

Note on the convex hull of the Stiefel manifold*

Kyle A. Gallivan[†] P.-A. Absil[‡]

July 19, 2010

Abstract

In this note, we characterize the convex hull of the Stiefel manifold and we find its strong convexity parameter. We also introduce the notion of roundness of a set and show that the Stiefel manifold is round.

Key words. convex hull, Stiefel manifold, strong convexity

1 Definitions and notation

Let p and n be positive integers with $p \leq n$. The (*compact*) *Stiefel manifold* is the set

$$\text{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\},$$

where I_p denotes the identity matrix of size p . We view $\text{St}(p, n)$ as a subset of $\mathbb{R}^{n \times p}$ endowed with the Frobenius norm. The fact that $\text{St}(p, n)$ has a natural manifold structure is not relevant here.

The *convex hull* of a subset \mathcal{V} of a vector space \mathcal{E} , denoted $\text{conv}(\mathcal{V})$, is the smallest convex set containing \mathcal{V} . The *closed convex hull* of a subset of \mathbb{R}^n is the closure of its convex hull. The *support function* of a nonempty subset \mathcal{S} of \mathbb{R}^n is the function $\mathfrak{S}_{\mathcal{S}} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$\mathbb{R}^n \ni x \mapsto \mathfrak{S}_{\mathcal{S}}(x) := \sup\{\langle s, x \rangle : s \in \mathcal{S}\}.$$

The *spectral norm* $\|X\|_2$ of a matrix $X \in \mathbb{R}^{n \times p}$ is its induced 2-norm, which is equal to the largest singular value of X and to the square root of the largest eigenvalue of $X^T X$. We let

$$\mathbb{B}_{\text{sp}}(p, n) := \{X \in \mathbb{R}^{n \times p} : \|X\|_2 \leq 1\} = \{X \in \mathbb{R}^{n \times p} : X^T X \preceq I_p\}$$

denote the unit spectral ball.

If \mathcal{E} is a normed vector space, the convex set $\mathcal{S} \subseteq \mathcal{E}$ is *strongly convex* if there exists a constant $\sigma > 0$ such that, for all $x, y \in \mathcal{S}$ and all $\alpha \in [0, 1]$, the following inclusion holds:

$$\alpha x + (1 - \alpha)y + \frac{\sigma}{2}\alpha(1 - \alpha)\|x - y\|^2 \mathcal{B} \subset \mathcal{S},$$

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

[†]Department of Mathematics, Florida State University, Tallahassee, Florida, 32306-4510, USA
<http://www.math.fsu.edu/~gallivan/>

[‡]Department of Mathematical Engineering, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
<http://www.inma.ucl.ac.be/~absil/>

where \mathcal{B} denotes the unit sphere in \mathcal{E} ; see [JNRS10]. Since \mathcal{S} is convex by assumption, “ $x, y \in \mathcal{S}$ ” can be equivalently replaced by “ $x, y \in \partial\mathcal{S}$ ” where $\partial\mathcal{S}$ denotes the boundary of the set. Strong convexity thus guarantees that a straight line cannot be embedded in any portion of the boundary of the convex set.

2 Convex hull of the Stiefel manifold

The next result shows that the convex hull of the Stiefel manifold is the (closed) unit ball in the spectral norm. The principle of the proof can be found in [JNRS10, §3.4].

Theorem 2.1 (Journée et al.) *The convex hull of the Stiefel manifold is the (closed) unit spectral ball:*

$$\text{conv}(\text{St}(p, n)) = \mathbb{B}_{\text{sp}}(p, n)$$

Proof. This proof is a detailed version of the argument given in [JNRS10, §3.4]. In view of proposition 2.1.1 and theorem 2.2.2 in [HUL01], two subsets of \mathbb{R}^n have the same closed convex hull if and only if they have the same support function. The support function of the unit spectral ball $\mathbb{B}_{\text{sp}}(p, n)$ is the trace norm:

$$S_{\mathbb{B}_{\text{sp}}(p, n)}(C) = \sum_{i=1}^p \sigma_i(C).$$

Equivalently stated, the dual norm of the spectral norm is the trace norm; see, e.g., [RFP07, prop. 2.1]. The support function of $\text{St}(p, n)$ is also the trace norm:

