

HIGH CURVATURE MEANS LOW-RANK: ON THE SECTIONAL CURVATURE OF GRASSMANN AND STIEFEL MANIFOLDS AND THE UNDERLYING MATRIX TRACE INEQUALITIES

RALF ZIMMERMANN* AND JAKOB STOYE†

Abstract. Methods and algorithms that work with data on nonlinear manifolds are collectively summarized under the term ‘Riemannian computing’. In practice, curvature can be a key limiting factor for the performance of Riemannian computing methods. Yet, curvature can also be a powerful tool in the theoretical analysis of Riemannian algorithms. In this work, we investigate the sectional curvature of the Stiefel and Grassmann manifold. On the Grassmannian, tight curvature bounds are known since the late 1960ies. On the Stiefel manifold under the canonical metric, it was believed that the sectional curvature does not exceed $5/4$. Under the Euclidean metric, the maximum was conjectured to be at 1. For both manifolds, the sectional curvature is given by the Frobenius norm of certain structured commutator brackets of skew-symmetric matrices. We provide refined inequalities for such terms and pay special attention to the maximizers of the curvature bounds. In this way, we prove for the Stiefel manifold that the global bounds of $5/4$ (canonical metric) and 1 (Euclidean metric) hold indeed. With this addition, a complete account of the curvature bounds in all admissible dimensions is obtained. We observe that ‘high curvature means low-rank’, more precisely, for the Stiefel and Grassmann manifolds under the canonical metric, the global curvature maximum is attained at tangent plane sections that are spanned by rank-two matrices, while the extreme curvature cases of the Euclidean Stiefel manifold occur for rank-one matrices. Numerical examples are included for illustration purposes.

Key words. Stiefel manifold, Grassmann manifold, orthogonal group, sectional curvature, Riemannian computing

AMS subject classifications. 15B10, 15B57, 15B30, 65F99, 22E70, 53C30, 53C80,

1. Introduction. Methods and algorithms that work with data on nonlinear manifolds are collectively summarised under the term Riemannian computing. Riemannian computing methods have established themselves as important tools in a large variety of applications, including computer vision, machine learning, and optimization, see [1, 2, 5, 9, 11, 20, 26, 27] and the anthologies [21, 28]. They also gain increasing attention in statistics and data science [24] and in numerical methods for differential equations [6, 7, 14, 17, 34].

A standard technique in designing Riemannian computing methods is to translate Euclidean algorithms to manifolds. The discrepancy between a linear space and a non-linear manifold is quantified by the concept of curvature. Therefore, curvature can also be seen as a decisive factor that separates Riemannian algorithms from their Euclidean counterparts. On the other hand, curvature estimates allow to compare Euclidean distances and Riemannian distances for embedded submanifolds [4], which can be a powerful tool in the theoretical analysis of Riemannian methods.

In this work, we investigate the sectional curvature on the Stiefel and Grassmann manifolds. Both of these manifolds have been subject to extensive investigations before. On the Grassmannian, tight curvature bounds are known since the seminal papers of Wong [31, 32]. The sectional curvature on the Stiefel manifold under a parametric family of Riemannian metrics introduced in [16] was extensively studied in [22]. On the Stiefel manifold under the canonical metric, it was believed that

*University of Southern Denmark, Department of Mathematics and Computer Science, Odense, Denmark (zimmermann@imadasdu.dk, <https://portal.findresearcher.sdu.dk/en/persons/zimmermann>, <https://orcid.org/0000-0003-1692-3996>)

†TU Braunschweig, Institute for Numerical Analysis, Braunschweig, Germany, (jakob.stoye@tu-braunschweig.de, <https://orcid.org/0009-0003-9119-0013>)

the sectional curvature does not exceed $5/4$, [27, p. 94,95], [22, Section 6, Table 2]. Propositions 6.1 and 6.2 from the same work report that under the Euclidean metric, the curvature range contains the interval $[-1/2, 1]$. Nguyen [22] shows that these bounds are tight for the Stiefel manifold $St(n, p)$ if $p = 2$, but a valid proof for general $p \geq 2$ was missing.

For the orthogonal group and, in turn, for the Stiefel and Grassmann manifolds as its quotient spaces, the sectional curvature is given by the Frobenius norm of certain structured commutator brackets of skew-symmetric matrices. Sharp estimates are provided by the matrix inequality of [33] and the special bounds for skew-symmetric matrices of [12, Lemma 2.5]. We provide refined inequalities for such terms and pay special attention to the maximizers of the curvature bounds. In this way, we prove the conjectured global curvature bounds of

$$0 \leq K_c \leq \frac{5}{4} \text{ (canonical metric)} \quad \text{and} \quad -\frac{1}{2} \leq K_e \leq 1 \text{ (Euclidean metric)}.$$

We also show that these bounds are sharp on all true Stiefel manifolds $St(n, p)$, where $p \geq 2, n \geq p + 2$ and provide a discussion of the special cases, where $p = 1, p = n$, and $p = n - 1$.

In doing so, we observe that ‘high curvature means low-rank’. To be precise, we prove that for the orthogonal group and the Stiefel and Grassmann manifolds under the canonical metric, the global curvature maximum is attained at tangent plane sections that are spanned by the same low-rank tangent matrices.¹ From the perspective of the orthogonal group, these are rank-four matrices, while they are rank-two matrices from the Stiefel or Grassmann view point. Under the Euclidean metric, the extreme curvature cases occur for rank-one matrices. Numerical experiments confirm that the curvature drops with increasing rank.

Organisation of the paper. Section 2 provides the required curvature formulas for matrix Lie groups and their quotient spaces. (For the reader’s convenience, some essentials of Lie group theory is gathered in Appendix C.1.) Readers not interested in the matrix manifold applications but only in the matrix norm inequalities may directly skip to Section 3. In this section, we also give a full account of the sharp Stiefel curvature bounds in all possible dimensions. Numerical experiments are discussed in Section 4. Section 5 concludes the paper.

Notational specifics. For $n \in \mathbb{N}$, the $(n \times n)$ -identity matrix is denoted by $I_n \in \mathbb{R}^{n \times n}$, or simply I , if the dimension is clear. The $(n \times n)$ special orthogonal group, i.e., the set of all square orthogonal matrices with determinant 1 is denoted by

$$SO(n) = \{Q \in \mathbb{R}^{n \times n} | Q^T Q = Q Q^T = I_n\}.$$

The orthogonal group $O(n)$ is $SO(n)$ joint with their ‘ $\det = -1$ ’-siblings. The sets of symmetric and skew-symmetric $(n \times n)$ -matrices are $\text{sym}(n) = \{A \in \mathbb{R}^{n \times n} | A^T = A\}$ and $\text{skew}(n) = \{A \in \mathbb{R}^{n \times n} | A^T = -A\}$, respectively. Overloading this notation, $\text{sym}(A) = \frac{1}{2}(A + A^T)$, $\text{skew}(A) = \frac{1}{2}(A - A^T)$ denote the symmetric and skew-symmetric parts of a matrix A . The Stiefel manifold is

$$St(n, p) = \{U \in \mathbb{R}^{n \times p} | U^T U = I_p\}.$$

¹Because Stiefel and Grassmann are quotient spaces of the orthogonal group $SO(n)$, in this case, Stiefel and Grassmann tangent vectors can be considered as special (namely, horizontal) $SO(n)$ -tangent vectors.

The Grassmann manifold is

$$Gr(n, p) = \{[U] \mid U \in St(n, p)\}, \text{ where } [U] = \{UR \mid R \in SO(p)\}.$$

A word of caution: The standard Euclidean inner product on $\mathbb{R}^{n \times p}$ is

$$\langle X, Y \rangle_F = \text{tr}(X^T Y).$$

The subscript is to emphasise the correspondence with the Frobenius norm $\|X\|_F = \sqrt{\langle X, X \rangle_F}$. To comply with standard conventions in the matrix manifolds and Riemannian computing literature, the Riemannian metric on $SO(n)$ will *not* be the Euclidean one, but the Euclidean one with a multiplicative factor of $\frac{1}{2}$, i.e.,

$$(1) \quad \langle X, Y \rangle_Q = \frac{1}{2} \text{tr}(X^T Y) = \frac{1}{2} \langle X, Y \rangle_F, \quad X, Y \in T_Q SO(n).$$

This factor is inherited by the Riemannian metrics of the Grassmann and the Stiefel manifold, when considered as quotients of $SO(n)$. It also makes the curvature results in this work compatible with those stated, e.g., in the seminal papers [31, 32]. To distinguish the Riemannian and the Euclidean metric, for the former, the location is always given as a subscript $\langle \cdot, \cdot \rangle_Q$.

Let $\mathcal{M}$ be a quotient of $SO(n)$ under a Lie group action. (We will only consider $\mathcal{M} = St(n, p)$ or $\mathcal{M} = Gr(n, p)$.) For a tangent vector $X \in T_I SO(n) = \mathfrak{so}(n) = \text{skew}(n)$, X_m denotes the projection onto the horizontal space associated with $\mathcal{M}$. When a distinction is necessary, we will also write $X_{m, \mathcal{M}}$ to emphasize, which quotient manifold is considered. Likewise, X_v and $X_{v, \mathcal{M}}$ denote the projection onto the vertical space associated with $\mathcal{M}$.

2. Curvature of Lie groups and quotients of Lie groups. This section recaps the basic curvature formulas for Lie groups and quotients of Lie groups. This is classic textbook material. Our main references are [10], [15] and [23], to which we refer for the details. Readers not familiar with Lie groups or quotients of Lie groups may want to read [Appendix C](#) first.

2.1. Lie group curvature formulae. Let $\mathcal{G}$ be a Lie group with Lie algebra $\mathfrak{g} = T_I \mathcal{G}$ (the tangent space at I) and a bi-invariant metric. Let $X, Y \in \mathfrak{g}$ be linearly independent tangent vectors. The sectional curvature associated with the two-dimensional subplane spanned by $\{X, Y\} \subset \mathfrak{g}$ is

$$K(X, Y) = \frac{1}{4} \frac{\|[X, Y]\|_I^2}{\|X\|_I^2 \|Y\|_I^2 - \langle X, Y \rangle_I},$$

where $[X, Y] = XY - YX$. It depends only on the subplane, in this context often referred to as ‘*the section*’, not on the spanning vectors X, Y . For convenience, we will only consider $X, Y \in \mathfrak{g}$ forming an orthonormal basis (ONB), i.e., $\|X\|_I = 1 = \|Y\|_I$, $\langle X, Y \rangle_I = 0$, so that the sectional curvature is computed as

$$(2) \quad K(X, Y) = \frac{1}{4} \|[X, Y]\|_I^2,$$

see [10, Prop. 21.19, p. 636].

Remark 1. For $\mathcal{G} = SO(n)$, we use the metric of (1). Hence, when translated to the Frobenius norm, one has

$$K^{SO}(X, Y) = \frac{1}{2} \| [X, Y] \|_F^2, \quad \text{for } \|X\|_F = 1 = \|Y\|_F, \quad \langle X, Y \rangle_F = 0.$$

In order to obtain curvature expressions for Lie group quotients, we need to introduce a few more terms. The *adjoint map* $\mathbf{Ad} : \mathcal{G} \rightarrow GL(\mathfrak{g})$ assigns to each group element $A \in \mathcal{G}$ the linear map $\mathbf{Ad}_A : \mathfrak{g} \mapsto ABA^{-1}$. Note that $\mathbf{Ad}_A(I) = I$. The differential of $\mathbf{Ad}_A$ at I is the map $\text{Ad}_A := d(\mathbf{Ad}_A)_I : T_I \mathcal{G} = \mathfrak{g} \rightarrow \mathfrak{g} = T_{\mathbf{Ad}_A(I)} \mathcal{G}$. For a matrix Lie group, $\text{Ad}_A(X) = AXA^{-1}$.

The homogeneous $\mathcal{G}$ -space $\mathcal{G}/\mathcal{H}$ is *reductive*, if the Lie algebra $\mathfrak{g}$ (the tangent space at the identity) can be split into

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m},$$

where $\mathfrak{h}$ is the Lie algebra of $\mathcal{H}$ (the tangent space at the identity) and $\mathfrak{m} \subset \mathfrak{g}$ is a complementary subspace such that $\text{Ad}_Q(\mathfrak{m}) \subseteq \mathfrak{m}$ for all $Q \in \mathcal{H}$. Note that while the setting above is more general, in the cases that we eventually consider in this work, $\mathfrak{m}$ will always be the *orthogonal* complement of $\mathfrak{h}$ with respect to a suitable metric. For a tangent vector $X \in \mathfrak{g}$, the projections onto $\mathfrak{h}$ and $\mathfrak{m}$ are denoted by $X_{\mathfrak{h}}$ and $X_{\mathfrak{m}}$, respectively.

