A comparison principle for functions of a uniformly random subspace

Joel A. Tropp

Abstract　　This note demonstrates that it is possible to bound the expectation of an arbitrary norm of a random matrix drawn from the Stiefel manifold in terms of the expected norm of a standard Gaussian matrix with the same dimensions. A related comparison holds for any convex function of a random matrix drawn from the Stiefel manifold. For certain norms, a reversed inequality is also valid.

1 Main result

Many problems in high-dimensional geometry concern the properties of a random k-dimensional subspace of the Euclidean space $\mathbb{R}^n$. For instance, the Johnson–Lindenstrauss Lemma [7] shows that, typically, the metric geometry of a collection of N points is preserved when we project the points onto a random subspace with dimension O(log N). Another famous example is Dvoretsky's Theorem [1,3,9], which states that, typically, the intersection between the unit ball of a Banach space with dimension N and a random subspace with dimension O(log N) is comparable with a Euclidean ball.

In geometric problems, it is often convenient to work with matrices rather than subspaces. Therefore, we introduce the Stiefel manifold,

$$\mathbb{V}_k^n := \{ \boldsymbol{Q} \in \mathbb{M}^{n \times k} : \boldsymbol{Q}^* \boldsymbol{Q} = \mathbf{I} \},$$

J. A. Tropp (✉)
California Institute of Technology, Annenberg Center, MC 305-16, Pasadena,
CA, 91125-5000, USA
e-mail: jtropp@cms.caltech.edu

which is the collection of real $n \times k$ matrices with orthonormal columns. The elements of the Stiefel manifold $\mathbb{V}_k^n$ are sometimes called k-frames in $\mathbb{R}^n$. The range of a k-frame in $\mathbb{R}^n$ determines a k-dimensional subspace of $\mathbb{R}^n$, but the mapping from k-frames to subspaces is not injective.

It is easy to check that each Stiefel manifold is invariant under orthogonal transformations on the left and the right. An important consequence is that the Stiefel manifold $\mathbb{V}_k^n$ admits an invariant Haar probability measure, which can be regarded as a uniform distribution on k-frames in $\mathbb{R}^n$. A matrix $\boldsymbol{Q}$ drawn from the Haar measure on $\mathbb{V}_k^n$ is called a random k-frame in $\mathbb{R}^n$.

It can be challenging to compute functions of a random k-frame $\boldsymbol{Q}$. The main reason is that the entries of the matrix $\boldsymbol{Q}$ are correlated on account of the orthonormality constraint $\boldsymbol{Q}^* \boldsymbol{Q} = \boldsymbol{I}$. Nevertheless, if we zoom in on a small part of the matrix, the local correlations are very weak because orthogonality is a global property. In other words, the entries of a small submatrix of $\boldsymbol{Q}$ are effectively *independent* for many practical purposes [6].

As a consequence of this observation, we might hope to replace certain calculations on a random k-frame by calculations on a random matrix with *independent* entries. An obvious candidate is a matrix $\boldsymbol{G} \in \mathbb{M}^{n \times k}$ whose entries are independent $N(0, n^{-1})$ random variables. We call the associated probability distribution on $\mathbb{M}^{n \times k}$ the *normalized Gaussian distribution*.

Why is this distribution a good proxy for a random k-frame in $\mathbb{R}^n$? First, a normalized Gaussian matrix $\boldsymbol{G}$ verifies $\mathbb{E}(\boldsymbol{G}^* \boldsymbol{G}) = \boldsymbol{I}$, so the columns of $\boldsymbol{G}$ are orthonormal on average. Second, the normalized Gaussian distribution is invariant under orthogonal transformations on the left and the right, so it shares many algebraic and geometric properties with a random k-frame. Furthermore, we have a wide variety of methods for working with Gaussian matrices, in contrast with the more limited set of techniques available for dealing with random k-frames.

These intuitions are well established in the random matrix literature, and many authors have developed detailed quantitative refinements. In particular, we mention Jiang's paper [6] and its references, which discuss the proportion of entries in a random orthogonal matrix that can be simultaneously approximated using independent standard normal variables. Subsequent work by Chatterjee and Meckes [2] demonstrates that the joint distribution of r (linearly independent) linear functionals of a random orthogonal matrix is close in Wasserstein distance to an appropriate Gaussian distribution, provided that $r = o(n)$.

