i\|^2}{n} \ge 1+2r\right) \le 2e^{-\beta n r^2/2} \tag{2.15}$$

for any $r \in (0, 1/4)$. Now we estimate the last probability in (2.14). Since $\|\mathbf{w}_i\|^2/n \sim \sum_{j=1}^{\beta(n-i+1)} \xi_j^2/(n\beta) \ge \sum_{j=1}^{\beta(n-m)} \xi_j^2/(n\beta)$ for any $1 \le i \le m$, we have

$$\begin{aligned}\max_{1\le i\le m} P\left(\frac{\|\mathbf{w}_i\|^2}{n} \le 1-r\right) \le P\left(\frac{\sum_{j=1}^{\beta(n-m)} \xi_j^2}{\beta(n-m)} \le b\right) &\le 2e^{-\beta(n-m)I(B)} \\ &= 2e^{-\beta(n-m)I(b)} \end{aligned}\tag{2.16}$$

by (2.8), where $b := n(1-r)/(n-m)$, $B = (-\infty, b]$ and $I(x)$ is as in Lemma 2.8. Here the fact that $I(x)$ is decreasing on $(0, 1)$, and that $0 < b < 1$ since $m \le nr/2$ are used in the last step. From (i) of Lemma 2.7,

$$(n-m)I(b) \ge (n-m)\cdot\frac{(1-b)^2}{4} \ge \frac{(nr-m)^2}{4(n-m)} \ge \frac{nr^2}{16} \tag{2.17}$$

for all $1 \le m \le nr/2$. The inequalities in (2.16) and (2.17) imply that

$$\max_{1\le i\le m} P\left(\frac{\|\mathbf{w}_i\|^2}{n} \le 1-r\right) \le 2e^{-\beta n r^2/16}. \tag{2.18}$$

The desired conclusion follows by combining (2.14), (2.15) and (2.18). ■

Proof of Theorem 1. By Lemma 2.2,

$$\begin{aligned} & P\left(\max_{1\le k\le m} \|\sqrt{n}\mathbf{e}_k - \mathbf{y}_k\| \ge rs+t\right) \\ \le\ & P\left(\max_{1\le k\le m} |a_k - 1| \ge r\right) + P\left(\max_{1\le k\le m} \|\mathbf{w}_k\| \ge s\right) + P\left(\max_{1\le k\le m} \|\mathbf{\Sigma}_{k-1}\mathbf{y}_k\| \ge t\right)\end{aligned}$$

for any $r > 0$, $s > 0$ and $t > 0$. Recall (ii) of Lemma 2.5, $\sum_{i=1}^{\beta(n-k+1)} \xi_i^2/\sum_{i=1}^{\beta n} \xi_i^2 \le 1$ for all $1 \le k \le n$. It follows from Lemma 2.9 that

$$\begin{aligned} P\left(\max_{1\le k\le m} \|\mathbf{w}_k\| \ge s\right) &\le m\cdot\max_{1\le k\le m} P(\|\mathbf{w}_k\| \ge s) \\ &\le mn\cdot P(|FN(0,1)| \ge s) \le mn\cdot\varphi_\beta(s)\cdot e^{-\beta s^2/2} \end{aligned}\tag{2.19}$$

for $1 \le k \le n$, where $\varphi_\beta(s)$ is as in Lemma 2.9. Now, (2.19) together with Lemmas 2.12 and 2.13 yields

$$\begin{aligned} & P\left(\max_{1\le k\le m} \|\sqrt{n}\mathbf{e}_k - \mathbf{y}_k\| \ge rs+t\right) \\ \le\ & 4me^{-\beta n r^2/16} + mn\cdot\varphi_\beta(s)\cdot e^{-\beta s^2/2} + 3mn\cdot\chi_\beta(n,t)\left(1+\frac{t^2}{3(m+t\sqrt{n})}\right)^{-\beta n/2}\end{aligned}$$

for any $r \in (0, 1/4),\ s > 0,\ t > 0$ and $1 \le m \le nr/2$, where $\chi_\beta(n,t)$ is as in Lemma 2.12. The result follows since $\varphi_\beta(s) \le \chi_\beta(n,s)$ for all $s > 0,\ n \ge 1$ and $\beta = 1, 2, 4$. ■