$$S_{\text{St}(p, n)}(C) = \sum_{i=1}^p \sigma_i(C);$$

see [JNRS10, prop. 7]. Hence the $\mathbb{B}_{\text{sp}}(p, n)$ and $\text{St}(p, n)$ have the same closed convex hull. Since the spectral norm is a norm, $\mathbb{B}_{\text{sp}}(p, n)$ is convex, and thus equal to its closed convex hull. Since $\text{St}(p, n)$ is compact, its convex hull is compact, hence its closed convex hull is its convex hull. Thus the convex hull of $\text{St}(p, n)$ is $\mathbb{B}_{\text{sp}}(p, n)$. \square

3 Convex hull of the Stiefel manifold – without support functions

In this section, we give an alternate proof of Theorem 2.1. It only applies to $p \leq n/2$. On the other hand, it does not need results pertaining to support functions, and it also yields the property

$$\text{conv}(\text{St}(p, n)) = \text{conv}_2(\text{St}(p, n)),$$

where

$$\text{conv}_2(S) := \{\alpha s_1 + (1 - \alpha)s_2 : s_1 \in S, s_2 \in S, \alpha \in [0, 1]\}$$

is the union of all line segments with extremities in S .

We will use the following results to show that $\text{conv}(\text{St}(p, n)) = \mathbb{B}_{\text{sp}}(p, n)$.

Lemma 3.1

$$\text{St}(p, n) \subseteq \text{conv}_2(\text{St}(p, n)) \subseteq \text{conv}(\text{St}(p, n)) \quad (1)$$

$$\text{St}(p, n) \subseteq \mathbb{B}_{\text{sp}}(p, n) \quad (2)$$

$$\mathbb{B}_{\text{sp}}(p, n) \text{ is convex} \quad (3)$$

Proof. The inclusions (1) and (2) follow immediately from the definitions. The convexity of $\mathbb{B}_{\text{sp}}(p, n)$ follows from the fact that $\mathbb{B}_{\text{sp}}(p, n)$ is a ball for a norm (specifically, the spectral norm). \square

We can now characterize $\text{conv}(\text{St}(p, n))$.

Theorem 3.2 *If $p \leq n/2$ then*

$$\text{conv}(\text{St}(p, n)) = \text{conv}_2(\text{St}(p, n)) = \mathbb{B}_{\text{sp}}(p, n).$$

Proof. Let $Q_1 \in \text{St}(p, n)$ and $Q_2 \in \text{St}(p, n)$. For $\alpha \in [0, 1]$, $Q_\alpha = \alpha Q_1 + (1 - \alpha)Q_2 \in \text{conv}_2(\text{St}(p, n))$. By Lemma 3.1, $\text{St}(p, n) \subseteq \mathbb{B}_{\text{sp}}(p, n)$ and $\mathbb{B}_{\text{sp}}(p, n)$ is convex. Therefore, $\text{conv}_2(\text{St}(p, n)) \subseteq \mathbb{B}_{\text{sp}}(p, n)$.

Now let $Z \in \mathbb{B}_{\text{sp}}(p, n)$ and $Z = U\Sigma V^T$ be a thin SVD with $\sigma_i \in [0, 1]$, $i = 1, \dots, p$, $U \in \mathbb{R}^{n \times p}$, and $V \in \mathbb{R}^{p \times p}$. Let $W \in \mathbb{R}^{n \times p}$ be such that $\begin{bmatrix} U & W \end{bmatrix}$ is orthonormal. Such a W exists since $p \leq n/2$. Let $Q_1 = (U\Sigma + W\sqrt{I_p - \Sigma^2})V^T$ and $Q_2 = (U\Sigma - W\sqrt{I_p - \Sigma^2})V^T$. Then $Q_1^T Q_1 = V(\Sigma^2 + (I_p - \Sigma^2))V^T = VV^T = I_p$. Likewise for Q_2 . It is easily verified that $Z = (Q_1 + Q_2)/2$ and so $\mathbb{B}_{\text{sp}}(p, n) \subseteq \text{conv}_2(\text{St}(p, n))$. Therefore, $\mathbb{B}_{\text{sp}}(p, n) = \text{conv}_2(\text{St}(p, n))$.