The quotient space $\mathcal{G}/\mathcal{H}$ is called *naturally reductive*, if

- (i) it is reductive with some decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$,
 - (ii) it has a $\mathcal{G}$ -invariant metric,
 - (iii) $\langle [X, Z]_{\mathfrak{m}}, Y \rangle_I = \langle X, [Z, Y]_{\mathfrak{m}} \rangle_I$ for all $X, Y, Z \in \mathfrak{m}$.
- Item (ii) means that the maps $\tau_G : \mathcal{G}/\mathcal{H} \rightarrow \mathcal{G}/\mathcal{H}, [A] \mapsto G \cdot [A]$ associated with the action $\mathcal{G} \times \mathcal{G}/\mathcal{H} \rightarrow \mathcal{G}/\mathcal{H}, (G, [A]) \mapsto G \cdot [A]$ are isometries. The precise condition is

$$\langle d(\tau_G)_{[A]}(X), d(\tau_G)_{[A]}(Y) \rangle_{\tau_G([A])} = \langle X, Y \rangle_{[A]} \quad \forall X, Y \in T_{[A]} \mathcal{G}/\mathcal{H}, [A] \in \mathcal{G}/\mathcal{H},$$

[10, Def. 23.5]. On the level of matrix representatives (and after identifying $T_{[A]} \mathcal{G}/\mathcal{H}$ with the horizontal space H_A), this boils down to the condition

$$\langle G\bar{X}, G\bar{Y} \rangle_{GA} = \langle \bar{X}, \bar{Y} \rangle_A \quad \text{for all } \bar{X}, \bar{Y} \in H_A, A \in \mathcal{G},$$

where $\bar{X}, \bar{Y} \in H_A$ are horizontal lifts of $X, Y \in T_{[A]} \mathcal{G}/\mathcal{H}$.

THEOREM 2 ([10], Prop. 23.29). *Let $\mathcal{G}$ be a connected Lie group such that the Lie algebra $\mathfrak{g}$ admits an $\text{Ad}(\mathcal{G})$ -invariant inner product $\langle \cdot, \cdot \rangle_I$. Let $\mathcal{G}/\mathcal{H}$ be a homogeneous space as above and let $\mathfrak{m} = \mathfrak{h}^\perp$ with respect to $\langle \cdot, \cdot \rangle_I$. Then*

1. *The space $\mathcal{G}/\mathcal{H}$ is reductive with respect to the decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$.*
2. *Under the $\mathcal{G}$ -invariant metric induced by the inner product, the homogeneous space $\mathcal{G}/\mathcal{H}$ is naturally reductive.*
3. *For an ONB spanned by $X, Y \in T_{[I]} \mathcal{G}/\mathcal{H} \cong \mathfrak{m}$, $X \perp Y$, $\|X\|_I = \|Y\|_I = 1$, the sectional curvature associated with the tangent plane spanned by X, Y is*

$$(3) \quad K(X, Y) = \frac{1}{4} \| [X, Y]_{\mathfrak{m}} \|_I^2 + \| [X, Y]_{\mathfrak{h}} \|_I^2.$$

Transport to arbitrary locations. Since the tangent space at an arbitrary location $A \in \mathcal{G}$ is given by the translates of the tangent space at the identity (see (36)), a tangent vector $\tilde{X} \in T_A \mathcal{G}$ at an arbitrary location $A \in \mathcal{G}$ is of the form $\tilde{X} = AX$ with

$X \in \mathfrak{g}$. For an ONB $\{\tilde{X} = AX, \tilde{Y} = AY\}$ that spans a tangent plane in $T_A\mathcal{G}$, the sectional curvature is

$$K^{\mathcal{G}}(\tilde{X}, \tilde{Y}) = \frac{1}{4} \|[X, Y]\|_I^2.$$

For quotient spaces, because of the natural reductive homogeneous space structure, the horizontal spaces H_A at arbitrary locations $A \in \mathcal{G}$ are translates of the horizontal space $\mathfrak{m} \cong T_{[I]}\mathcal{G}/\mathcal{H}$. Therefore, the tangent space of the quotient $\mathcal{G}/\mathcal{H}$ is identified with

$$T_{[A]}\mathcal{G}/\mathcal{H} \cong H_A = A\mathfrak{m}.$$

Let $\tilde{X}, \tilde{Y} \in T_{[A]}\mathcal{G}/\mathcal{H}$ be an ONB of a tangent plane in the quotient space. Let $\bar{X}, \bar{Y} \in H_A \subset T_A\mathcal{G}$ be horizontal lifts of $\tilde{X}, \tilde{Y} \in T_{[A]}\mathcal{G}/\mathcal{H} \cong A\mathfrak{m}$ with $\bar{X} = AX, \bar{Y} = AY, X, Y \in \mathfrak{m}$. Then the sectional curvature at $[A] \in T_{[A]}\mathcal{G}/\mathcal{H}$ with respect to the tangent plane is given by

$$(4) \quad K^{\mathcal{G}/\mathcal{H}}(\tilde{X}, \tilde{Y}) = \frac{1}{4} \|[X, Y]_{\mathfrak{m}}\|_I^2 + \|[X, Y]_{\mathfrak{h}}\|_I^2.$$

For details, see [10, §19–23].

2.2. Canonical curvature formulae on Grassmann and Stiefel. In this section we recap the expressions for the sectional curvature on the Grassmann and the Stiefel manifold. Both the Stiefel manifold $St(n, p)$ and the Grassmann manifold $Gr(n, p)$ are considered as quotients of $SO(n)$. The Riemannian metric on the total space $SO(n)$ is

$$(5) \quad \langle X, Y \rangle_Q = \frac{1}{2} \operatorname{tr}(X^T Y), \quad X, Y \in T_Q SO(n) = Q \operatorname{skew}(n).$$

All conditions of [Theorem 2](#) are fulfilled for $\mathcal{G} = SO(n)$ with this metric and its quotient spaces $St(n, p)$ and $Gr(n, p)$.

Grassmann sectional curvature. The Grassmann manifold of linear subspaces can be realized as the quotient space $Gr(n, p) = SO(n)/(SO(p) \times SO(n-p))$. The canonical projection and the left cosets are

$$\begin{aligned} \Pi^{Gr} &: SO(n) \rightarrow Gr(n, p) \\ [Q] =: \Pi^{Gr}(Q) &= \{Q \begin{pmatrix} S & 0 \\ 0 & R \end{pmatrix} \mid \begin{pmatrix} S & 0 \\ 0 & R \end{pmatrix} \in SO(p) \times SO(n-p)\}. \end{aligned}$$

By splitting $Q = (U \mid U_{\perp})$ with $U \in \mathbb{R}^{n \times p}, U_{\perp} \in \mathbb{R}^{n \times (n-p)}, [Q] \in Gr(n, p)$ is uniquely represented by $\operatorname{ran}(U) \subset \mathbb{R}^n$.

When lifting $[Q] \in Gr(n, p)$ to $Q \in SO(n)$ (in practice, this is nothing but fixing a representative Q for the equivalence class $[Q]$), the vertical space is represented by

$$V_Q^{Gr} = \{\bar{X} = Q \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} \mid A \in \operatorname{skew}(p), C \in \operatorname{skew}(n-p)\} = Q\mathfrak{h}^{Gr}.$$

The associated horizontal space is

$$T_{[Q]}Gr(n, p) \cong H_Q^{Gr} = \{\bar{X} = Q \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix} \mid B \in \mathbb{R}^{(n-p) \times p}\} = Q\mathfrak{m}^{Gr}.$$

For Grassmann tangent vectors (already represented in matrix form by their lifts)

$X = \begin{pmatrix} 0 & -B_1^T \\ B_1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & -B_2^T \\ B_2 & 0 \end{pmatrix} \in \mathfrak{m}^{Gr} \cong T_{[I]}Gr(n, m)$, it holds that X, Y are

orthonormal w.r.t. (5) if and only if $B_1, B_2 \in \mathbb{R}^{(n-p) \times p}$ are orthonormal w.r.t. the standard Euclidean inner product $\langle \cdot, \cdot \rangle_F$. For such orthonormal X, Y , a straightforward evaluation of (3) yields

$$\begin{aligned}
 K^{Gr}(X, Y) &= \frac{1}{4} \|[X, Y]_{\mathfrak{m}^{Gr}}\|_{[I]}^2 + \|[X, Y]_{\mathfrak{h}^{Gr}}\|_{[I]}^2 \\
 (6) \quad &= \frac{1}{2} \|B_1^T B_2 - B_2^T B_1\|_F^2 + \frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 \\
 (7) \quad &= \text{tr}(B_1^T B_2 B_2^T B_1) + \text{tr}(B_1 B_2^T B_2 B_1^T) - 2 \text{tr}(B_1^T B_2 B_1^T B_2).
 \end{aligned}$$

Stiefel sectional curvature. The Stiefel manifold of orthonormal p -frames can be realized as the quotient space $St(n, p) = SO(n)/SO(n-p)$. The canonical projection and the left cosets are

$$\begin{aligned}
 \Pi^{St} &: SO(n) \rightarrow St(n, p) \\
 [Q] =: \Pi^{St}(Q) &= \{Q \begin{pmatrix} I & 0 \\ 0 & R \end{pmatrix} \mid R \in SO(n-p)\}.
 \end{aligned}$$

By splitting $Q = (U \mid U_\perp)$, $[Q] \in St(n, p)$ is uniquely determined by $U \in \mathbb{R}^{n \times p}$. When lifting $[Q] \in St(n, p)$ to $Q \in SO(n)$, the vertical space is represented by

$$V_Q^{St} = \{\bar{X} \in \mathbb{R}^{n \times n} \mid \bar{X} = Q \begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix}, C \in \text{skew}(n-p)\} = Q\mathfrak{h}^{St}.$$

The associated horizontal space is

$$T_{[Q]} St(n, p) \cong H_Q^{St} = \{\bar{X} = Q \begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix} \mid A \in \text{skew}(p), B \in \mathbb{R}^{(n-p) \times p}\} = Q\mathfrak{m}^{St}.$$

For tangent vectors $X = \begin{pmatrix} A_1 & -B_1^T \\ B_1 & 0 \end{pmatrix}, Y = \begin{pmatrix} A_2 & -B_2^T \\ B_2 & 0 \end{pmatrix} \in \mathfrak{m}^{St} \cong T_{[I]} St(n, p)$ that are orthonormal w.r.t. (5), a straightforward evaluation of (3) yields

$$\begin{aligned}
 K_c^{St}(X, Y) &= \frac{1}{4} \|[X, Y]_{\mathfrak{m}^{St}}\|_{[I]}^2 + \|[X, Y]_{\mathfrak{h}^{St}}\|_{[I]}^2 \\
 &= \frac{1}{2} \|B_2 B_1^T - B_1 B_2^T\|_F^2 + \frac{1}{4} \|B_1 A_2 - B_2 A_1\|_F^2 \\
 (8) \quad &+ \frac{1}{8} \|[A_1, A_2] - (B_1^T B_2 - B_2^T B_1)\|_F^2.
 \end{aligned}$$

This is in line with the result of [22, Prop 4.2, eq. (34)].

3. Matrix norm inequalities and curvature estimates. In this section, we investigate the extremal behavior of the sectional curvature on the Grassmann and the Stiefel manifold. It is known since [32], that the sectional curvature on the Grassmann manifold ranges in the interval $[0, 2]$ with both the lower and the upper bound attained. The upper bound can be established by using the matrix inequality

$$(9) \quad \|B_1 B_2^T - B_2 B_1^T\|_F^2 \leq 2 \|B_1\|_F^2 \|B_2\|_F^2$$

of Wu and Chen, [33].

