arXiv:1102.0534v1 [math.PR] 2 Feb 2011

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 size. 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 [JL84] shows that, typically, the metric geometry of a collection of N points is preserved when we project them onto a random subspace with dimension $\mathrm{O}(\log N)$. Another famous example is Dvoretsky's Theorem [Dvo61, Mil71, Bal97], which states that, typically, the intersection between the unit ball of a Banach space with dimension N and a random subspace with dimension $\mathrm{O}(\log N)$ 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*,

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

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 because of the orthonormality constraint $\boldsymbol{Q}^*\boldsymbol{Q} = \mathbf{I}$. 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 [Jia06].

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 $\mathrm{N}(0, n^{-1})$ random variables. We call this the *normalized Gaussian distribution* on $\mathbb{M}^{n\times k}$.

Why is this distribution a good proxy for a random k-frame in $\mathbb{R}^n$? First, a normalized Gaussian matrix $\boldsymbol{G}$ verifies $\mathbb{E}(\boldsymbol{G}^*\boldsymbol{G}) = \mathbf{I}$, so the columns of $\boldsymbol{G}$ are orthonormal on average. Second, the normalized Gaussian distribution is invariant under orthogonal transformations from the left and the right, so it share many algebraic and geometric properties with a random k-frame. Furthermore,

Date: 28 January 2011.
2010 *Mathematics Subject Classification.* Primary: 60B20.

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 [Jia06] and its references, which discuss the proportion of entries in a random orthogonal matrix that can be simultaneously approximated using independent standard normal variables. The present note has more limited aspirations, but it should complement earlier work.

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 $\mathrm{N}(0, n^{-1})$ entries. For each nonnegative sublinear function $|\cdot|$ on $\mathbb{M}^{n\times k}$ and each weakly increasing, convex function $\Phi : \mathbb{R} \to \mathbb{R}$,*

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

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

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

Note that the leading constant in the bound is asymptotic to one when $k = \mathrm{o}(n)$. When $k = n$, however, the constant 1.5 cannot be improved within our method. We establish Theorem 1 in Section 3 as a consequence of a more comprehensive result, Theorem 4, 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 $\|\cdot\|_{\mathrm{F}}$ for the Frobenius norm, and consider the weakly increasing, convex function

$$\Phi(t) := \left((t)_+ - \sqrt{k}\right)_+ \quad \text{where} \quad (\cdot)_+ := \max\{0, \cdot\}.$$

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

$$\mathbb{E}\,\Phi(\beta\,\|\boldsymbol{G}\|_{\mathrm{F}}) = \mathbb{E}\left(\beta\,\|\boldsymbol{G}\|_{\mathrm{F}} - \sqrt{k}\right)_+ > 0$$

because there is always a positive probability that $\beta\,\|\boldsymbol{G}\|_{\mathrm{F}} \ge \sqrt{k} + 1$. Meanwhile, the Frobenius norm of a random k-frame $\boldsymbol{Q}$ in $\mathbb{R}^n$ satisfies $\|\boldsymbol{Q}\|_{\mathrm{F}} = \sqrt{k}$, so

$$\mathbb{E}\,\Phi(\|\boldsymbol{Q}\|_{\mathrm{F}}) = \mathbb{E}\,\Phi(\sqrt{k}) = 0.$$

Inexorably,

$$\mathbb{E}\,\Phi(\beta\,\|\boldsymbol{G}\|_{\mathrm{F}}) \le \mathbb{E}\,\Phi(\|\boldsymbol{Q}\|_{\mathrm{F}}) \quad \Longrightarrow \quad \beta \le 0.$$

Therefore, it is impossible to control $\Phi(\beta\,|\boldsymbol{G}|)$ using $\Phi(|\boldsymbol{Q}|)$ without additional restrictions. But turn to Section 4 for some partial converses.

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, which is essentially due to Gordon [LT91, Thm. 3.20].

Proposition 2 (Spectral Norm)**.** *Assume that $k \le n$, and let $\boldsymbol{G} \in \mathbb{M}^{n\times k}$ be a normalized Gaussian matrix. Then*

$$\mathbb{E}\,\|\boldsymbol{G}\|_{\ell_2^k \to \ell_2^n} \le 1 + \sqrt{k/n}.$$

2.1. **Spectral norm of a random orthogonal matrix.** The first example is trivial. Consider a uniformly random orthogonal matrix $\boldsymbol{Q} \in \mathbb{V}_n^n$. Evidently, its spectral norm $\|\boldsymbol{Q}\| = 1$. Let $\boldsymbol{G} \in \mathbb{M}^{n\times n}$ be a normalized Gaussian matrix. Theorem 1 and Proposition 2 ensure that

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

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

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

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

Jiang [Jia05] has shown that a sequence $\{\boldsymbol{Q}^{(n)}\}$ of random orthogonal matrices almost surely satisfies

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

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 $\boldsymbol{Q} \in \mathbb{V}_k^n$, and let $\boldsymbol{G} \in \mathbb{M}^{n\times k}$ 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 $\mathscr{L}_j(\boldsymbol{Q})$ approximately have unit Euclidean norm. We obtain

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

because of Theorem 1 and Proposition 2.