We argue that there is a general comparison principle for random k-frames and normalized Gaussian matrices of the same size. Recall that a convex function is called *sublinear* when it is positive homogeneous. Norms, in particular, are sublinear. Theorem 1 ensures that the expectation of a nonnegative sublinear function of a random k-frame is dominated by that of a normalized Gaussian matrix. This result also allows us to study moments and, therefore, tail behavior.

Theorem 1 (Sublinear Comparison Principle) *Assume that $k = \rho n$ for $\rho \in (0, 1]$. Let $\boldsymbol{Q}$ be uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$, and let $\boldsymbol{G} \in \mathbb{M}^{n \times k}$ be a matrix with independent $N(0, n^{-1})$ entries. For each nonnegative, sublinear, convex*

function $|\cdot|$ on $\mathbb{M}^{n \times k}$ and each weakly increasing, convex function $\Phi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\,\Phi(|\boldsymbol{Q}|) \leq \mathbb{E}\,\Phi((1 + \rho/2)\,|\boldsymbol{G}|).$$

In particular, for all $k \leq n$,

$$\mathbb{E}\,\Phi(|\boldsymbol{Q}|) \leq \mathbb{E}\,\Phi(1.5\,|\boldsymbol{G}|).$$

Note that the leading constant in the bound is asymptotic to one when $k = \mathrm{o}(n)$. Conversely, Sect. 2 identifies situations where the leading constant must be at least one. We establish Theorem 1 in Sect. 3 as a consequence of a more comprehensive result, Theorem 5, for convex functions of a random k-frame.

A simple example suffices to show that Theorem 1 does not admit a matching lower bound, no matter what comparison factor β we allow. Indeed, suppose that we fix a positive number β. Write $\|\cdot\|$ for the spectral norm (i.e., the operator norm between two Hilbert spaces), and consider the weakly increasing, convex function

$$\Phi(t) := (t - 1)_+ \quad \text{where} \quad (a)_+ := \max\{0, a\}.$$

For a normalized Gaussian matrix $\boldsymbol{G} \in \mathbb{M}^{n \times k}$, we compute that

$$\mathbb{E}\,\Phi(\beta\,\|\boldsymbol{G}\|) = \mathbb{E}\,(\beta\,\|\boldsymbol{G}\| - 1)_+ > 0$$

because there is always a positive probability that $\beta\,\|\boldsymbol{G}\| \geq 2$. Meanwhile, the spectral norm of a random k-frame $\boldsymbol{Q}$ in $\mathbb{R}^n$ satisfies $\|\boldsymbol{Q}\| = 1$, so

$$\mathbb{E}\,\Phi(\|\boldsymbol{Q}\|) = \mathbb{E}\,\Phi(1) = 0.$$

Inexorably,

$$\mathbb{E}\,\Phi(\beta\,\|\boldsymbol{G}\|) \leq \mathbb{E}\,\Phi(\|\boldsymbol{Q}\|) \quad \Longrightarrow \quad \beta \leq 0.$$

Therefore, it is impossible to control $\Phi(\beta\,|\boldsymbol{G}|)$ using $\Phi(|\boldsymbol{Q}|)$ unless we impose additional restrictions. Turn to Sect. 4 for some conditions under which we can reverse the comparison in Theorem 1.

One of the anonymous referees has made a valuable point that deserves amplification. Note that a random orthogonal matrix with dimension one is a scalar Rademacher variable, while a normalized Gaussian matrix with dimension one is a scalar Gaussian variable. From this perspective, Theorem 1 resembles a noncommutative version of the classical comparison between Rademacher series and Gaussian series in a Banach space [8, Sec. 4.2]. Let us state an extension of Theorem 1 that makes this connection explicit.