Applying (9) to the terms (6) and keeping $\|B_1\|_F = 1 = \|B_2\|_F$ in mind immediately gives $0 \leq K^{Gr}(X, Y) \leq 2$. A key idea in [33] is to exploit the skew-symmetry of

the real Schur form of $B_1 B_2^T - B_2 B_1^T$. They also show by algebraic means that the inequality is sharp if and only if

$$B_1 = \left(\begin{array}{cc|c} 0 & 1 & \mathbf{0} \\ \lambda & \mu & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right), \quad B_2 = \left(\begin{array}{cc|c} \lambda & \mu & \mathbf{0} \\ 0 & -1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right),$$

up to an orthogonal transformation and scaling. Hence, under the normalization constraint $\|B_1\|_F = 1 = \|B_2\|_F$, both terms in (6) attain their upper bound of 2 simultaneously for the rank-2 matrices

$$B_1 = \frac{1}{\sqrt{2}} \left(\begin{array}{cc|c} 0 & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right), \quad B_2 = \frac{1}{\sqrt{2}} \left(\begin{array}{cc|c} 1 & 0 & \mathbf{0} \\ 0 & -1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right).$$

Observe that $\langle B_1, B_2 \rangle_F = 0$. Hence, this matrix pair forms an ONB and so do the associated Grassmann tangent vectors.

To gain more insight on how the various trace terms contribute to the overall curvature value, we will re-establish this result via an optimization approach. To this end, we will consider the three trace terms

$$\text{tr}(B_1^T B_2 B_2^T B_1), \quad \text{tr}(B_1 B_2^T B_2 B_1^T), \quad \text{and} \quad -2 \text{tr}(B_1^T B_2 B_1^T B_2) \quad B_1, B_2 \in \mathbb{R}^{(n-p) \times p},$$

in the curvature expression (7) separately. Eventually, this will lead to an alternative proof and a refinement of the matrix inequality of Wu and Chen, [33]. We start with a preparatory lemma that improves on the classical submultiplicativity property $\|AB\|_F \leq \|A\|_F \|B\|_F$.

LEMMA 3. *Let $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{m \times p}$. Then*

$$(10) \quad \|AB\|_F \leq \min\{\|A\|_2 \|B\|_F, \|A\|_F \|B\|_2\}.$$

If either A or B is skew-symmetric, then

$$(11) \quad \|AB\|_F \leq \frac{1}{\sqrt{2}} \|A\|_F \|B\|_F.$$

Proof. On (10): First, consider $D \in \mathbb{R}^{m \times m}$ diagonal. Write $B = (b_1, \dots, b_p)$ column-wise. It holds

$$\begin{aligned} \|DB\|_F^2 &= \text{tr}(B^T D^2 B) = \sum_{j=1}^p b_j^T D^2 b_j = \sum_{j=1}^p \|b_j\|_2^2 \frac{b_j^T D^2 b_j}{b_j^T b_j} \\ &\leq \max_j \{d_j^2\} \sum_{j=1}^p \|b_j\|_2^2 = \|D\|_2^2 \|B\|_F^2. \end{aligned}$$

The general case can be reduced to this case. Let $U \Sigma V^T = A$ be the full SVD of A .

$$\begin{aligned} \|AB\|_F^2 &= \|U \Sigma V^T B\|_F^2 = \text{tr}(B^T V \Sigma^T \Sigma V^T B) \\ &= \|\sqrt{\Sigma^T \Sigma} V^T B\|_F^2 \stackrel{D=\sqrt{\Sigma^T \Sigma}}{=} \sigma_1^2 \|V^T B\|_F^2 = \|A\|_2^2 \|B\|_F^2. \end{aligned}$$

When working with the SVD of B , the same argument may be applied to $\|B^T A^T\|_F^2$ and yields $\|B^T A^T\|_F^2 \leq \|B^T\|_2^2 \|A^T\|_F^2 = \|B\|_2^2 \|A\|_F^2$.

The second inequality (11) comes as a corollary. W.l.o.g. assume that $A \in \text{skew}(m)$. Then, all eigenvalues of A are either zero or form purely imaginary, complex conjugate pairs. The singular values of A are the absolute values of the eigenvalues of A . Therefore, the singular values also come in pairs. Let $\sigma_1 = \sigma_2 \geq \dots \geq \sigma_{r-1} = \sigma_r > 0$ be the non-zero singular values of A . We obtain

$$\|A\|_2^2 = \sigma_1^2 = \frac{1}{2} (\sigma_1^2 + \sigma_2^2) \leq \frac{1}{2} (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2) = \frac{1}{2} \|A\|_F^2.$$

Hence, (10) yields $\|AB\|_F \leq \|A\|_2 \|B\|_F \leq \frac{1}{\sqrt{2}} \|A\|_F \|B\|_F$. $\square$

Alternatively, (10) can also be established as a consequence of [30, Lemma 1].

The following lemma is obvious.

LEMMA 4. Let $m \geq p$ and consider $B_2 \in \mathbb{R}^{m \times p}$ as fixed. Let $U\Sigma V^T = B_2$ be the (full) SVD of B_2 with $\Sigma = \begin{pmatrix} \Sigma_p \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times p}$ and $\Sigma_p = \text{diag}(\sigma_1, \dots, \sigma_p)$.

(1.) The global optimum of

$$\max_{B_1 \in \mathbb{R}^{m \times p}} \text{tr}(B_1^T B_2 B_2^T B_1) \quad \text{s.t.} \quad \|B_1\|_F = 1$$

is σ_1^2 and is attained for any normalized B_1 such that $\tilde{B}_1 := U^T B_1 V$ only features scaled copies of the first unit vector $e_1 = (\pm 1, 0, \dots, 0)^T \in \mathbb{R}^m$ as columns.

(2.) The global optimum of

$$\max_{B_1 \in \mathbb{R}^{m \times p}} \text{tr}(B_1 B_2^T B_2 B_1^T) \quad \text{s.t.} \quad \|B_1\|_F = 1$$

is σ_1^2 and is attained for any normalized B_1 such that $\tilde{B}_1 := U^T B_1 V$ only features scaled copies of the first unit vector $e_1 = (\pm 1, 0, \dots, 0)^T \in \mathbb{R}^p$ as rows.

The next lemma concerns the third trace term in (7).

LEMMA 5. Let $m \geq p$ and let $B_2 = U\Sigma V^T \in \mathbb{R}^{m \times p}$ (not necessarily normalized). Then

$$\begin{aligned} \max_{B_1 \in \mathbb{R}^{m \times p}, \|B_1\|_F=1} \text{tr}(B_1^T B_2 B_1^T B_2) &= \sigma_1^2, \\ \min_{B_1 \in \mathbb{R}^{m \times p}, \|B_1\|_F=1} \text{tr}(B_1^T B_2 B_1^T B_2) &= -\sigma_1 \sigma_2. \end{aligned}$$

The global maximum and minimum are attained for $B_1^+ = U\tilde{B}_1^+ V^T$ and $B_1^- = U\tilde{B}_1^- V^T$, where

$$(12) \quad \tilde{B}_1^+ = \pm \left(\begin{array}{cc|c} 1 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right), \text{ and } \tilde{B}_1^- = \pm \frac{1}{\sqrt{2}} \left(\begin{array}{cc|c} 0 & 1 & \mathbf{0} \\ -1 & 0 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right), \text{ respectively.}$$

For B_1, B_2 both of unit Frobenius norm,

$$-\frac{1}{2} \leq \text{tr}(B_1^T B_2 B_1^T B_2) \leq 1.$$

In this case, the global extrema are attained for B_1^+, B_1^- as above and B_2^+ of rank one ($\sigma_1 = 1$) and B_2^- of rank 2 ($\sigma_1 = \sigma_2 = \frac{1}{\sqrt{2}}$).

Remark 6. The extrema may not be isolated nor are they necessarily unique (up to the trace-preserving transformation). If $m = p$, $B_2 = \sigma I$, then any normalized symmetric B_1 yields the maximum $\text{tr}(B_1^T B_2 B_1^T B_2) = \sigma^2 \|B_1\|_F^2 = \sigma_1^2$. Likewise, any normalized skew-symmetric B_1 yields the minimum $\text{tr}(B_1^T B_2 B_1^T B_2) = -\sigma^2 \|B_1\|_F^2 = -\sigma^2 = (-\sigma_1 \sigma_2)$. However, for both B_1 and B_2 of unit Frobenius norm, the global extrema are attained only for the matrices in (12) (up to the trace preserving transformation).

Proof (Lemma 5). Let $B_2 = U \Sigma V^T \in \mathbb{R}^{m \times p}$ be the full SVD of B_2 . Under the norm-preserving bijection $B \mapsto U^T B V =: \tilde{B}$, the trace optimization problem $\min_{B \in \mathbb{R}^{m \times p}, \|B\|_F=1} \pm \text{tr}(B^T B_2 B^T B_2)$ becomes $\min_{\tilde{B} \in \mathbb{R}^{m \times p}, \|\tilde{B}\|_F=1} \pm \text{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma)$.

Write $\Sigma = \begin{pmatrix} \Sigma_p \\ \mathbf{0} \end{pmatrix}$ and $\tilde{B} = \begin{pmatrix} \tilde{B}_p \\ \tilde{B}_{m-p} \end{pmatrix}$, with empty lower blocks if $m = p$. As a preliminary, observe that $\text{tr}(\tilde{B}_p^T \tilde{B}_p^T) = \langle \tilde{B}_p, \tilde{B}_p^T \rangle_F = \sum_{j=1}^p \sum_{k=1}^p b_{jk} b_{kj}$. It holds

$$\begin{aligned} (13) \quad \text{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma) &= \sum_{j=1}^p \sum_{k=1}^p \sigma_j \sigma_k b_{jk} b_{kj} \\ &= \sigma_1^2 b_{11}^2 + \sum_{k=2}^p \sigma_1 \sigma_k b_{1k} b_{k1} + \sum_{j=2}^p \left(\sigma_j \sigma_1 b_{j1} b_{1j} + \sum_{k=2}^p \sigma_j \sigma_k b_{jk} b_{kj} \right) \\ &\leq \sigma_1^2 \langle \tilde{B}_p, \tilde{B}_p^T \rangle_F \leq \sigma_1^2 \|\tilde{B}\|_F^2 = \sigma_1^2. \end{aligned}$$

The maximum is attained, if all weight in $\tilde{B}$ is put on the upper diagonal entry, i.e., for $\tilde{B}^+$ as in the statement of the lemma. If $\sigma_1 > \sigma_2$, then the maximum is isolated.

Now on the minimum. Reconsider (13) and extract the ‘diagonal terms’,

$$\text{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma) = \sum_{j=1}^p \sum_{k=1, k \neq j}^p \sigma_j \sigma_k b_{jk} b_{kj} + \sum_{j=1}^p \sigma_j^2 b_{jj}^2.$$

One sees that the diagonal terms in $\tilde{B}_p$ only make nonnegative contributions to the trace total. All remaining terms become non-positive, if b_{jk} and b_{kj} are of opposite sign for all $k \neq j$. In this case and with $b_{jj} = 0$, it holds

$$\begin{aligned} \text{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma) &= - \sum_{j=1}^p \sum_{k=1, k \neq j}^p \sigma_j \sigma_k b_{jk} b_{kj} \geq -\sigma_1 \sigma_2 \sum_{j=1}^p \sum_{k=1, k \neq j}^p b_{jk} b_{kj} \\ &\geq -\sigma_1 \sigma_2 \langle \tilde{B}_p, \tilde{B}_p^T \rangle_F \geq -\sigma_1 \sigma_2 \|\tilde{B}\|_F^2 = -\sigma_1 \sigma_2. \end{aligned}$$

The pairing $\sigma_1 \sigma_2$ of the largest singular values features only as a factor in front of the product $b_{12} b_{21}$. Therefore, the estimate is sharp if all weight in $\tilde{B}$ is placed on these

terms. Consider $\tilde{B} = \left(\begin{array}{cc|c} 0 & b_{12} & \mathbf{0} \\ b_{21} & 0 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right)$ with $b_{12}^2 + b_{21}^2 = 1$. This yields

$$\begin{aligned} \text{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma) &= \text{tr} \left(\begin{pmatrix} 0 & b_{12} \\ b_{21} & 0 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} 0 & b_{12} \\ b_{21} & 0 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \right) \\ &= 2\sigma_1 \sigma_2 b_{12} b_{21}. \end{aligned}$$

The product $b_{12} b_{21}$ gets extremal for $b_{12}, b_{21} = \pm \frac{1}{\sqrt{2}}$, the associated trace minimum is $-\sigma_1 \sigma_2$ and is attained when b_{12} and b_{21} feature opposite signs. The global minimum is isolated, if $\sigma_2 > \sigma_3$.

Under the additional normalization of $\|B_2\|_F = 1$, the global minimum of $-\frac{1}{2}$ is attained, if $\sigma_1 = \sigma_2 = \frac{1}{\sqrt{2}}$ (which enforces $\sigma_3, \dots, \sigma_p = 0$). The global maximum of a value of 1 is attained if $\sigma_1 = 1$ (which enforces $\sigma_2, \dots, \sigma_p = 0$). $\square$

The next theorem includes (and refines) the inequality of Wu and Chen. The proof is different and is based on an optimization approach.