This estimate is interesting because it applies for all values of j and k. Note that the leading constant is asymptotic to one whenever $k = \mathrm{o}(n)$. In contrast, we recall Jiang's result [Jia06] that the total-variation distance between the distributions of $\mathscr{L}_j(\boldsymbol{Q})$ and $\mathscr{L}_j(\boldsymbol{G})$ does not vanish unless $j, k = \mathrm{o}(\sqrt{n})$. A related fact is that, under a natural coupling of $\boldsymbol{Q}$ and $\boldsymbol{G}$, the matrix ℓ_∞-norm distance between $\mathscr{L}_j(\boldsymbol{Q})$ and $\mathscr{L}_j(\boldsymbol{G})$ does not vanish in probability unless $k = \mathrm{o}(n/\log n)$.

3. Proof of the 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 $\mathrm{N}(0,1)$ entries. See Muirhead's book [Mui82] for a detailed derivation of this result.

Proposition 3 (The Bartlett Decomposition)**.** *Assume that $k \le n$, and let $\boldsymbol{\Gamma} \in \mathbb{M}^{n\times k}$ be a standard Gaussian matrix. Then*

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

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

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

where the diagonal entries $X_i^2 \sim \chi^2_{n-i+1}$ and 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 4 (Convex Comparison Principle)**.** *Assume that $k \leq n$. Let $\boldsymbol{Q} \in \mathbb{M}^{n\times k}$ be uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$, and let $\boldsymbol{\Gamma} \in \mathbb{M}^{n\times k}$ be a standard Gaussian matrix. For each convex function $f : \mathbb{M}^{n\times k} \to \mathbb{R}$, it holds that*

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

and $X_i^2 \sim \chi^2_{n-i+1}$. Similarly, for each concave function $g : \mathbb{M}^{n\times k} \to \mathbb{R}$, it holds that

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

Proof. The result is a direct consequence of the Bartlett decomposition and Jensen's inequality. Define $\boldsymbol{\Gamma}$, $\boldsymbol{Q}$, and $\boldsymbol{R}$ as in the statement of Proposition 3. Let $\boldsymbol{P} \in \mathbb{M}^{k\times k}$ be a uniformly random permutation matrix, independent from everything else.

First, observe that

$$\mathbb{E}(\boldsymbol{PRP}^T) = (\mathbb{E}\operatorname{tr}(\boldsymbol{R}))\cdot \mathbf{I} = \alpha\mathbf{I} \quad \text{where} \quad \alpha := \frac{1}{k}\sum\nolimits_{i=1}^{k} \mathbb{E}(X_i).$$

Jensen's inequality allows that

$$\mathbb{E}\, f(\boldsymbol{Q}) = \mathbb{E}\, f(\alpha^{-1}\boldsymbol{Q}(\mathbb{E}\,\boldsymbol{PRP}^T)) \leq \mathbb{E}\, f(\alpha^{-1}\boldsymbol{QPRP}^T).$$

Since $\boldsymbol{Q}$ is uniformly distributed on the Stiefel manifold, $\boldsymbol{Q} \sim \boldsymbol{QS}$ for each fixed permutation $\boldsymbol{S}$. It follows that

$$\mathbb{E}[f(\alpha^{-1}\boldsymbol{QPRP}^T) \,|\, \boldsymbol{P}] = \mathbb{E}[f(\alpha^{-1}\boldsymbol{QRP}^T) \,|\, \boldsymbol{P}] = \mathbb{E}[f(\alpha^{-1}\boldsymbol{\Gamma P}^T) \,|\, \boldsymbol{P}] = \mathbb{E}\, f(\alpha^{-1}\boldsymbol{\Gamma})$$

because $\boldsymbol{\Gamma} \sim \boldsymbol{\Gamma S}^T$ for each fixed permutation $\boldsymbol{S}$. Combining the last two relations, we reach

$$\mathbb{E}\, f(\boldsymbol{Q}) \leq \mathbb{E}\,\mathbb{E}[f(\alpha^{-1}\boldsymbol{QPRP}^T) \,|\, \boldsymbol{P}] = \mathbb{E}\, f(\alpha^{-1}\boldsymbol{\Gamma}).$$

The proof for concave functions is analogous. □

For Theorem 4 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 5 (Chi-Square Moments)**.** *Let Ξ^2 be 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 of Proposition 5, standard inequalities for this ratio of gamma functions allow us to estimate the constant α in terms of elementary operations and radicals.