Theorem 2 (Noncommutative Gaussian Comparison Principle) *Fix a sequence of square matrices $\{A_j : j = 1, \ldots, J\} \subset \mathbb{M}^{n \times n}$. Consider an independent family $\{\boldsymbol{Q}_j : j = 1, \ldots, J\} \subset \mathbb{M}^{n \times n}$ of random orthogonal matrices and an independent*

family $\{G_j : j = 1, \ldots, J\} \subset \mathbb{M}^{n \times n}$ *of normalized Gaussian matrices. For each nonnegative, sublinear, convex function* $|\cdot|$ *on* $\mathbb{M}^{n \times n}$ *and each weakly increasing, convex function* $\Phi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E} \, \Phi \left(\left| \sum_{j=1}^{J} Q_j A_j \right| \right) \leq \mathbb{E} \, \Phi \left(1.5 \left| \sum_{j=1}^{J} G_j A_j \right| \right).$$

We can complete the proof of Theorem 2 using an obvious variation on the arguments behind Theorem 1. We omit further details out of consideration for the reader's patience.

2 A few examples

Before proceeding with the proof of Theorem 1, we present some applications that may be interesting. We need the following result [8, Thm. 3.20], which is due to Gordon [4].

Proposition 3 (Spectral Norm of a Gaussian Matrix) *Let* $G \in \mathbb{M}^{n \times k}$ *be a random matrix with independent* $\mathrm{N}(0, n^{-1})$ *entries. Then*

$$\mathbb{E} \, \|G\| \leq 1 + \sqrt{k/n}.$$

2.1 How good are the constants?

Consider a uniformly random orthogonal matrix $Q \in \mathbb{V}_n^n$. Evidently, its spectral norm $\|Q\| = 1$. Let $G \in \mathbb{M}^{n \times n}$ be a normalized Gaussian matrix. Theorem 1 and Proposition 3 ensure that

$$1 = \mathbb{E} \, \|Q\| \leq 1.5 \, \mathbb{E} \, \|G\| \leq 3.$$

Thus, the constant 1.5 in Theorem 1 cannot generally be improved by a factor greater than three.

Next, we specialize to the trivial case where $k = n = 1$. Let Q be a Rademacher random variable, and let G be a standard Gaussian random variable. Theorem 1 implies that

$$1 = \mathbb{E} \, |Q| \leq 1.5 \, \mathbb{E} \, |G| = 1.5 \sqrt{\frac{2}{\pi}} < 1.2.$$

Therefore, we cannot reduce the constant 1.5 below $\sqrt{\pi/2} \approx 1.253$ if we demand a result that holds when n is small.

Finally, consider the case where $k = 1$. Let q be a random unit vector in $\mathbb{R}^n$, and let g be a vector in $\mathbb{R}^n$ with independent $\mathrm{N}(0, n^{-1})$ entries. Applying Theorem 1 with

the Euclidean norm, we obtain

$$1 = \mathbb{E} \|\mathbf{q}\|_2 \leq \left(1 + \frac{1}{2n}\right) \cdot \mathbb{E} \|\mathbf{g}\|_2 \leq 1 + \frac{1}{2n}.$$

This example demonstrates that the best constant in Theorem 1 is at least one when $k = 1$ and n is large. Related examples show that the best constant is at least one as long as $k = \mathrm{o}(n)$.

2.2 Maximum entry of a random orthogonal matrix

Consider a uniformly random orthogonal matrix $\mathbf{Q} \in \mathbb{V}_n^n$, and let $\mathbf{G} \in \mathbb{M}^{n \times n}$ be a normalized Gaussian matrix. Using Theorem 1 and a standard bound for the maximum of standard Gaussian variables, we estimate that

$$\mathbb{E} \max_{i,j} |\mathbf{Q}_{ij}| \leq 1.5\, \mathbb{E} \max_{i,j} |\mathbf{G}_{ij}| \leq 1.5\sqrt{\frac{2\log(n^2) + 1}{n}} = 3\sqrt{\frac{\log(n) + 1/4}{n}}$$

Jiang [5] has shown that, almost surely, a sequence $\{\mathbf{Q}^{(n)}\}$ of random orthogonal matrices with $\mathbf{Q}^{(n)} \in \mathbb{V}_n^n$ has the limiting behavior

$$\liminf_{n \to \infty} \sqrt{\frac{n}{\log n}} \cdot \max_{i,j} |\mathbf{Q}_{ij}^{(n)}| = 2 \quad \text{and} \quad \limsup_{n \to \infty} \sqrt{\frac{n}{\log n}} \cdot \max_{i,j} |\mathbf{Q}_{ij}^{(n)}| = \sqrt{6}.$$

We see that our simple estimate is not sharp, but it is very reasonable.