THEOREM 7. *For $B_1, B_2 \in \mathbb{R}^{m \times p}$, with SVDs $B_1 = U_1 P V_1^T \in \mathbb{R}^{m \times p}$, $B_2 = U_2 \Sigma V_2^T \in \mathbb{R}^{m \times p}$, where the upper left diagonal blocks of the singular values matrices P, Σ are $P_p = \text{diag}(\rho_1, \dots, \rho_p)$, and $\Sigma_p = \text{diag}(\sigma_1, \dots, \sigma_p)$, respectively, it holds*

$$(14) \quad \text{tr}(B_1^T B_2 B_2^T B_1) - \text{tr}(B_1^T B_2 B_1^T B_2) \leq \min\{\|B_1\|_F^2(\sigma_1^2 + \sigma_2^2), \|B_2\|_F^2(\rho_1^2 + \rho_2^2)\},$$

$$(15) \quad \text{tr}(B_1 B_2^T B_2 B_1^T) - \text{tr}(B_1^T B_2 B_1^T B_2) \leq \min\{\|B_1\|_F^2(\sigma_1^2 + \sigma_2^2), \|B_2\|_F^2(\rho_1^2 + \rho_2^2)\}.$$

As a consequence,

$$(16) \quad \frac{1}{2}\|B_1^T B_2 - B_2^T B_1\|_F^2 \leq \|B_1\|_F^2 \|B_2\|_F^2, \quad \frac{1}{2}\|B_1 B_2^T - B_2 B_1^T\|_F^2 \leq \|B_1\|_F^2 \|B_2\|_F^2.$$

Note that (16) is the Wu-Chen matrix inequality of [33]. While (14), (15) do not give tighter general bounds, they can be significantly tighter in special situations. For example, if $B_2 \in St(m, p)$ is column-orthogonal, then all singular values of B_2 are 1 and $\|B_2\|_F^2 = \text{tr}(I) = m$. The bound of (14) is $\|B_1^T B_2 - B_2^T B_1\|_F \leq \sqrt{2}\sqrt{2}\|B_1\|_F$, while that of (16) gives $\|B_1^T B_2 - B_2^T B_1\|_F \leq \sqrt{2}\sqrt{m}\|B_1\|_F$, which grows with the dimension m .

Proof. It holds $\frac{1}{2}\|B_1^T B_2 - B_2^T B_1\|_F^2 = \text{tr}(B_1^T B_2 B_2^T B_1) - \text{tr}(B_1^T B_2 B_1^T B_2)$ and $\frac{1}{2}\|B_1 B_2^T - B_2 B_1^T\|_F^2 = \text{tr}(B_1 B_2^T B_2 B_1^T) - \text{tr}(B_1^T B_2 B_1^T B_2)$. We normalize and apply again the coordinate change based on the SVD data of B_2 to obtain

$$\begin{aligned} & \text{tr}(B_1^T B_2 B_2^T B_1) - \text{tr}(B_1^T B_2 B_1^T B_2) \\ &= \|B_1\|_F^2 \text{tr}\left(\frac{B_1^T}{\|B_1\|_F} B_2 B_2^T \frac{B_1}{\|B_1\|_F}\right) - \|B_1\|_F^2 \text{tr}\left(\frac{B_1^T}{\|B_1\|_F} B_2 \frac{B_1^T}{\|B_1\|_F} B_2\right) \\ &= \|B_1\|_F^2 \left(\text{tr}(\tilde{B}^T \Sigma \Sigma^T \tilde{B}) - \text{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma)\right), \quad \tilde{B} = U^T (B_1 / \|B_1\|_F) V, \quad \|\tilde{B}\|_F = 1. \end{aligned}$$

By the preparatory lemmata, it is clear that all weight in the (normalised) matrix $\tilde{B}$ must be concentrated on the upper (2×2) -diagonal block. Therefore, we look for $\tilde{B} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with the optimal balance between the parameters a, b, c, d that maximizes the combination of trace terms. Yet, a quick calculation shows that the diagonal entries a, d do not contribute to the result. We have

$$\text{tr}(\tilde{B}^T \Sigma \Sigma^T \tilde{B}) - \text{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma) = \sigma_1^2 b^2 + \sigma_2^2 c^2 - 2\sigma_1 \sigma_2 b c = (\sigma_1 b - \sigma_2 c)^2.$$

To maximize this term, all weight in $\tilde{B}$ must be on the off-diagonal terms so that $a, d = 0$ and $b^2 + c^2 = 1$. Let $S^1 = \{(b, c) \in \mathbb{R}^2 \mid b^2 + c^2 = 1\}$ be the unit circle, parameterized by $\gamma : t \mapsto (b(t), c(t)) = (\cos(t), \sin(t))$. The function $f : [-\pi, \pi] \rightarrow \mathbb{R}, t \mapsto (\sigma_1 \cos(t) - \sigma_2 \sin(t))^2$ attains its global maximum on S^1 at $t_* = \arctan\left(-\frac{\sigma_2}{\sigma_1}\right)$.

This yields $b_* = \frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}, c_* = \frac{-\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ and a global maximum of

$$(17) \quad (\sigma_1 b_* - \sigma_2 c_*)^2 = (\sigma_1^2 + \sigma_2^2).$$

Combined, it holds

$$\mathrm{tr}(B_1^T B_2 B_2^T B_1) - \mathrm{tr}(B_1^T B_2 B_1^T B_2) \leq \|B_1\|_F^2 (\sigma_1^2 + \sigma_2^2) \leq \|B_1\|_F^2 \|B_2\|_F^2.$$

The roles of B_1 and B_2 can be exchanged. The same reasoning applies to (15) but with the transpose of B_1, B_2 . This yields

$$(18) \quad \mathrm{tr}(\tilde{B} \Sigma^T \Sigma \tilde{B}^T) - \mathrm{tr}(\tilde{B}^T \Sigma \tilde{B}^T \Sigma) = (\sigma_2 b - \sigma_1 c)^2.$$

The global optimum is the same value of $(\sigma_1^2 + \sigma_2^2)$, but is attained for $b_* = \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$, $c_* = \frac{-\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}$, which corresponds to the transpose of the maximizer of (14). Both inequalities become simultaneously sharp for the same input pair $\tilde{B}, \Sigma$, if $\sigma_1 = \sigma_2$. $\square$

3.1. The classical Grassmann sectional curvature bounds. With [Theorem 7](#), the global bound of $K^{Gr} \leq 2$ of the sectional curvature on $Gr(n, p)$ is an immediate consequence of (6), (7) (as was already clear from [33], and clear to Wong in 1968, [32]). The curvature formulas (6), (7) hold for $\{B_1, B_2\}$ forming an ONB. With $B_2 = U \Sigma V^T$, under the transformation $B_1 \mapsto U^T B_1 V =: \tilde{B}_1$, this is equivalent to $\{\tilde{B}_1, \Sigma\}$ being an ONB. According to [Theorem 7](#), the matrices, for which both matrix inequalities are simultaneously sharp are

$$(19) \quad \tilde{B}_1 = \pm \left(\begin{array}{cc|c} 0 & \frac{1}{\sqrt{2}} & \mathbf{0} \\ \frac{-1}{\sqrt{2}} & 0 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right), \text{ and } \tilde{\Sigma} = \left(\begin{array}{cc|c} \sigma & 0 & \mathbf{0} \\ 0 & \sigma & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right).$$

This pair of matrices is orthogonal; it is an ONB if $\sigma = \frac{1}{\sqrt{2}}$. Observe that $\tilde{B}_1^T \Sigma$ is skew-symmetric. The tangent plane spanned by the associated tangent vectors is the only tangent plane of maximum sectional curvature.

3.2. Bounds on the canonical sectional curvature on Stiefel. In this section, we establish a global upper bound on the sectional curvature on $St(n, p)$ under the canonical metric for all Stiefel manifolds $n \geq p$ that are at least two-dimensional. (Otherwise, the concept of sectional curvature does not apply.) We start from (8) for an orthonormal pair of Stiefel tangents

$$X = \begin{pmatrix} A_1 & -B_1^T \\ B_1 & \mathbf{0} \end{pmatrix}, \quad Y = \begin{pmatrix} A_2 & -B_2^T \\ B_2 & \mathbf{0} \end{pmatrix} \in T_{[I]} St(n, p).$$

Orthonormality in the canonical Stiefel metric means that

$$\frac{1}{2} \|A_j\|_F^2 + \|B_j\|_F^2 = 1, \quad j = 1, 2, \quad \frac{1}{2} \mathrm{tr}(A_1^T A_2) + \mathrm{tr}(B_1^T B_2) = 0.$$

From [12, Lemma 2.5], we have norm bounds for the commutator bracket of skew-symmetric matrices,

$$(20) \quad \|[A_1, A_2]\|_F^2 \leq \|A_1\|_F^2 \|A_2\|_F^2, \quad \forall p \geq 4,$$

$$(21) \quad \|[A_1, A_2]\|_F^2 \leq \frac{1}{2} \|A_1\|_F^2 \|A_2\|_F^2, \quad p = 3,$$

$$(22) \quad \|[A_1, A_2]\|_F^2 = 0, \quad p = 2.$$

The last one is obvious, because (2×2) -skew symmetric matrices necessarily commute; (21) is a straightforward consequence of the interplay between skew-symmetric (3×3) -matrices and their representation with 3-vectors, see [Appendix B](#) for a short recap.

Write $\alpha_1 = \|A_1\|_F, \alpha_2 = \|A_2\|_F, \beta_1 = \|B_1\|_F, \beta_2 = \|B_2\|_F$ and note that for normalized tangent vectors $\beta_j^2 = 1 - \frac{1}{2}\alpha_j^2, \beta_j \in [0, 1], \alpha_j \in [0, \sqrt{2}]$.

Because A_1, A_2 are skew-symmetric, eq. (11) of Lemma 3 yields the following refined estimate for the ‘one-fourth’-term in (8)

$$\begin{aligned} \frac{1}{4}\|B_1A_2 - B_2A_1\|_F^2 &\leq \frac{1}{4}\left(\frac{1}{\sqrt{2}}\|B_1\|_F\|A_2\|_F + \frac{1}{\sqrt{2}}\|B_2\|_F\|A_1\|_F\right)^2 \\ &= \frac{1}{8}(\beta_1^2\alpha_2^2 + \beta_2^2\alpha_1^2 + 2\alpha_1\alpha_2\beta_1\beta_2). \end{aligned}$$

With the above inequalities and Theorem 7, a global bound for the Stiefel curvature is

$$\begin{aligned} K_c^{St}(X, Y) &= \frac{1}{2}\|B_2B_1^T - B_1B_2^T\|_F^2 + \frac{1}{4}\|B_1A_2 - B_2A_1\|_F^2 \\ &\quad + \frac{1}{8}\|[A_1, A_2] - (B_1^TB_2 - B_2^TB_1)\|_F^2 \\ (23) \quad &\leq \frac{5}{4} + \frac{5}{16}\alpha_1^2\alpha_2^2 - \frac{1}{2}(\alpha_1^2 + \alpha_2^2) + \frac{1+\sqrt{2}}{4}\alpha_1\alpha_2\sqrt{1-\frac{1}{2}\alpha_1^2}\sqrt{1-\frac{1}{2}\alpha_2^2}. \end{aligned}$$

This function in (α_1, α_2) has an isolated local maximum at $\alpha_1 = 0 = \alpha_2$ with a corresponding function value of $\frac{5}{4}$, which is the global maximum in the admissible range of $(\alpha_1, \alpha_2) \in [0, \sqrt{2}]^2$. For a verification see Appendix A. For most Stiefel manifolds, more precisely, for all Stiefel manifolds with $p \geq 2, n \geq p+2$, this bound is tight. This main result is detailed in the next theorem. For the sake of completeness, the remaining cases are also included, even though they are not new. The cases $p=2, n=3$ and $p=2, n>3$ are explicitly treated in [22, Prop. 6.1] for a parametric family of metrics [16, 35] that include the canonical metric as a special case. Since $St(n, 1) \cong S^{n-1}$, $St(n, n-1) \cong O(n)$ and $St(n, n) \cong SO(n)$, one may argue that $St(n, p), p \geq 2, n \geq p+2$ are the only ‘true’ Stiefel manifolds.