Proof of Corollary 1.1. Choose $r = 1/\log n,\ s = (\log n)^{3/4},\ t = t,\ m = m_n' = [\delta n/\log n]$ for some $\delta < \min\{1/4, t^2/100\}$ in Theorem 1. It is easy to see that $t^2/(3(m+\sqrt{n})) \ge t^2(\log n)/(4n\delta)$ and $1/s \le 1$ for all $\beta = 1,\ 2,\ 4$ as n large enough. Noting that

$$\chi_\beta(n,t) \le \chi_\beta(n, (\log n)^{3/4}) \le n^3$$

for $\beta = 1, 2, 4$ as n is large enough, we get

$$\begin{aligned} P(\epsilon_n(m_n) \ge 3t) &\le P(\epsilon_n(m_n') \ge 3t) \\ &\le 4ne^{-n/(4\log n)^2} + n^2 \cdot \chi_\beta(n,\ (\log n)^{3/4}) \cdot e^{-(\log n)^{3/2}/2} \\ &\qquad + 3n^2 \cdot \chi_\beta(n,\ t)\left(1 + \frac{t^2}{4\delta}\frac{\log n}{n}\right)^{-n/2} \to 0 \end{aligned}$$

as $n \to \infty$ by the choice of δ. ■

3 Appendix

Recall that $\mathbb{H}$ is the set of real quaternions stated in the Introduction. For $a = a_1+a_2i+a_3j+a_4k \in \mathbb{H}$, its conjugate $a^* = a_1 - a_2i - a_3j - a_4k$, and its norm $|a| = (a_1^2 + a_2^2 + a_3^2 + a_4^2)^{1/2}$. Let $\mathbf{A} = (a_{pq})$ be a matrix with entries $a_{pq} \in \mathbb{H}$ for all p and q, the conjugate of $\mathbf{A}$ is $\mathbf{A}^* = (a_{qp}^*)$. We say a square matrix $\mathbf{A}$ is self-dual if $\mathbf{A}^* = \mathbf{A}$.

The following statements and lemmas about matrices of quaternion entries, as in the linear algebras, look quite familiar. However, since the multiplication operation of the elements in $\mathbb{H}$ is not commutative, the statements need to be verified. It seems hard to find their proofs in the literature, we collect and prove them next.

Lemma 3.1 *Let $a, b \in \mathbb{H}$, and $\mathbf{A}$ and $\mathbf{B}$ are matrices of quaternion entries. Then (i) $(ab)^* = b^*a^*$; (ii) $|ab| = |a| \cdot |b|$; (iii) $(ab)^{-1} = b^{-1}a^{-1}$ if $a \ne 0$ and $b \ne 0$; (iv) $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$.*

Proof. (i) Let $a = \alpha_1+\alpha_2i+\alpha_3j+\alpha_4k$ and $b = \beta_1+\beta_2i+\beta_3j+\beta_4k$, where $\alpha_i,\ \beta_i \in \mathbb{R}$ for $i = 1,\ 2,\ 3,\ 4$. Then, by the identities that $i^2 = j^2 = k^2 = -1,\ ij = k,\ jk = i,\ ki = j,\ ij = -ji,\ ik = -ki$ and $jk = -kj$, we have

$$\begin{aligned} ab = (\alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3 - \alpha_4\beta_4) &+ (\alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_4 - \alpha_4\beta_3)i \\ &+ (\alpha_1\beta_3 + \alpha_3\beta_1 + \alpha_4\beta_2 - \alpha_2\beta_4)j \\ &+ (\alpha_1\beta_4 + \alpha_4\beta_1 + \alpha_2\beta_3 - \alpha_3\beta_2)k. \end{aligned} \tag{3.1}$$

Thus,

$$\begin{aligned}(ab)^* = (\alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3 - \alpha_4\beta_4) &- (\alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_4 - \alpha_4\beta_3)i \\ &- (\alpha_1\beta_3 + \alpha_3\beta_1 + \alpha_4\beta_2 - \alpha_2\beta_4)j \\ &- (\alpha_1\beta_4 + \alpha_4\beta_1 + \alpha_2\beta_3 - \alpha_3\beta_2)k.\end{aligned}$$

Notice that $a^* = \alpha_1 - \alpha_2 i - \alpha_3 j - \alpha_4 k$ and $b^* = \beta_1 - \beta_2 i - \beta_3 j - \beta_4 k$. Applying formula (3.1) to b^*a^*, we see $(ab)^* = b^*a^*$.

(ii) Note that $(|a|\cdot|b|)^2 = (\sum_{p=1}^4 \alpha_p^2)\cdot\sum_{p=1}^4 \beta_p^2$. It is trivial to verify that $|ab|^2 = (|a|\cdot|b|)^2$ by (3.1).