We have

$$\text{St}(p, n) \subseteq \mathbb{B}_{\text{sp}}(p, n) = \text{conv}_2(\text{St}(p, n)) \subseteq \text{conv}(\text{St}(p, n)).$$

By Lemma 3.1, $\mathbb{B}_{\text{sp}}(p, n)$ is convex and therefore so is $\text{conv}_2(\text{St}(p, n))$. $\text{conv}_2(\text{St}(p, n))$ is then a convex subset of $\text{conv}(\text{St}(p, n))$ that contains $\text{St}(p, n)$. Since $\text{conv}(\text{St}(p, n))$ is the smallest convex set that contains $\text{St}(p, n)$ it follows that $\text{conv}(\text{St}(p, n)) = \text{conv}_2(\text{St}(p, n)) = \mathbb{B}_{\text{sp}}(p, n)$. \square

So $\text{conv}(\text{St}(p, n))$ is characterized by the singular values of matrices in $\mathbb{R}^{n \times p}$ and only requires all convex combinations of pairs of matrices from $\text{St}(p, n)$.

4 $\text{conv}(\text{St}(p, n))$ is not strongly convex for $p > 1$

When $p = 1$, $\text{St}(1, n)$ is the unit sphere in \mathbb{R}^n which is clearly strongly convex.

This is not the case for $p > 1$, as pointed out in [JNRS10, §3.4]. A detailed proof of this result follows.

Theorem 4.1 *Let $n \geq 2$. If $p \geq 2$ then $\text{conv}(\text{St}(p, n))$ is not strongly convex, in the sense given in Section 1.*

Proof. Define $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times p}$ as

$$X = UI_n V^T, \quad Y = U \begin{bmatrix} I_k & 0 \\ 0 & 0_{n-k} \end{bmatrix} V^T$$

with $k \geq 1$. Note that $X \in \partial\text{conv}(\text{St}(p, n))$ and $Y \in \partial\text{conv}(\text{St}(p, n))$. Let $0 \leq \alpha \leq 1$. We have

$$Z_\alpha = \alpha X + (1 - \alpha)Y = U \begin{bmatrix} I_k & 0 \\ 0 & \alpha I_{n-k} \end{bmatrix} V^T$$

$$\therefore Z_\alpha \in \partial\text{conv}(\text{St}(p, n)).$$

So the line is completely contained in $\partial\text{conv}(\text{St}(p, n))$ hence there is no room to fit a ball around the points on the line. It follows that $\text{conv}(\text{St}(p, n))$ is not strongly convex. \square

5 $\text{St}(p, n)$ is “round”

We introduce the notion of roundness, which is weaker than strong convexity. The roundness of a set measures how far away the boundary of the set is from containing aligned points.

Definition 5.1 *Let \mathcal{E} be a normed vector space and let \mathcal{S} be a subset of \mathcal{E} , not necessarily convex. The value $\rho > 0$ is a roundness lower bound of \mathcal{S} if, for all $x, y \in \partial\mathcal{S}$ and all $\alpha \in [0, 1]$, the following holds:*

$$\alpha x + (1 - \alpha)y + \frac{\rho}{2}\alpha(1 - \alpha)\|x - y\|^2\mathcal{B} \cap \partial\mathcal{S} = \emptyset,$$

where \mathcal{B} is the unit ball in \mathcal{E} . The supremum of the roundness lower bounds is termed the roundness of \mathcal{S} and denoted by $\rho_{\mathcal{S}}$. A round set is a set whose roundness is nonzero.

It is clear, in view of the previous section, that $\text{conv}(\text{St}(p, n))$ is not round. For the general case $p \leq n$, we can obtain a bound on $\rho_{\text{St}(p, n)}$ using the following result that follows directly from the definition of roundness.

Lemma 5.2 *Let \mathcal{E} be a normed vector space. If $\mathcal{V} \subseteq \mathcal{S} \subseteq \mathcal{E}$, then*

$$\rho_{\mathcal{S}} \leq \rho_{\mathcal{V}}.$$

We then have the following result.