THEOREM 8. *The sectional curvature under the canonical metric on the Stiefel manifold $St(n, p), n \geq p$, is globally bounded by*

$$0 \leq K_c^{St}(X, Y) \leq \frac{5}{4}.$$

1. For $p \geq 2, n \geq p+2$, the bound is sharp. Up to trace-preserving transformations, the maximum curvature is attained only for the tangent plane spanned by $\left\{X = \begin{pmatrix} 0 & -B_1^T \\ B_1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & -B_2^T \\ B_2 & 0 \end{pmatrix}\right\}$, where

$$(24) \quad B_1 = \pm \frac{1}{\sqrt{2}} \left(\begin{array}{cc|c} 0 & 1 & 0 \\ -1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array} \right), \text{ and } B_2 = \pm \frac{1}{\sqrt{2}} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array} \right).$$

2. If $\text{rank}(B_1) = 1$ or $\text{rank}(B_2) = 1$ holds for the B -blocks in X or Y , respectively, then

$$K_c^{St}(X, Y) \leq 1.$$

The bound is attained for the matrices stated in eq. (25) below.

3. For $p=1$ and $n \geq 3$,

$$K_c^{St}(X, Y) \leq 1$$

and the bound is sharp. It is attained for the matrices specified in eq. (25) if these are reduced to their first column.

4. For $(n, p) = (3, 2)$ or $(n, p) = (3, 3)$, the sectional curvature has a constant value of

$$K_c^{St}(X, Y) \equiv \frac{1}{4}.$$

5. For $n \geq 4$ and $p = n - 1$ or $p = n$,

$$K_c^{St}(X, Y) \leq \frac{1}{2}$$

and the bound is sharp.

For $n = 2, p = 2$ or $n = 2, p = 1$, $St(n, p)$ is one-dimensional so that the concept of sectional curvatures does not apply.

Proof. On 1.: The global bound of $\frac{5}{4}$ is already established by the maximum of (23). A direct calculation shows that the bound is attained for the tangent plane associated with the matrices in (24). Hence, the bound is sharp for all Stiefel manifolds that fit these matrices dimension-wise. Moreover, the matrices in (24) are the only matrices (up to trace-preserving transformations) for which the bound is achieved. This can be established analogously to the Grassmann manifold case, since the maximum curvature is attained for zero skew-symmetric blocks $A_j = 0$.

On 2.: If one of the matrices B_1 or B_2 is of rank one, then from the proof of Theorem 7 one can deduce that

$$\frac{1}{2}\|B_2 B_1^T - B_1 B_2^T\|_F^2 + \frac{1}{2}\|B_1^T B_2 - B_2^T B_1\|_F^2 \leq \|B_1\|_F^2 \|B_2\|_F^2.$$

Compared to the general case, the bound is improved by a factor of 2. As before, the maximum curvature is attained for zero skew-symmetric blocks $A_j = 0$. Hence, the curvature is bounded by

$$\begin{aligned} K_c^{St}(X, Y) &\leq \frac{1}{2}\|B_2 B_1^T - B_1 B_2^T\|_F^2 + \frac{1}{8}\|(B_1^T B_2 - B_2^T B_1)\|_F^2 \\ &= \frac{3}{8}\|B_2 B_1^T - B_1 B_2^T\|_F^2 \\ &\quad + \frac{1}{4}\left(\frac{1}{2}\|B_2 B_1^T - B_1 B_2^T\|_F^2 + \frac{1}{2}\|(B_1^T B_2 - B_2^T B_1)\|_F^2\right) \\ &\leq \frac{3}{4}\|B_1\|_F^2 \|B_2\|_F^2 + \frac{1}{4}\|B_1\|_F^2 \|B_2\|_F^2 \leq 1. \end{aligned}$$

The rank-one bound is attained for

$$(25) \quad B_1 = \pm \left(\begin{array}{cc|c} 0 & 0 & \mathbf{0} \\ 1 & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right), \text{ and } B_2 = \pm \left(\begin{array}{cc|c} 1 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right).$$

The matrix B_1 may feature multiple copies of the canonical unit vector e_1^T as rows.

On 3.: This is a direct consequence of item 2.

On 4.: See [22, Prop. 6.1] or see Appendix B for a direct proof.

On 5.: This is clear because $St(n, n-1) \cong O(n)$, $St(n, n) \cong SO(n)$. In both these cases, $[X, Y]_{\mathfrak{m}} = [X, Y]$, $[X, Y]_{\mathfrak{h}} = 0$ so that the curvature formula (3) reduces to (2). For $O(n)$ and $SO(n)$, the formula from Remark 1 applies. The global bound of $\frac{1}{2}$ stems from (20). $\square$

Remark 9. The bound of [Theorem 8](#) was already correctly guessed in [\[27, p. 94, 95\]](#), but the derivation there is not correct. It was overlooked that tangent matrices X of unit norm w.r.t. the Riemannian metric $1 = \|X\|^2 = \frac{1}{2}\|X\|_F^2$ have a Frobenius norm of $\|X\|_F = \sqrt{2}$. In [\[22, Prop. 6.2\]](#) it is shown that the sectional curvature range of the Stiefel manifold under the canonical metric *includes* the interval $[0, \frac{5}{4}]$. But to argue that this is the exact range, the author resorted to the result of [\[27, p. 94, 95\]](#), which was lacking a valid proof.

3.3. Bounds on the Euclidean sectional curvature on Stiefel. Besides the canonical metric that follows from equipping the orthogonal group with the Euclidean metric $\langle X, Y \rangle_Q = \frac{1}{2}\langle X, Y \rangle_F$ and the quotient space representation $St(n, p) = O(n)/O(n-p)$, the most important metric for applications is obtained from considering the Stiefel manifold as an embedded submanifold of the Euclidean matrix space, $St(n, p) \subset \mathbb{R}^{n \times p}$. Tangent vectors are then rectangular matrices $\Delta = Q \begin{pmatrix} A \\ B \end{pmatrix} = UA + U_\perp B \in \mathbb{R}^{n \times p}$, where $Q = \begin{pmatrix} U & U_\perp \end{pmatrix} \in O(n)$, $A \in \text{skew}(p)$, $B \in \mathbb{R}^{(n-p) \times p}$. The natural Riemannian metric in this setting is obtained from restricting the Euclidean inner product² to the Stiefel tangent spaces. This yields the so-called *Euclidean metric* on $St(n, p) \subset \mathbb{R}^{n \times p}$,

$$\langle \Delta, \tilde{\Delta} \rangle_e^{St} = \text{tr} \left(\begin{pmatrix} A^T & B^T \end{pmatrix} Q^T Q \begin{pmatrix} \tilde{A} \\ \tilde{B} \end{pmatrix} \right) = \text{tr}(A^T \tilde{A}) + \text{tr}(B^T \tilde{B}).$$

The differences between this metric and the canonical one have been discussed in [\[9, p. 313\]](#). Most notably, this metric is not directly related to a quotient space construction so that the formulas of [Section 2](#) do not readily apply. One option for calculating curvatures is via direct computations and the geometry of submanifolds, see [\[19, Section 8\]](#). Yet, the canonical and the Euclidean metric are included in a parametric family of metrics considered in [\[16, Def. 3.1, Remark 1\]](#), where the authors show that all these metrics stem from a bi-invariant metric associated with a different quotient space representation, namely $St(n, p) \cong (O(n) \times O(p))/(O(n-p) \times O(p))$. The associated sectional curvature is computed in [\[22\]](#). We omit the details and quote only the final result for the Stiefel manifold under the Euclidean metric: For a tangent plane section spanned by an ONB $\{X = Q \begin{pmatrix} A_1 \\ B_1 \end{pmatrix}, Y = \begin{pmatrix} A_2 \\ B_2 \end{pmatrix}\}$,

$$\begin{aligned} K_e^{St}(X, Y) &= \|B_1 A_2 - B_2 A_1\|_F^2 + \frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 - \frac{1}{2} \|B_1^T B_2 - B_2^T B_1\|_F^2 \\ (26) \quad &+ \frac{1}{4} \|[A_1, A_2] - (B_2^T B_1 - B_1^T B_2)\|_F^2, \end{aligned}$$

see [\[22, Prop 4.2, eq. \(34\)\]](#) with $\alpha = 1$. In the same work, it is shown that the interval $[-\frac{1}{2}, 1]$ is contained in the range of the Euclidean curvature and it is conjectured that $-\frac{1}{2} \leq K_e^{St}(X, Y) \leq 1$. We confirm this conjecture. The Euclidean sectional curvature formula shares many terms with the canonical one from [\(8\)](#) but with different factors. The occurrence of a negative contribution is annoying in view of estimates. As before, we restrict our considerations to plane sections spanned by ONBs. Shifting the base point has no impact on the curvature. Hence, w.l.o.g. $Q = I$, $\{X = \begin{pmatrix} A_1 \\ B_1 \end{pmatrix}, Y = \begin{pmatrix} A_2 \\ B_2 \end{pmatrix}\}$

²Caution: Here *without* the factor $\frac{1}{2}$!

$\begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \} \subset T_{[I]} St(n, p)$. The matrices X, Y form a (Riemannian) ONB if and only if

$$1 = \|A_i\|_F^2 + \|B_i\|_F^2, \quad i = 1, 2, \quad 0 = \text{tr}(A_1^T A_2) + \text{tr}(B_1^T B_2).$$

The lower bound on the Euclidean Stiefel curvature. For an ONB X, Y consisting of subblocks A_i, B_i as above, let $\alpha_i = \|A_i\|_F, \beta_i = \|B_i\|_F, i = 1, 2$ so that $\alpha_i^2 + \beta_i^2 = 1$. By the reversed triangle inequality,

$$\begin{aligned} K_e^{St}(X, Y) &\geq \|B_1 A_2 - B_2 A_1\|_F^2 + \frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 - \frac{1}{2} \|B_1^T B_2 - B_2^T B_1\|_F^2 \\ &\quad + \frac{1}{4} \left(\| [A_1, A_2] \|_F^2 + \| B_2^T B_1 - B_1^T B_2 \|_F^2 \right. \\ &\quad \left. - 2 \| [A_1, A_2] \|_F \| B_2^T B_1 - B_1^T B_2 \|_F \right) \\ &= \|B_1 A_2 - B_2 A_1\|_F^2 + \frac{1}{4} \| [A_1, A_2] \|_F^2 - \frac{1}{2} \| [A_1, A_2] \|_F \| B_1^T B_2 - B_2^T B_1 \|_F \\ &\quad + \frac{1}{2} \| B_1 B_2^T - B_2 B_1^T \|_F^2 - \frac{1}{4} \| B_1^T B_2 - B_2^T B_1 \|_F^2 \\ &\geq -\frac{1}{2} \| [A_1, A_2] \|_F \| B_1^T B_2 - B_2^T B_1 \|_F - \frac{1}{4} \| B_1^T B_2 - B_2^T B_1 \|_F^2 \\ &\geq -\frac{1}{2} \alpha_1 \alpha_2 \sqrt{2} \beta_1 \beta_2 - \frac{1}{2} \beta_1^2 \beta_2^2 \\ &= -\frac{1}{2} \left(\beta_1 \beta_2 \left(\sqrt{2} \sqrt{1 - \beta_1^2} \sqrt{1 - \beta_2^2} + \beta_1 \beta_2 \right) \right). \end{aligned}$$

where we have used the Wu-Chen matrix inequality (16). The minimum of this function in the admissible range of $(\beta_1, \beta_2) \in [0, 1]^2$ is attained at the corner point $\beta_1 = 1 = \beta_2$, so that

$$K_e^{St}(X, Y) \geq -\frac{1}{2}.$$

As in the case of the canonical metric, the extrema are attained for tangent vectors with zero A -blocks. For a section spanned by such an ONB, the curvature formula reduces to

$$(27) \quad K_e^{St}(X, Y) = \frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 - \frac{1}{4} \|B_1^T B_2 - B_2^T B_1\|_F^2.$$

The lower bound is attained for a normalized, orthogonal matrix pair B_1, B_2 such that the first positive term vanishes, while the second negative term becomes extremal. A corresponding matrix pair can be read off from (17) and (18), when choosing $\sigma_1 = 1 = b_*$ and $\sigma_2 = 0 = c_*$. The associated matrices that span an extremal tangent plane section are $X = \begin{pmatrix} \mathbf{0} \\ B_1 \end{pmatrix}, Y = \begin{pmatrix} \mathbf{0} \\ B_2 \end{pmatrix}$, where

$$(28) \quad B_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

For these matrices, $\frac{1}{2} \|B_1^T B_2 - B_2^T B_1\|_F^2 = 1$, while $\frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 = 0$. Note that this matrix pair is a maximizer of the trace term in Lemma 4, item (1).