2.3 Spectral norm of a submatrix of a random k-frame

Consider a uniformly random k-frame $\mathbf{Q} \in \mathbb{V}_k^n$, and let $\mathbf{G} \in \mathbb{M}^{n \times k}$ be a normalized Gaussian matrix. Define the linear map $\mathscr{L}_j$ that restricts an $n \times k$ matrix to its first j rows and rescales it by $\sqrt{n/j}$. As a consequence, the columns of the $j \times k$ matrix $\mathscr{L}_j(\mathbf{Q})$ approximately have unit Euclidean norm. We may compute that

$$\mathbb{E} \|\mathscr{L}_j(\mathbf{Q})\| \leq (1 + (k/2n))\, \mathbb{E} \|\mathscr{L}_j(\mathbf{G})\| \leq (1 + (k/2n))(1 + \sqrt{k/j})$$

because of Theorem 1 and Proposition 3.

This estimate is interesting because it applies for all values of j and k. Note that the leading constant $1 + (k/2n)$ is asymptotic to one whenever $k = \mathrm{o}(n)$. In contrast, we recall Jiang's result [6] that the total-variation distance between the distributions of $\mathscr{L}_j(\mathbf{Q})$ and $\mathscr{L}_j(\mathbf{G})$ vanishes if and only if $j, k = \mathrm{o}(\sqrt{n})$. A related fact is that, under a natural coupling of $\mathbf{Q}$ and $\mathbf{G}$, the matrix ℓ_∞-norm distance between $\mathscr{L}_j(\mathbf{Q})$ and $\mathscr{L}_j(\mathbf{G})$ vanishes in probability if and only if $k = \mathrm{o}(n/\log n)$.

3 Proof of the sublinear comparison principle

The main tool in our proof is a well-known theorem of Bartlett that describes the statistical properties of the QR decomposition of a standard Gaussian matrix, i.e., a matrix with independent N(0, 1) entries. See Muirhead's book [10] for a detailed derivation of this result.

Proposition 4 (The Bartlett Decomposition) *Assume that $k \leq n$, and let $\mathbf{\Gamma} \in \mathbb{M}^{n \times k}$ be a standard Gaussian matrix. Then*

$$\mathbf{\Gamma}_{n \times k} \sim \mathbf{Q}_{n \times k} \, \mathbf{R}_{k \times k}.$$

The factors $\mathbf{Q}$ and $\mathbf{R}$ are statistically independent. The matrix $\mathbf{Q}$ is uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$. The matrix $\mathbf{R}$ is a random upper-triangular matrix of the form

$$\mathbf{R} = \begin{bmatrix} X_1 & Y_{12} & Y_{13} & \ldots & Y_{1k} \\ & X_2 & Y_{23} & \ldots & Y_{2k} \\ & & \ddots & \ddots & \vdots \\ & & & X_{k-1} & Y_{k-1,k} \\ & & & & X_k \end{bmatrix}_{k \times k}.$$

The diagonal entries are nonnegative and $X_i^2 \sim \chi_{n-i+1}^2$; the super-diagonal entries $Y_{ij} \sim \mathrm{N}(0, 1)$. Furthermore, all these random variables are mutually independent.

We may now establish a comparison principle for a general convex function of a random k-frame.