(iii) Since $cc^* = c^*c = |c|^2$ for any $0 \neq c \in \mathbb{H}$, then $c^{-1} = c^*/|c|^2$. It follows from (i) and (ii) that

$$(ab)^{-1} = \frac{(ab)^*}{|ab|^2} = \frac{b^*a^*}{|a|^2|b|^2} = \frac{b^*}{|b|^2}\cdot\frac{a^*}{|a|^2} = b^{-1}a^{-1}.$$

(iv) Let $\mathbf{A} = (a_{pq})$ and $\mathbf{B} = (b_{pq})$. For matrix $\mathbf{C} = (c_{pq})$, we use $\mathbf{C}_{pq}$ to denote the (p,q)-entry of $\mathbf{C}$. Recall $\mathbf{C}^* = (c_{qp}^*)$. First, obviously, $(\mathbf{B}^*\mathbf{A}^*)_{pq} = \sum_k b_{kp}^* a_{qk}^*$. It is easy to see that

$$((\mathbf{AB})^*)_{pq} = ((\mathbf{AB})_{qp})^* = (\sum_k a_{qk}b_{kp})^* = \sum_k b_{kp}^* a_{qk}^*,$$

where (i) is used in the last step. Thus, $((\mathbf{AB})^*)_{pq} = (\mathbf{B}^*\mathbf{A}^*)_{pq}$ for any p and q. That is, $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$. ■

Let $\mathbf{A} = (a_{ij})_{n\times n}$ with $a_{ij} \in \mathbb{H}$ for all $1 \le i, j \le n$. Define $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$. Obviously, $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$ for any $\mathbf{A}$ and $\mathbf{B}$. If $\mathbf{A}^* = \mathbf{A}$, then $\{a_{ii} : 1 \le i \le n\}$ are real numbers, so $\mathrm{tr}(\mathbf{A})$ is a real number.

Lemma 3.2 *Let $\mathbf{A}^* = \mathbf{A}$. Then $tr(\mathbf{U}^*\mathbf{A}\mathbf{U}) = tr(\mathbf{A})$ for any $\mathbf{U} \in Sp(n)$. In particular, $tr(\mathbf{A}) = \sum_{i=1}^n \lambda_i$, where $\{\lambda_i; 1 \le i \le n\}$ are eigenvalues of $\mathbf{A}$, that is, there exists $\mathbf{x}_i \in \mathbb{H}^n$ such that $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$ for all $i = 1, 2, \cdots, n$.*

Proof. By the spectral theorem for hermitian quaternion-real matrices (see, e.g., Theorem 2 on p.145 from [8] or Theorem 3.2.1 on p.59 from [15]), there exists a matrix $\mathbf{V} \in Sp(n)$ such that $\mathbf{A} = \mathbf{V}^*\,\mathrm{diag}(\lambda_1,\cdots,\lambda_n)\,\mathbf{V}$, where λ_i, $1 \le i \le n$, are eigenvalues of $\mathbf{A}$, and they all are real numbers. So it suffices to prove the first conclusion. We claim

$$\mathrm{Re}(\mathrm{tr}(\mathbf{ST})) = \mathrm{Re}(\mathrm{tr}(\mathbf{TS})) \tag{3.2}$$

for any two multiplicable (not necessarily square) matrices $\mathbf{S}$ and $\mathbf{T}$, where $\mathrm{Re}(a+bi+cj+dk) = a$ for any a, b, c, $d \in \mathbb{R}$. If this holds, then

$$\mathrm{tr}(\mathbf{U}^*\mathbf{A}\mathbf{U}) = \mathrm{Re}(\mathrm{tr}(\mathbf{U}^*\mathbf{A}\mathbf{U})) = \mathrm{Re}(\mathrm{tr}(\mathbf{A}\mathbf{U}\mathbf{U}^*)) = \mathrm{Re}(\mathrm{tr}(\mathbf{A})) = \mathrm{tr}(\mathbf{A}).$$

We now verify (3.2). Let $\mathbf{S} = (s_{ij})$ and $\mathbf{T} = (t_{ij})$. Recall (i) in the proof of Lemma 3.1 and (3.1), we have $\mathrm{Re}(a+b) = \mathrm{Re}(a) + \mathrm{Re}(b)$ and $\mathrm{Re}(ab) = \mathrm{Re}(ba)$ for any $a, b \in \mathbb{H}$. Therefore,

$$\mathrm{Re}(\mathrm{tr}(\mathbf{ST})) = \mathrm{Re}(\sum_{i,j} s_{ij}t_{ji}) = \mathrm{Re}(\sum_{i,j} t_{ji}s_{ij}) = \mathrm{Re}(\mathrm{tr}(\mathbf{TS})). \qquad ■$$

Let $\mathbf{e} \in \mathbb{H}^n$ be a unit vector, since $(\mathbf{e}\mathbf{e}^*)^* = \mathbf{e}\mathbf{e}^*$, by (3.2),

$$\mathrm{tr}(\mathbf{e}\mathbf{e}^*) = \mathrm{Re}(\mathrm{tr}(\mathbf{e}\mathbf{e}^*)) = \mathrm{Re}(\mathrm{tr}(\mathbf{e}^*\mathbf{e})) = 1. \tag{3.3}$$