The upper bound on the Euclidean Stiefel curvature. Continuing the above notation and introducing $\gamma = \|B_1^T B_2 - B_2^T B_1\|_F \in [0, \sqrt{2}\beta_1\beta_2]$, we obtain for the upper bound on the Euclidean Stiefel curvature

$$\begin{aligned}
K_e^{St}(X, Y) &\leq \|B_1 A_2 - B_2 A_1\|_F^2 + \frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 - \frac{1}{2} \|B_1^T B_2 - B_2^T B_1\|_F^2 \\
&\quad + \frac{1}{4} \left(\| [A_1, A_2] \|_F^2 + \|B_1^T B_2 - B_2^T B_1\|_F^2 \right. \\
&\quad \left. + 2 \| [A_1, A_2] \|_F \|B_1^T B_2 - B_2^T B_1\|_F \right) \\
&\leq \frac{1}{2} \alpha_2^2 \beta_1^2 + \frac{1}{2} \alpha_1^2 \beta_2^2 + \alpha_1 \alpha_2 \beta_1 \beta_2 + \beta_1^2 \beta_2^2 - \frac{1}{2} \gamma^2 \\
&\quad + \frac{1}{4} (\alpha_1^2 \alpha_2^2 + \gamma^2 + 2\alpha_1 \alpha_2 \gamma) \\
&= \frac{1}{2} (1 - \beta_2^2) \beta_1^2 + \frac{1}{2} (1 - \beta_1^2) \beta_2^2 + \sqrt{1 - \beta_1^2} \sqrt{1 - \beta_2^2} \beta_1 \beta_2 + \beta_1^2 \beta_2^2 \\
(29) \quad &\quad - \frac{1}{2} \gamma^2 + \frac{1}{4} \left((1 - \beta_1^2)(1 - \beta_2^2) + \gamma^2 + 2\gamma \sqrt{1 - \beta_1^2} \sqrt{1 - \beta_2^2} \right).
\end{aligned}$$

When considering all other parameters as fixed, the last expression is a parabola in γ ; the parametric maximum is $\gamma_* = \gamma_*(\beta_1, \beta_2) = \sqrt{1 - \beta_1^2} \sqrt{1 - \beta_2^2}$. Inserting this γ in (29) yields

$$\begin{aligned}
K_e^{St}(X, Y) &\leq \frac{1}{2} (1 - \beta_2^2) \beta_1^2 + \frac{1}{2} (1 - \beta_1^2) \beta_2^2 + \sqrt{1 - \beta_1^2} \sqrt{1 - \beta_2^2} \beta_1 \beta_2 + \beta_1^2 \beta_2^2 \\
&\quad + \frac{1}{2} (1 - \beta_1^2)(1 - \beta_2^2).
\end{aligned}$$

This expression has its global maximum at $\beta_1 = 1 = \beta_2$, which can be verified by elementary means similar to the discussion of (23). We omit the details. Again, the associated matrix maximizers feature a zero A -block. Reconsidering (27), we can utilize (17) and (18) to find a maximizing matrix pair. This time one needs to choose $\sigma_1 = 1 = c_*$ and $\sigma_2 = 0 = b_*$. (Mind the "transposition duality" that has been mentioned before.) The associated matrices that span an extremal tangent plane section are $X = \begin{pmatrix} \mathbf{0} \\ B_1 \end{pmatrix}, Y = \begin{pmatrix} \mathbf{0} \\ B_2 \end{pmatrix}$, where

$$(30) \quad B_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

For these matrices, $\frac{1}{2} \|B_1^T B_2 - B_2^T B_1\|_F^2 = 0$, while $\frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 = 1$. Note that this matrix pair is a maximizer of the trace term in Lemma 4, item (2). Unlike the canonical metric, the Euclidean metric achieves the cases of extreme curvature for tangent plane sections spanned by rank-1 matrices.

The next theorem summarizes the findings, extended by the cases of special dimensions.

THEOREM 10. *The sectional curvature under the Euclidean metric on the Stiefel manifold $St(n, p)$, $n \geq p$, is globally bounded by*

$$-\frac{1}{2} \leq K_e^{St}(X, Y) \leq 1.$$

1. For $p \geq 2, n \geq p + 2$, the bound is sharp. Up to trace-preserving transformations, the maximum and minimum curvature is attained for the tangent plane associated with the matrices of (30) and (28), respectively.
2. For $p = 1$ and $n \geq 3$, the sectional curvature has a constant value of

$$K_e^{St}(X, Y) \equiv 1.$$

3. For $p = n \geq 4$, it holds

$$0 \leq K_e^{St}(X, Y) \leq \frac{1}{4}.$$

4. For $p = n = 3$, it holds

$$K_e^{St}(X, Y) \equiv \frac{1}{8}.$$

5. For $(n, p) = (3, 2)$ the sectional curvature is bounded by

$$-\frac{1}{2} \leq K_e^{St}(X, Y) \leq \frac{1}{2}.$$

The bounds are sharp. The lower bound is sharp and is attained for the matrices from (28), but reduced to their first row. The upper bound is attained, e.g., for

$$X = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ -1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 \end{pmatrix}.$$

6. For $n \geq 4$ and $(n, p) = (n, n - 1)$ the sectional curvature is bounded by

$$-\frac{1}{2} \leq K_e^{St}(X, Y) \leq \frac{2}{3}.$$

The lower bound is sharp and is attained for the matrices from (28), but reduced to their first row.

Conjecture 1. For $n \geq 4$ and $(n, p) = (n, n - 1)$, sharp bounds on the sectional curvature of the Stiefel manifold under the Euclidean metric are

$$-\frac{1}{2} \leq K_e^{St}(X, Y) \leq \frac{1}{2}.$$

Proof. (of Theorem 10) The global bounds have already been established by considerations that are independent of the dimensions. However, the bounds are not sharp in all dimensions.

On 1.: In the cases, $p \geq 2, n \geq p + 2$, both extremal pairs from (30) and (28) fit into the tangent space and thus provide examples in which the global lower and upper bounds are reached.

On 2.: For $p = 1$ and $n \geq 3$, the A -block in a tangent vector is a skew-symmetric (1×1) -matrix, i.e., $A = \mathbf{0}$. Moreover, the B -blocks are $((n - 1) \times 1)$ -column vectors. Hence, $\|B_1^T B_2 - B_2^T B_1\|_F = 0$. The curvature formula reduces to

$$K_e^{St}(X, Y) = \frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2.$$

Because X, Y form an ONB, we have in this case $B_i^T B_i = 1$ and $B_1^T B_2 = 0$. As a consequence, $\frac{1}{2} \|B_1 B_2^T - B_2 B_1^T\|_F^2 = \frac{1}{2} (2\|B_1\|_2^2 \|B_2\|_2^2 - 2\langle B_1, B_2 \rangle_2^2) = 1$.

On 3.: For $p = n \geq 4$, the B -blocks of the spanning tangent vectors vanish. The curvature formula becomes

$$K_e^{St}(X, Y) = \frac{1}{4} \| [A_1, A_2] \|_F^2.$$

Now, the upper bound of $\frac{1}{4}$ is from (20), while the lower bound of 0 is obvious. The upper bound is attained, e.g., for the (4×4) skew-symmetric matrices A_1, A_2 listed at the beginning of Subsection 4.3. For a full discussion of the sharpness of the bound, see [12]. The lower bound is attained for any orthogonal, commuting pair of skew-symmetric matrices, e.g.,

$$A_1 = \frac{1}{\sqrt{2}} \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right), \quad A_2 = \frac{1}{\sqrt{2}} \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{array} \right).$$

Of course, the curvature is the same for all matrices of higher dimensions that feature the above examples as sub-blocks and are otherwise filled up with zeros.

On 4.: This is in analogy to item 4. of Theorem 8. The argument is the same as outlined in Appendix B.

On 5.: For $(n, p) = (3, 2)$, the matrix blocks $A_1, A_2 \in \text{skew}(2)$ and $B_1, B_2 \in \mathbb{R}^{1 \times 2}$ can be written as

$$A_1 = \alpha \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad A_2 = a \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad B_1 = (\beta_1 \quad \beta_2), \quad B_2 = (b_1 \quad b_2),$$

for $\alpha, \beta_1, \beta_2, a, b_1, b_2 \in \mathbb{R}$. The curvature formula reduces to

$$K_e^{St}(X, Y) = \|B_1 A_2 - B_2 A_1\|_F^2 - \frac{1}{4} \|B_1^T B_2 - B_2^T B_1\|_F^2$$

and the orthogonality constraints

$$1 = \|A_i\|_F^2 + \|B_i\|_F^2, \quad i = 1, 2, \quad 0 = \text{tr}(A_1^T A_2) + \text{tr}(B_1^T B_2)$$

translate to

$$\alpha^2 = \frac{1}{2} - \frac{1}{2}(\beta_1^2 + \beta_2^2), \quad a^2 = \frac{1}{2} - \frac{1}{2}(b_1^2 + b_2^2), \quad \beta_1 b_1 + \beta_2 b_2 = -2\alpha a.$$

By squaring both sides of the equation $\beta_1 b_1 + \beta_2 b_2 = -2\alpha a$ and exploiting the equations for α^2 and a^2 , we obtain

$$(\beta_1^2 + \beta_2^2) + (b_1^2 + b_2^2) - 1 = (\beta_1 b_2 - \beta_2 b_1)^2 \geq 0$$

and therefore $(\beta_1^2 + \beta_2^2) + (b_1^2 + b_2^2) \geq 1$. By exploiting the equations resulting from the orthogonality constraints once again, the bound on the curvature is obtained:

$$K_e^{St}(X, Y) = \frac{3}{2} - ((\beta_1^2 + \beta_2^2) + (b_1^2 + b_2^2)) \leq \frac{1}{2}.$$

On the other hand, $-\frac{1}{2} \leq K_e^{St}(X, Y)$ applies because $(\beta_1^2 + \beta_2^2) + (b_1^2 + b_2^2) \leq 2$.

For an alternative, but more involved argument, see [22, Prop. 6.1], which gives tight bounds in the case $p = 2$ for a family of metrics including the Euclidean and the

canonical one.

On 6.: For $p = n - 1$, the matrix blocks $B_1, B_2 \in \mathbb{R}^{(n-(n-1)) \times (n-1)}$ are row vectors in $\mathbb{R}^{1 \times (n-1)}$. The global lower curvature bound of $K_e \geq -\frac{1}{2}$ continues to hold and the sharpness of the bound is confirmed by the example of the tangent space section associated with the matrix blocks from (28), which fit into this setting dimension-wise, when reduced to their first row and possible filled up with zeros. By taking $\|B_1 B_2^T - B_2 B_1^T\|_F^2 = 0$ into account, the curvature formula reads

$$\begin{aligned} K_e^{St}(X, Y) &= \|B_1 A_2 - B_2 A_1\|_F^2 - \frac{1}{2} \|B_1^T B_2 - B_2^T B_1\|_F^2 \\ &\quad + \frac{1}{4} \|[A_1, A_2] - (B_2^T B_1 - B_1^T B_2)\|_F^2. \end{aligned}$$

The upper bound of $K_e \leq \frac{2}{3}$ can be established in the same manner as before. We omit the details, on the one hand because this is straightforward, on the other hand, because we do not believe that this bound is sharp anyways. $\square$

3.4. Impact on the injectivity radius of the Stiefel manifold. On a Riemannian manifold $\mathcal{M}$, the injectivity radius is the largest possible radius within which geodesics are unique and lengths-minimizing, regardless of where you start from. In loose words, as long as you stay within the injectivity radius when travelling along a geodesic, you are guaranteed not to travel unnecessary distances. This concept is formalized with introducing the Riemannian exponential map at a point $p \in \mathcal{M}$, which sends a tangent vector $v \in T_p \mathcal{M}$, to the endpoint of the geodesic on the unit interval that starts from p with velocity v . For a precise definition of the notion of the injectivity radius, we refer the reader to [8, Chap. 13].