Theorem 5 (Convex Comparison Principle) *Assume that $k \leq n$. Let $\mathbf{Q} \in \mathbb{M}^{n \times k}$ be uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$, and let $\mathbf{\Gamma} \in \mathbb{M}^{n \times k}$ be a standard Gaussian matrix. For each convex function $f : \mathbb{M}^{n \times k} \to \mathbb{R}$, it holds that*

$$\mathbb{E} \, f(\mathbf{Q}) \leq \mathbb{E} \, f(\alpha^{-1} \mathbf{\Gamma}) \quad where \quad \alpha := \alpha(k, n) := \frac{1}{k} \sum_{i=1}^{k} \mathbb{E}(X_i)$$

and X_i^2 is the nonnegative square root of a χ_{n-i+1}^2 random variable. Similarly, for each concave function $g : \mathbb{M}^{n \times k} \to \mathbb{R}$, it holds that

$$\mathbb{E} \, g(\mathbf{Q}) \geq \mathbb{E} \, g(\alpha^{-1} \mathbf{\Gamma}).$$

Proof The result is a direct consequence of the Bartlett decomposition and Jensen's inequality. Define $\mathbf{\Gamma}$, $\mathbf{Q}$, and $\mathbf{R}$ as in the statement of Proposition 4. Let $\mathbf{P} \in \mathbb{M}^{k \times k}$ be a uniformly random permutation matrix, independent from everything else.

First, observe that

$$\mathbb{E}(\mathbf{P} \mathbf{R} \mathbf{P}^T) = (\mathbb{E} \, \bar{\mathrm{tr}}(\mathbf{R})) \cdot \mathbf{I} = \alpha \mathbf{I} \quad where \quad \alpha := \frac{1}{k} \sum_{i=1}^{k} \mathbb{E}(X_i).$$

The symbol $\bar{\text{tr}}$ denotes the normalized trace. Since the function f is convex, Jensen's inequality allows that

$$\mathbb{E}\,f(\boldsymbol{Q}) = \mathbb{E}\,f(\alpha^{-1}\boldsymbol{Q}(\mathbb{E}\,\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^T)) \le \mathbb{E}\,f(\alpha^{-1}\boldsymbol{Q}\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^T).$$

It remains to simplify the random matrix in the last expression.

Recall that the Haar distribution on the Stiefel manifold $\mathbb{V}_k^n$ and the normalized Gaussian distribution on $\mathbb{M}^{n\times k}$ are both invariant under orthogonal transformations. Therefore, $\boldsymbol{Q} \sim \boldsymbol{Q}\boldsymbol{S}$ and $\boldsymbol{\Gamma} \sim \boldsymbol{\Gamma}\boldsymbol{S}^T$ for each fixed permutation matrix $\boldsymbol{S}$. It follows that

$$\begin{aligned}
\mathbb{E}[f(\alpha^{-1}\boldsymbol{Q}\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^T) \mid \boldsymbol{P}] &= \mathbb{E}[f(\alpha^{-1}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{P}^T) \mid \boldsymbol{P}] \\
&= \mathbb{E}[f(\alpha^{-1}\boldsymbol{\Gamma}\boldsymbol{P}^T) \mid \boldsymbol{P}] = \mathbb{E}\,f(\alpha^{-1}\boldsymbol{\Gamma}),
\end{aligned}$$

where we have also used the fact that $\boldsymbol{Q}$ and $\boldsymbol{R}$ are statistically independent. Combining the last two displayed formulas with the tower property of conditional expectation, we reach

$$\mathbb{E}\,f(\boldsymbol{Q}) \le \mathbb{E}\,\mathbb{E}[f(\alpha^{-1}\boldsymbol{Q}\boldsymbol{P}\boldsymbol{R}\boldsymbol{P}^T) \mid \boldsymbol{P}] = \mathbb{E}\,f(\alpha^{-1}\boldsymbol{\Gamma}).$$

The proof for concave functions is analogous. □

For Theorem 5 to be useful, we need to make some estimates for the constant $\alpha(k, n)$ that arises in the argument. To that end, we state without proof a simple result on the moments of a chi-square random variable.

Proposition 6 (Chi-Square Moments) *Let X be the nonnegative square root of a chi-square random variable with p degrees of freedom. Then*

$$\mathbb{E}(\Xi) = \frac{\sqrt{2}\cdot\Gamma((p+1)/2)}{\Gamma(p/2)}.$$

Given the identity from Proposition 6, standard inequalities for this ratio of gamma functions allow us to estimate the constant α in terms of elementary operations and radicals.