Theorem 5.3 *The roundness of the Stiefel manifold, $\rho_{\text{St}(p, n)}$, satisfies*

$$\frac{1}{\sqrt{p}} \leq \rho_{\text{St}(p, n)} \leq 1$$

Proof. The Stiefel manifold $\text{St}(p, n)$ is a subset of $\mathcal{B}_{\sqrt{p}}$, the sphere of radius \sqrt{p} in $\mathbb{R}^{n \times p}$. So $\rho_{\text{St}(p, n)}$ cannot be smaller than $\rho_{\mathcal{B}_{\sqrt{p}}}$. It is shown in [JNRS10] that the convexity parameter of the sphere of radius r is $\frac{1}{r}$, from which one can deduce the lower bound.

For the upper bound, consider $X \in \text{St}(p, n)$ and $Y \in \text{St}(p, n)$ such that only their first column differ. This yields a problem on the unit sphere in \mathbb{R}^n and therefore

$$\rho_{\text{St}(p, n)} \leq 1.$$

\square

If we require $p \leq n/2$ as in Theorem 3.2, then $\rho_{\text{St}(p, n)}$ can be determined.

Theorem 5.4 *If $p \leq n/2$ then the roundness of $\text{St}(p, n)$ is*

$$\rho_{\text{St}(p,n)} = \frac{1}{\sqrt{p}}.$$

Proof.

It is easily shown that $\text{St}(p, n) = \partial\text{St}(p, n)$. It follows that the set of pairwise convex combinations in Definition 5.1 is $\text{conv}_2(\text{St}(p, n)) = \mathbb{B}_{\text{sp}}(p, n)$ given in Section 1. By Lemma 3.1 we know $\text{conv}_2(\text{St}(p, n)) = \mathbb{B}_{\text{sp}}(p, n) = \text{conv}(\text{St}(p, n))$ and therefore our weaker definition checks the intersection of balls around lines in $\text{conv}(\text{St}(p, n))$ with $\partial\text{St}(p, n)$ rather than $\partial\text{conv}(\text{St}(p, n))$.

Let $Z = U\Sigma V^T$, $Q_1 = (U\Sigma + W\sqrt{I_p - \Sigma^2})V^T$, and $Q_2 = (U\Sigma - W\sqrt{I_p - \Sigma^2})V^T$ be as in the proof of Theorem 3.2. Recall that Q_1 and Q_2 belong to $\text{St}(p, n)$ and that $Z = \alpha Q_1 + (1 - \alpha)Q_2$ with $\alpha = 1/2$. Observe that UV^T belongs to $\text{St}(p, n)$. Hence, from the definition of $\rho_{\text{St}(p,n)}$, we must have

$$\|Z - UV^T\| \geq \frac{\rho_{\text{St}(p,n)}}{2} \frac{1}{4} \|Q_1 - Q_2\|^2.$$

This yields

$$\rho_{\text{St}(p,n)} \leq \frac{2\sqrt{\sum_{i=1}^p (1 - \sigma_i)^2}}{\sum_{i=1}^p (1 - \sigma_i^2)},$$

where the σ_i 's are arbitrary in $[0, 1]$. The right-hand side, for $\sigma_1 = \dots = \sigma_p =: s$, is equal to

$$\frac{1\sqrt{p}(1 - s)}{p(1 - s^2)} = \frac{2}{\sqrt{p}(1 + s)}.$$

As s goes to 1, this goes to $\frac{1}{\sqrt{p}}$. Hence $\rho_{\text{St}(p,n)} \leq \frac{1}{\sqrt{p}}$. Since we also know that $\rho_{\text{St}(p,n)} \geq \frac{1}{\sqrt{p}}$, the result follows. \square

References

- [HUL01] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Grundlehren Text Editions. Springer-Verlag, Berlin, 2001. Abridged version of it Convex analysis and minimization algorithms. I [Springer, Berlin, 1993; MR1261420 (95m:90001)] and it II [ibid.; MR1295240 (95m:90002)].
- [JNRS10] Michel Journée, Yurii Nesterov, Peter Richtárik, and Rodolphe Sepulchre. Generalized power method for sparse principal component analysis. *J. Mach. Learn. Res.*, 11:517–553, 2010.
- [RFP07] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, 2007. arXiv:0706.4138.