A classical result from Riemannian geometry relates the injectivity radius with the sectional curvature:

THEOREM 11 (Klingenberg, stated as Lemma 6.4.7 in [25]). *Let $\mathcal{M}$ be a compact Riemannian manifold with sectional curvatures bounded by $K(X, Y) \leq C$, where $C > 0$. Then the injectivity radius $\text{inj}(p)$ at any $p \in \mathcal{M}$ satisfies*

$$\text{inj}(p) \geq \min \left\{ \frac{\pi}{\sqrt{C}}, \frac{1}{2} l_p \right\},$$

where l_p is the length of a shortest closed geodesic starting from p . For the global injectivity radius, it holds

$$\text{inj}(\mathcal{M}) \geq \frac{\pi}{\sqrt{C}} \quad \text{or} \quad \text{inj}(\mathcal{M}) = \frac{1}{2} l,$$

where l is the length of a shortest closed geodesic on $\mathcal{M}$.

From Theorem 8 and Theorem 10, one obtains for $p \geq 2, n \geq p + 2$.

1. $\text{inj}(St(n, p)) \geq \sqrt{\frac{4}{5}} \pi$ (canonical metric).
2. $\text{inj}(St(n, p)) = \pi$ unless there is a closed geodesic of length strictly smaller than 2π (Euclidean metric).

Under the canonical metric, it is clear that closed geodesics have a length of at least 2π , see [27, Chapter 5] or the recent preprints [3], [29]. A closed geodesic on the unit interval $[0, 1]$ of length 2π on the Stiefel manifold $St(4, 2)$ under the Euclidean metric is

$$c(t) = \exp_m \left(t \begin{pmatrix} 2A & -B^T \\ B & \mathbf{0} \end{pmatrix} \right) \begin{pmatrix} \exp_m(-tA) \\ \mathbf{0} \end{pmatrix}, \quad A = \mathbf{0} \in \text{skew}(2), \quad B = \begin{pmatrix} 2\pi & 0 \\ 0 & 0 \end{pmatrix}.$$

The form of this geodesic follows the representation from [35, Prop. 1]. Note that $c(0) = c(1) = \begin{pmatrix} I_2 \\ \mathbf{0} \end{pmatrix}$ and that the length of $c|_{[0,1]}$ is $l = 2\pi$. Of course, this geodesic can be embedded in higher dimensions of $p \geq 2, n \geq p + 2$. Examples of closed geodesics and their length with respect to the metrics of the family from [16] are given in [3, Section 6]. We believe that the above example is a shortest closed geodesic under the Euclidean metric. A more detailed discussion is beyond the scope of this paper, but will be provided in [29]. Closely related investigations can be found in [3].

4. Numerical Experiments. In this section, we illustrate the behavior of the sectional curvature on $SO(n), Gr(n, p)$ and $St(n, p)$ at special parametric sections, where the rank of the spanning matrices increase with the parameter. We also investigate the behavior of the generic sectional curvature on these manifolds when increasing the dimension p . For $St(n, p)$ both the canonical and the Euclidean metric feature in the experiments.

4.1. Experiment 1. In the first experiment, we start with the rank-one matrices

$$B_1 = \left(\begin{array}{cc|c} 0 & 1 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ \hline \mathbf{0} & & \mathbf{0} \end{array} \right) \text{ and } B_2 = \left(\begin{array}{cc|c} 1 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ \hline \mathbf{0} & & \mathbf{0} \end{array} \right) \in \mathbb{R}^{10 \times 10}.$$

Then we fill the remaining diagonal entries of B_2 and the super- and sub-diagonal entries of B_1 one after the other in the following way

$$B_1(u) = \left(\begin{array}{cc|cc|cc} 0 & 1 & & & & \\ -u_2 & 0 & & & & \\ \hline & & \ddots & & & \\ & & & 0 & u_9 & \\ & & & -u_{10} & 0 & \end{array} \right) \text{ and } B_2(u) = \begin{pmatrix} 1 & & & & \\ & u_2 & & & \\ & & \ddots & & \\ & & & & u_{10} \end{pmatrix} \in \mathbb{R}^{10 \times 10}.$$

In the beginning, $u_2 = u_3 = \dots = u_{10} = 0$. Then, we first increase $u_2 \in [0, 1]$ linearly until the upper bound 1 is reached. Then, we keep $u_2 = 1$ and let u_3 run through $[0, 1]$ and so on. For each matrix pair $B_1(u), B_2(u)$ under this procedure, we compute the tangent vectors

$$(31) \quad X(u) = \begin{pmatrix} 0 & -B_1(u)^T \\ B_1(u) & 0 \end{pmatrix}, \quad Y(u) = \begin{pmatrix} 0 & -B_2(u)^T \\ B_2(u) & 0 \end{pmatrix}$$

and the sectional curvatures associated with the tangent planes spanned by $X(u), Y(u)$ for the special orthogonal group, the Stiefel manifold and the Grassmann manifold,

$$K^{SO}(X(u), Y(u)), \quad K_c^{St}(X(u), Y(u)), \quad K_e^{St}(X(u), Y(u)), \quad K^{Gr}(X(u), Y(u)).$$

It is understood that the matrices $X(u), Y(u)$ are normalized according to the metric of the respective manifold. To catch both the maximal and the minimal case of the Stiefel curvature under the Euclidean metric, the experiment is performed twice. Once with $B_1(u)$ as above and once with $B_1^T(u)$. The results are displayed in Figure 1. The figure confirms that for the matrices under consideration, the global curvature maximum on all manifolds under consideration is reached for tangent planes spanned by X, Y that feature zero skew-symmetric A -blocks. For $SO(n), Gr(n, p)$ and $St(n, p)$ under the canonical metric, the maximum sectional curvature occurs for the rank-two subblocks B_1, B_2 from (24). Under the Euclidean metric the curvature maximum and minimum occur for the rank-one matrices B_1, B_2 from (30) and (28), respectively.

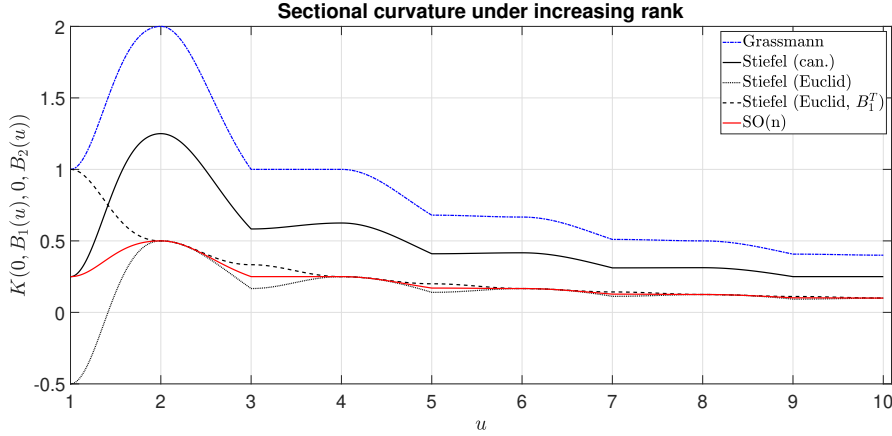


FIG. 1. (Corresponding to Subsection 4.1.) Sectional curvature on $SO(20)$, $St(20,10)$, $Gr(20,10)$ for the tangent sections defined by (31). For $SO(20)$, $Gr(20,10)$, and $St(20,10)$ under the canonical metric, the respective global sectional curvature maximum is attained for the matrices with subblocks B_1 , B_2 as stated in Theorem 8. For $St(n,p)$ under the Euclidean metric, the extreme curvature cases occur for the rank-one matrices from (30) (max) and (28) (min).

4.2. Experiment 2: Curvature associated with pseudo-random sections. Next, we run an experiment with random-matrices of increasing dimension. To this end, we create random skew-symmetric matrices

$$X(A_1, B_1, C_1) = \begin{pmatrix} A_1 & -B_1^T \\ B_1 & C_1 \end{pmatrix}, \quad Y(A_2, B_2, C_2) = \begin{pmatrix} A_2 & -B_2^T \\ B_2 & C_2 \end{pmatrix} \in \text{skew}(2p).$$

We start from $p = 2$. For each $p = 2, \dots, 1000$, a number of 100 random pairs $X, Y \in \text{skew}(2p)$ are computed and the sectional curvature of

- the tangent plane spanned by $X, Y \in T_I SO(2p)$,
- the tangent plane spanned by $X_{\text{m}st} = X(A_1, B_1, \mathbf{0}), Y_{\text{m}st} = Y(A_2, B_2, \mathbf{0}) \in T_{[I]} St(2p, p)$ (canonical metric),
- the tangent plane spanned by $X_{\text{eucl}} = (A_1^T, B_1^T)^T, Y_{\text{eucl}} = (A_2^T, B_2^T)^T \in T_{(I,0)^T} St(2p, p)$ (Euclidean metric),
- and the tangent plane spanned by $X_{\text{m}Gr} = X(\mathbf{0}, B_1, \mathbf{0}), Y_{\text{m}Gr} = Y(\mathbf{0}, B_2, \mathbf{0}) \in T_{[I]} Gr(2p, p)$,

respectively, is computed. The result is then averaged over the number of 100 random runs. Figure 2 displays the sectional curvature versus the block dimension p on a logarithmic scale. It can be seen that the curvature of these "random" planes is largest for $p = 2$ and drops considerably, when p is increased. To be precise, for the canonical Stiefel data, the averaged random curvature at $p = 2$ is $K = 0.33$, while it is $K = 8.8 \cdot 10^{-3}$ for $p = 100$ and $K = 8.9 \cdot 10^{-4}$ for $p = 1000$.

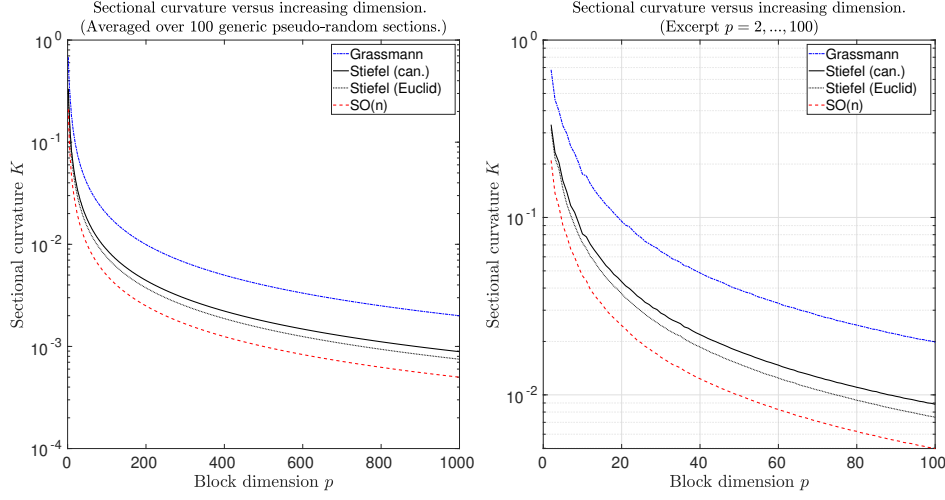


FIG. 2. (Corresponding to [Subsection 4.2.](#)) Averaged sectional curvature for random tangent sections $X = X(A_1, B_1, C_1), Y = Y(A_2, B_2, C_2) \in \text{skew}(2p)$. The dimension of the sub-blocks is p . For Grassmann and Stiefel (canonical metric), X, Y are projected onto the respective horizontal space. For Stiefel under the Euclidean metric, $X = \begin{pmatrix} A_1 \\ B_1 \end{pmatrix}, Y = \begin{pmatrix} A_2 \\ B_2 \end{pmatrix}$ are formed. In all cases, the tangent vectors are orthonormalized according to the respective metric.

4.3. Experiment 3: Impact of the blocks A and B. Consider the special matrices

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

$$B_1(u, v) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & u \\ 0 & 0 & -v & 0 \end{pmatrix}, \quad B_2(u, v) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & u & 0 \\ 0 & 0 & 0 & v \end{pmatrix}$$

The matrix pair A_1, A_2 maximizes the commutator norm $\|[A_1, A_2]\|_F$. For $(u, v) = (0, 0)$, the matrix pair B_1, B_2 makes the Wu-Chen inequality sharp. Let

$$K_c^{St}(A_1, B_1, A_2, B_2) := K_c^{St}\left(\begin{pmatrix} A_1 & -B_1^T \\ B_1 & \mathbf{0} \end{pmatrix}, \begin{pmatrix} A_2 & -B_2^T \\ B_2 & \mathbf{0} \end{pmatrix}\right),$$

$$K_e^{St}(A_1, B_1, A_2, B_2) := K_e^{St}\left(\begin{pmatrix} A_1 \\ B_1 \end{pmatrix}, \begin{pmatrix} A_2 \\ B_2 \end{pmatrix}\right).$$

Figure 3 displays the function

$$[0, 1]^2 \rightarrow \mathbb{R}, (u, v) \mapsto K_m^{St}(\mathbf{0}, B_1(u, v), \mathbf{0}, B_2(u, v)), \quad m \in \{c, e\}$$

under the canonical metric and the Euclidean metric. In both cases, the curvature decreases, when the lower subblocks B_1, B_2 get filled up. Figure 4 displays the function

$$[0, 1] \rightarrow \mathbb{R}, u \mapsto K_m^{St}(uA_1, (1-u)B_1(0, 0), uA_2, (1-u)B_2(0, 0)), \quad m \in \{c, e\}$$

under the canonical metric and the Euclidean metric. The figure illustrates that the sectional curvature decreases in both cases, when the weight in the normalized tangent vectors is shifted from the B -blocks to the A -blocks.