Lemma 7 (Estimates for the Constant) *The constant $\alpha(k, n)$ defined in Theorem 5 satisfies*

$$\frac{1}{k}\sum_{i=0}^{k-1}\sqrt{n-(i+1/2)} \le \alpha(k, n) \le \frac{1}{k}\sum_{i=0}^{k-1}\sqrt{n-i}.$$

Proof We require bounds for

$$\alpha = \frac{1}{k}\sum_{i=1}^{k}\mathbb{E}(X_i) \quad \text{where} \quad X_i^2 \sim \chi_{n-i+1}^2 \quad \text{and} \quad X_i \ge 0.$$

Proposition 6 states that

$$\mathbb{E}(X_i) = \frac{\sqrt{2} \cdot \Gamma((p_i + 1)/2)}{\Gamma(p_i/2)} \quad \text{for } p_i = n - i + 1.$$

This ratio of gamma functions appears frequently, and the following bounds are available.

$$\sqrt{p - 1/2} < \frac{\sqrt{2} \cdot \Gamma((p + 1)/2)}{\Gamma(p/2)} < \sqrt{p} \quad \text{for } p \geq 1/2.$$

Combine these relations and reindex the sums to reach the result.

The upper bound can be obtained directly from Jensen's inequality and the basic properties of a chi-square variable: $\mathbb{E}(X_i) \leq [\mathbb{E}(X_i^2)]^{1/2} = \sqrt{n - i + 1}$. In contrast, the lower bound seems to require hard analysis. □

For practical purposes, it is valuable to simplify the estimates from Lemma 7 even more. To accomplish this task, we interpret the sums in terms of basic integral approximations.

Lemma 8 (Simplified Estimates) *The constant $\alpha(k, n)$ defined in Theorem 5 satisfies*

$$\frac{2}{3k} \left[n^{3/2} - (n - k)^{3/2} \right] \leq \alpha(k, n) \leq \frac{2}{3k} \left[n^{3/2} - (n - k)^{3/2} \right] + \frac{1}{2k} \left[\sqrt{n} - \sqrt{n - k} \right].$$

The minimum value for the lower bound occurs when $k = n$, and

$$\frac{2}{3} \sqrt{n} \leq \alpha(n, n) \leq \frac{2}{3} \sqrt{n} + o(1) \quad as \ n \to \infty.$$

Furthermore, when we express $k = \rho n$ for $\rho \in (0, 1]$, it holds that

$$\frac{1}{\alpha(\rho n, n)} \leq \frac{1}{\sqrt{n}} \cdot (1 + \rho/2).$$

Proof Fix the parameters k and n. Define the real-valued function $h(x) = \sqrt{n - x}$, and observe that h is concave and decreasing on its natural domain. The lower bound for α from Lemma 7 implies that

$$\alpha \geq \frac{1}{k} \sum_{i=0}^{k-1} h(i + 1/2) \geq \frac{1}{k} \int_0^k h(x) \, dx.$$

To justify the second inequality, we observe that the sum corresponds with the midpoint-rule approximation to the integral. Because the integrand is concave, the midpoint rule must overestimate the integral. Evaluate the integral to obtain the stated lower bound.

To see that the minimum value for the lower bound occurs when $k = n$, notice that

$$k \longmapsto \frac{1}{k} \int_0^k h(x)\,dx$$

is the running average of a decreasing function. Of course, the running average also decreases.

Next, we use the relation $k = \rho n$ to simplify (the reciprocal of) the lower bound, which yields

$$\frac{1}{\alpha(\rho n, n)} \leq 1.5 \cdot n^{-1/2} \cdot \frac{\rho}{1 - (1-\rho)^{3/2}} \leq n^{-1/2} \cdot (1 + \rho/2).$$

The second inequality holds because the fraction defines a convex function of ρ on the interval $(0, 1]$, so we may bound it above by the chord $\rho \mapsto (2 + \rho)/3$ connecting the endpoints.

The proof of the upper bound follows from a related principle: The trapezoidal rule underestimates the integral of a concave function. Lemma 7 ensures that

$$\alpha \leq \frac{1}{k} \sum_{i=0}^{k-1} h(i) \leq \frac{1}{k} \left[\int_0^k h(x)\,dx + \frac{1}{2}\,(h(0) - h(k)) \right].$$

Here, we have applied the trapezoidal rule on the interval $[0, k]$ and then redistributed the terms associated with the endpoints. Evaluate the integral to complete the bound.