Lemma 6 (Estimates for the Constant)**.** *The constant $\alpha(k,n)$ defined in Theorem 4 satisfies*

$$\frac{1}{k}\sum\nolimits_{i=0}^{k-1}\sqrt{n-(i+1/2)} \quad \leq \quad \alpha(k,n) \quad \leq \quad \frac{1}{k}\sum\nolimits_{i=0}^{k-1}\sqrt{n-i}.$$

Proof. We require bounds for

$$\alpha = \frac{1}{k}\sum\nolimits_{i=1}^{k}\mathbb{E}(X_i) \quad \text{where} \quad X_i^2 \sim \chi^2_{n-i+1}.$$

Proposition 5 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 χ^2 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 6 even more. To accomplish this task, we interpret the sums in terms of basic integral approximations.

Lemma 7 (Simplified Estimates)**.** *The constant $\alpha(k,n)$ defined in Theorem 4 satisfies*

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

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

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

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

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

Proof. Define the real 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 6 implies that

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

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 sum 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

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

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 is 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 its endpoints.

The proof of the upper bound follows from a related principle: The trapezoidal rule underestimates the integral of a concave function. Lemma 6 ensures that

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

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. □

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 4 demonstrates that

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

For $k = \rho n$, Lemma 7 ensures that the constant α satisfies

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

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

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

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

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 [Mui82].

Proposition 8 (Polar Factorization)**.** *Assume that $k \le n$. Let $\boldsymbol{\Gamma} \in \mathbb{M}^{n\times k}$ be a standard Gaussian matrix. Then*

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

where the factors $\boldsymbol{Q}$ and $\boldsymbol{W}$ are statistically independent. The matrix $\boldsymbol{Q}$ is uniformly distributed on the Stiefel manifold $\mathbb{V}_k^n$. The matrix $\boldsymbol{W}$ 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 $|\!|\!|\boldsymbol{AB}|\!|\!| \le |\!|\!|\boldsymbol{A}|\!|\!| \cdot \|\boldsymbol{B}\|$, where $\|\cdot\|$ is the spectral norm.

Theorem 9 (Partial Converse I)**.** *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 normalized Gaussian matrix. For each right operator ideal norm $|\!|\!|\cdot|\!|\!|$, it holds that*

$$\mathbb{E}\,|\!|\!|\boldsymbol{G}|\!|\!| \le (1+\sqrt{\rho})\cdot\mathbb{E}\,|\!|\!|\boldsymbol{Q}|\!|\!|\,.$$

Proof. The proof uses the polar factorization of the Gaussian matrix described in Proposition 8. For a standard Gaussian matrix $\boldsymbol{\Gamma} \in \mathbb{M}^{n\times k}$,

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

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

$$n^{-1/2}\,\mathbb{E}\,\|\boldsymbol{W}\| = n^{-1/2}\,\mathbb{E}\,\big(\,\|\boldsymbol{\Gamma}^*\boldsymbol{\Gamma}\|^{1/2}\,\big) = n^{-1/2}\,\mathbb{E}\,\|\boldsymbol{\Gamma}\| = \mathbb{E}\,\|\boldsymbol{G}\| \le 1 + \sqrt{k/n},$$

where the last bound follows from Gordon's result, Proposition 2. □

We have a second result that holds for other types of operator norms, but the constant here is no longer explicit. We omit the proof, which, by now, should be obvious.

Theorem 10 (Partial Converse II)**.** *Assume that $k \le n$. 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*

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

where $\boldsymbol{T}$ is either the upper-triangular matrix $\boldsymbol{R}$ defined in Proposition 3 or the Wishart square root $\boldsymbol{W}$ defined in Proposition 8.

Acknowledgments

The author would like to thank Ben Recht and Michael Todd for encouraging him to refine and present these results. Alex Gittens and Tiefeng Jiang provided useful comments on a preliminary draft of this article. This work has been supported in part by ONR awards N00014-08-1-0883 and N00014-11-1-0025, AFOSR award FA9550-09-1-0643, and a Sloan Fellowship. Some of the research took place at Banff International Research Station (BIRS).

References

[Bal97] K. Ball. An elementary introduction to modern convex geometry. In *Flavors of Geometry*, number 31 in Math. Sci. Res. Inst. Publ., pages 1–58. Cambridge Univ. Press, 1997.

[Dvo61] A. Dvoretsky. Some results on convex bodies and Banach spaces. In *Proc. Intl. Symp. Linear Spaces*, pages 123–160, Jerusalem, 1961.

[Jia05] T. Jiang. Maxima of entries of Haar distributed matrices. *Probab. Theory Related Fields*, 131:121–144, 2005.

[Jia06] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.*, 34(4):1497–1529, 2006.

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.*, 26:189–206, 1984.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin, 1991.

[Mil71] V. D. Milman. A new proof of A. Dvoretsky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen*, 5(4):28–37, 1971.

[Mui82] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, NY, 1982.