Lemma 3.3 *Let the entries of an $n \times n$ matrix $\mathbf{A}$ be in $\mathbb{H}$ such that $\mathbf{A}^* = \mathbf{A}$ and $\mathbf{A}^2 = \mathbf{A}$. Then there exist an integer $0 \le r \le n$ and a matrix $\mathbf{O} \in Sp(n)$ such that*

$$\mathbf{A} = \mathbf{O}^* \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{O}.$$

Proof. By the spectral theorem for hermitian quaternion-real matrices (see, e.g., Theorem 2 on p.145 from [8] or Theorem 3.2.1 on p.59 from [15]), there exists a matrix $\mathbf{B} \in Sp(n)$ such that $\mathbf{A} = \mathbf{B}^* \,\mathrm{diag}(\lambda_1, \cdots, \lambda_n)\, \mathbf{B}$, where λ_i, $1 \le i \le n$, are real numbers. Since $\mathbf{A}^2 = \mathbf{A}$, we have $\mathrm{diag}(\lambda_1^2, \cdots, \lambda_n^2) = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. This says that $\lambda_i = 0$ or 1 for any $i = 1, 2, \cdots, n$. Write $\mathrm{diag}(\lambda_1, \cdots, \lambda_n) = \mathbf{P}^* \,\mathrm{diag}(1, 1, \cdots, 1, 0, \cdots, 0)\, \mathbf{P}$ for some (real) permutation matrix $\mathbf{P}$. The conclusion follows by taking $\mathbf{O} = \mathbf{P}\mathbf{B} \in Sp(n)$. ■

Acknowledgements. The author thanks Xue Ding and Danning Li for helpful discussions on the proofs of the main result in this paper.

References

[1] Borel, E. (1906). Introduction géometrique à quelques théories physiques, Gauthier-Villars, Paris.

[2] Collins, B. (2003). Intégrales Matricielles et Probabilitiés Non-commutatives. Thèse de Doctorat of Université Paris 6.

[3] Dembo, A. and Zeitouni, O. (1998). Large Deviations Techniques and Applications. Springer, 2nd edition.

[4] D'Aristotle, A., Diaconis, P. and Newman, C. (2003). Brownian Motion and the Classical Groups, Probability, Statisitca and their applications: Papers in Honor of Rabii Bhattacharaya. Edited by K. Athreya, et al. 97-116. Beechwood, OH: Institute of Mathematical Statistics (2003).

[5] Diaconis, P., Eaton, M. and Lauritzen, L. (1992). Finite deFinetti theorem in linear models and multivariate analysis. Scand. J. Statist. 19(4), 289-315.

[6] Diaconis, P. and Freedman, D. (1987). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, Vol. 23, 397-423.

[7] Donoho, D. and Huo, X. (2001). Uncertainty principles and ideal atomic decomposition. IEEE Trans. Information Theory, Vol. 47(7), 289-315.

[8] Dyson, E. (1962). Statistical theory of the energy levels of complex systems I. Journal of Math. Phys. 3(1) 140-156.

[9] Gallardo, L. (1983). Au sujet du contenu probabiliste d'un lemma d'Henri Poincaré. Annals de l'université de clemont, Vol. 69, 192-197.

[10] Jiang, T. (2009). A variance formula for quantum conductance. Physics Letters A 373(25), 2117-2121.

[11] Jiang, T. (2009). Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles. Probability Theory and Related Fields 144(1), 221-246.

[12] Jiang, T. (2006). How many entries of a typical orthogonal matrix can be approximated by independent normals? Ann. Probab. 34(4), 1497-1529.

[13] Jiang, T. (2005). Maxima of Entries of Haar Distributed Matrices. Probability Theory and Related Fields, 131, 121-144.

[14] Knapp, A. W. (2002). Lie Groups: Beyond An Introduction. Birkhauser Boston; 2nd ed.

[15] Mehta, M. L. (1991). Random Matrices, 2nd Ed., Academic Press, Boston.

[16] Mezzadri, F. (2007). How to generate random matrices from the classical compact groups. Notices to the AMS, Volume 54(5), 592-604.

[17] Procesi, C. (2005). Lie Groups: An Approach through Invariants and Representations. Springer.

[18] Stam, A.J. (1982). Limit theorems for uniform distributions on high dimensional Euclidean spaces. J. Appl. Prob., Vol, 19, 221-228.

[19] Yor, M. (1985). Inégalitiés de martingales continus arrêtès à un temps quelconques I. Springer Lecture Notes in Math., No. 1118.