$\square$

We are now prepared to establish the main result.

Proof of Theorem 1 Let $\boldsymbol{Q}$ be a random matrix distributed uniformly on the Stiefel manifold $\mathbb{V}_k^n$, and let $\boldsymbol{G} \in \mathbb{M}^{n \times k}$ be a normalized Gaussian matrix. We can write $\boldsymbol{G} = n^{-1/2}\boldsymbol{\Gamma}$ where $\boldsymbol{\Gamma}$ is a standard normal matrix.

Suppose that $|\cdot|$ is a nonnegative, sublinear, convex function and that Φ is a weakly increasing, convex function. Then the function $\boldsymbol{M} \mapsto \Phi(|\boldsymbol{M}|)$ is also convex. Theorem 5 demonstrates that

$$\mathbb{E}\,\Phi(|\boldsymbol{Q}|) \leq \mathbb{E}\,\Phi\left(\left|\alpha^{-1}\boldsymbol{\Gamma}\right|\right) = \mathbb{E}\,\Phi\left(\alpha^{-1}\sqrt{n} \cdot |\boldsymbol{G}|\right).$$

For $k = \rho n$, Lemma 8 ensures that the constant α satisfies

$$\alpha^{-1}\sqrt{n} \leq 1 + \rho/2.$$

Given that the function Φ is increasing and $|\boldsymbol{G}| \geq 0$, we conclude that

$$\mathbb{E}\,\Phi(|\boldsymbol{Q}|) \leq \Phi\left(\alpha^{-1}\sqrt{n} \cdot |\boldsymbol{G}|\right) \leq \Phi((1 + \rho/2) \cdot |\boldsymbol{G}|).$$

This argument establishes the main part of the theorem. To establish the remaining assertion, we simply assign $\rho = 1$, the maximum value allowed. $\qquad\square$

4 Partial converses

There are at least a few situations where it is possible to reverse the inequality of Theorem 1. To develop these results, we record another basic observation about Gaussian matrices [10].

Proposition 9 (Polar Factorization) *Assume that $k \leq n$. Let $\mathbf{\Gamma} \in \mathbb{M}^{n \times k}$ be a standard Gaussian matrix. Then*

$$\mathbf{\Gamma}_{n \times k} \sim \mathbf{Q}_{n \times k} \mathbf{W}_{k \times k}.$$

The factors $\mathbf{Q}$ and $\mathbf{W}$ are statistically independent. The matrix $\mathbf{Q}$ is uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$, and the matrix $\mathbf{W}$ is the positive square root of a $k \times k$ Wishart matrix with n degrees of freedom.

The first converse concerns a right operator ideal norm; that is, a norm $\|\|\cdot\|\|$ that satisfies the relation $\|\|\mathbf{AB}\|\| \leq \|\|\mathbf{A}\|\| \cdot \|\mathbf{B}\|$, where $\|\cdot\|$ is the spectral norm.

Theorem 10 (Partial Converse I) *Assume that $k = \rho n$ for $\rho \in (0, 1]$. Let $\mathbf{Q}$ be uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$, and let $\mathbf{G} \in \mathbb{M}^{n \times k}$ be a normalized Gaussian matrix. For each right operator ideal norm $\|\|\cdot\|\|$, it holds that*

$$\mathbb{E}\,\|\|\mathbf{G}\|\| \leq (1 + \sqrt{\rho}) \cdot \mathbb{E}\,\|\|\mathbf{Q}\|\|.$$

Proof The proof uses the polar factorization of the Gaussian matrix described in Proposition 9. For a standard Gaussian matrix $\mathbf{\Gamma} \in \mathbb{M}^{n \times k}$,

$$\mathbb{E}\,\|\|\mathbf{G}\|\| = n^{-1/2}\,\mathbb{E}\,\|\|\mathbf{\Gamma}\|\| = n^{-1/2}\,\mathbb{E}\,\|\|\mathbf{QW}\|\| \leq n^{-1/2}\,\mathbb{E}(\|\|\mathbf{Q}\|\| \cdot \|\mathbf{W}\|)$$
$$= n^{-1/2}(\mathbb{E}\,\|\|\mathbf{Q}\|\|) \cdot (\mathbb{E}\,\|\mathbf{W}\|).$$

The last relation relies on the independence of the polar factors. To continue, we note that the Wishart square root $\mathbf{W}$ has the same distribution as $(\mathbf{\Gamma}^*\mathbf{\Gamma})^{1/2}$. Therefore,

$$n^{-1/2}\,\mathbb{E}\,\|\mathbf{W}\| = n^{-1/2}\,\mathbb{E}\left(\|\mathbf{\Gamma}^*\mathbf{\Gamma}\|^{1/2}\right) = n^{-1/2}\,\mathbb{E}\,\|\mathbf{\Gamma}\| = \mathbb{E}\,\|\mathbf{G}\| \leq 1 + \sqrt{k/n},$$

where the last bound follows from Gordon's result, Proposition 3. $\qquad\square$

A version of Theorem 10 also holds for higher moments:

$$\mathbb{E}(\|\|\mathbf{G}\|\|^m) \leq \mathbb{E}(\|\mathbf{G}\|^m) \cdot \mathbb{E}(\|\|\mathbf{Q}\|\|^m) \leq C\sqrt{m} \cdot (1 + \sqrt{\rho}) \cdot \mathbb{E}(\|\|\mathbf{Q}\|\|^m) \quad \text{when } m \geq 1.$$

The second inequality holds because moments of a Gaussian series are equivalent [8, Cor. 3.2].

We have a second result for other types of operator norms. We omit the proof, which, by now, should be obvious.

Theorem 11 (Partial Converse II) *Assume that $k \leq n$. Let $\boldsymbol{Q}$ be uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$, and let $\boldsymbol{G} \in \mathbb{M}^{k \times n}$ be a normalized Gaussian matrix. Suppose that $\|\cdot\|_Y$ is a norm on $\mathbb{R}^k$ and $\|\cdot\|_Z$ is a norm on $\mathbb{R}^n$. Then*

$$\mathbb{E}\,\|\boldsymbol{G}\|_{Y \to Z} \leq (n^{-1/2}\,\mathbb{E}\,\|\boldsymbol{T}\|_{Y \to Y}) \cdot (\mathbb{E}\,\|\boldsymbol{Q}\|_{Y \to Z})$$

where $\boldsymbol{T}$ is either the upper-triangular matrix $\boldsymbol{R}$ defined in Proposition 4 *or the Wishart square root $\boldsymbol{W}$ defined in Proposition* 9.

References

1. Ball, K.: An elementary introduction to modern convex geometry. In: Flavors of Geometry. Math. Sci. Res. Inst. Publ., vol. 31, pp. 1–58. Cambridge University Press (1997)
2. Chatterjee, S., Meckes, E.: Multivariate normal approximation using exchangeable pairs. Alea **4**, 257–283 (2008)
3. Dvoretsky, A.: Some results on convex bodies and Banach spaces. In: Proc. Intl. Symp. Linear Spaces, pp. 123–160, Jerusalem (1961)
4. Gordon, Y.: Gaussian processes and almost spherical sections of convex bodies. Ann. Probab. **16**(1), 180–188 (1988)
5. Jiang, T.: Maxima of entries of Haar distributed matrices. Probab. Theory Relat. Fields **131**, 121–144 (2005)
6. Jiang, T.: How many entries of a typical orthogonal matrix can be approximated by independent normals?. Ann. Probab. **34**(4), 1497–1529 (2006)
7. Johnson, W.B., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. Contemp. Math. **26**, 189–206 (1984)
8. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: Isoperimetry and Processes. Springer, Berlin (1991)
9. Milman, V.D.: A new proof of A. Dvoretsky's theorem on cross-sections of convex bodies. Funkcional. Anal. i Priložen **5**(4), 28–37 (1971)
10. Muirhead, R.J.: Aspects of Multivariate Statistical Theory. Wiley, New York (1982)