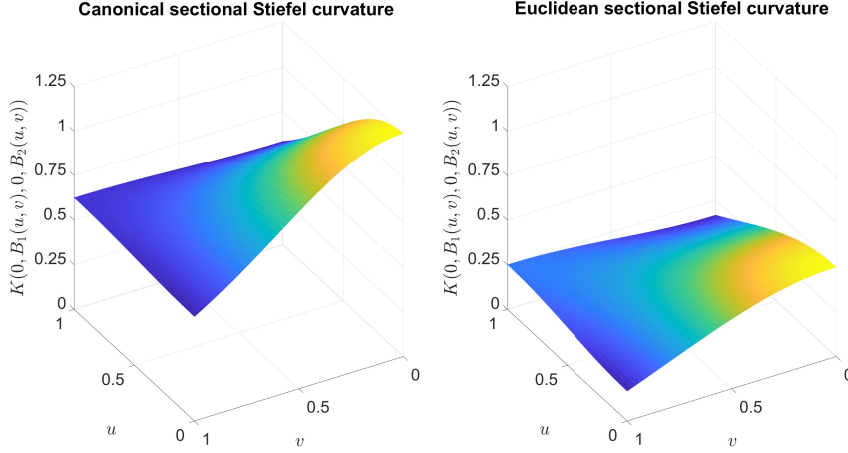


FIG. 3. Sectional curvature on Stiefel for the tangent sections spanned by the matrix blocks of Subsection 4.3. Left: canonical metric, Right: Euclidean metric.

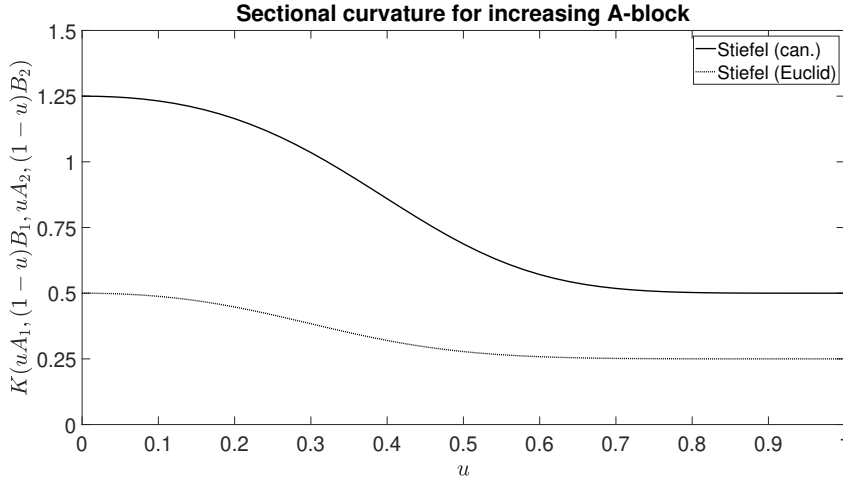


FIG. 4. Sectional curvature on Stiefel when the weight in the spanning normalized tangent matrices is shifted from the B -blocks to the A -blocks.

5. Summary. This paper resumes the investigation of the sectional curvature on the Stiefel and Grassmann manifolds. In the Grassmann case, tight bounds have been known since the work of Wong [32] from 1968. We pay special attention to the maximizers of the curvature bounds and provide refined matrix trace inequalities for this purpose.

An extensive study of the sectional curvature on the Stiefel manifold equipped with a parametric family of Riemannian metrics has been conducted in [22]. However, since the formulae apply to all members of the family, it can be tedious to extract specific formulae for a particular metric. Moreover, tight bounds were only obtained for Stiefel manifolds $St(n, p)$ with $p = 2$.

Under the canonical metric, which is a member of the parametric family, we confirm prior conjectures regarding the sectional curvature. Specifically, we establish that the curvature on the Stiefel manifold equipped with this metric globally does not exceed $5/4$. With this addition, we now have a complete account of the curvature bounds in all admissible dimensions.

Under the Euclidean metric, we prove sharp global bounds on the sectional curvature in all admissible dimensions with one single exception: For the special case $p = n - 1$ and $n \geq 4$, we could only establish a sharp lower bound of $K_e \geq -1/2$ and an upper bound of $K_e \leq 2/3$. We believe that the true upper bound is at $K_e \leq 1/2$. We share this belief with the author of [22].

We also show that the sections that maximise the Grassmann curvature are exactly those for which the Stiefel curvature is maximised under the canonical metric, and that these tangent space sections are necessarily spanned by special rank-two matrices. Under the Euclidean metric, the extreme cases occur for tangent space sections spanned by special rank-one matrices. This supports the observation that 'high curvature means low-rank', which is illustrated by numerical experiments that reveal a decrease in curvature with increasing the rank.

Acknowledgements. The authors would like to thank Prof. P.-A. Absil, University of Louvain, for stimulating conversations on the subject.

Appendix A. The global maximum of the curvature bound of (23). We formally verify that the function in (23) that bounds the sectional curvature of the Stiefel manifold

$$(32) \quad \tilde{f} : (\alpha_1, \alpha_2) \mapsto \frac{5}{4} + \frac{5}{16}\alpha_1^2\alpha_2^2 - \frac{1}{2}(\alpha_1^2 + \alpha_2^2) + \frac{1+\sqrt{2}}{4}\alpha_1\alpha_2\sqrt{1-\frac{1}{2}\alpha_1^2}\sqrt{1-\frac{1}{2}\alpha_2^2}$$

has its global maximum at $(0, 0)$ for $(\alpha_1, \alpha_2) \in [0, \sqrt{2}]^2$.

Consider the transformation $\alpha : [0, \frac{\pi}{2}] \rightarrow [0, \sqrt{2}]$, $r \mapsto \alpha(r) = \sqrt{2}\sin(r)$ so that $\sqrt{1-\frac{1}{2}\alpha^2} = \cos(r)$ and

$$\alpha\sqrt{1-\frac{1}{2}\alpha^2} = \sqrt{2}\sin(r)\cos(r) = \frac{\sqrt{2}}{2}\sin(2r).$$

With parameterizing $\alpha_1 = \alpha_1(r)$, $\alpha_2 = \alpha_2(s)$ in this form, the task is equivalent to showing that the global maximum of $f(r, s) := \tilde{f}(\alpha_1(r), \alpha_2(s))$,

$$f(r, s) = \frac{5}{4} + \frac{5}{4}\sin(r)^2\sin(s)^2 - \sin(r)^2 - \sin(s)^2 + \frac{1+\sqrt{2}}{8}\sin(2r)\sin(2s),$$

in the admissible range $[0, \frac{\pi}{2}]^2$ is at $(0, 0)$. The gradient of f is

$$\nabla f(r, s) = \begin{bmatrix} f_r(r, s) \\ f_s(r, s) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sin(r)\cos(r)(5\sin(s)^2 - 4) + \frac{1+\sqrt{2}}{4}\cos(2r)\sin(2s) \\ \frac{1}{2}\sin(s)\cos(s)(5\sin(r)^2 - 4) + \frac{1+\sqrt{2}}{4}\cos(2s)\sin(2r) \end{bmatrix}.$$

The condition $f_r(r, s) = 0$ gives

$$(33) \quad \frac{1}{2} \sin(r) \cos(r) (4 - 5 \sin(s)^2) = \frac{1 + \sqrt{2}}{4} \cos(2r) \sin(2s).$$

We assume $r, s \neq \frac{\pi}{4}$ and $\sin(s)^2, \sin(r)^2 \neq \frac{4}{5}$ and tackle those special cases afterwards. With excluding those cases we equivalently obtain

$$(34) \quad \frac{1}{2} \frac{\sin(r) \cos(r)}{\cos(2r)} = \frac{1 + \sqrt{2}}{4} \frac{\sin(2s)}{4 - 5 \sin(s)^2}.$$

The left-hand side is greater than zero for $r < \frac{\pi}{4}$ and smaller than zero for $r > \frac{\pi}{4}$ and in each case monotonically increasing. The same holds for the right-hand side with $\sin(s)^2 < \frac{4}{5}$ and $\sin(s)^2 > \frac{4}{5}$. We only consider the case where both sides are greater than zero. The other cases can be tackled analogously. Due to the monotonicity of both sides, it immediately follows that there is at most one pair (r, s) that fulfills the equation. Suppose (r, s) is a pair satisfying the equation. We will show that in this case r has to be equal to s . Assume $r \neq s$. Let $s < r < \frac{\pi}{4}$ and investigate $f_s(r, s) = 0$ which is equivalent to

$$(35) \quad \frac{1}{2} \frac{\sin(s) \cos(s)}{\cos(2s)} = \frac{1 + \sqrt{2}}{4} \frac{\sin(2r)}{4 - 5 \sin(r)^2}.$$

With $s < \frac{\pi}{4}$ we are also in the case where both sides have to be positive in order for the equation to hold. The equation (35) for f_s is the same as the equation (34) for f_r , but with the roles for r and s reversed. Thus, for the pair (r, s) fulfilling the equation (35), $r < s$ applies. This contradicts the above assumption. The case $r < s$ can be tackled in the same way. In summary, for the equation to hold, r must be equal to s so that candidates for extrema lie necessarily on the diagonal $r = s$. Along this diagonal, the function (32) bounding the sectional curvature becomes a parabola in α^2

$$\tilde{f} : \alpha \mapsto \frac{5}{4} + \frac{3 - 2\sqrt{2}}{16} \alpha^4 - \frac{3 - \sqrt{2}}{4} \alpha^2.$$

It is easy to show that $\tilde{f}$ is monotonically strictly decreasing for $\alpha \in [0, \sqrt{2}]$ and therefore the maximum is at $\alpha = 0$.

Now, we tackle the remaining cases $r, s = \frac{\pi}{4}$ and $\sin(s)^2, \sin(r)^2 = \frac{4}{5}$. At $r = \frac{\pi}{4}$, the condition $f_r(r, s) = 0$ yields $\sin(s)^2 = \frac{4}{5}$. But in this case $f_s(r, s) \neq 0$. The other combinations can be treated analogously (and are also not describing any extreme points). This completes the analysis and verifies that $\tilde{f}$ has its unique global maximum in the admissible range $[0, \sqrt{2}]^2$ at $(0, 0)$. Figure 5 displays the function $\tilde{f}$.

Appendix B. The sectional curvature of special $St(n, p)$. For the sake of completeness, we give a direct proof of item 4 of Theorem 8:

On $St(3, 2)$ and $St(3, 3)$, the sectional curvature is $K^{St}(X, Y) \equiv \frac{1}{4}$. The result is already contained in [22, Prop. 6.1]. It follows also immediately from Remark 1 combined with (21).

Proof. On $St(3, 2)$, tangent vectors are skew-symmetric (3×3) -matrices,

$$X = \begin{pmatrix} A & -B^T \\ B & \mathbf{0} \end{pmatrix} = \left(\begin{array}{cc|c} 0 & -a & -b_1 \\ a & 0 & -b_2 \\ b_1 & b_2 & 0 \end{array} \right).$$

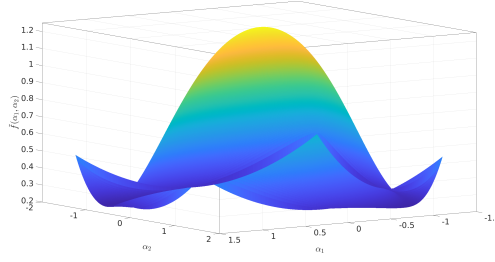


FIG. 5. Plot of the function $\tilde{f}$ from (32) that bounds the Stiefel sectional curvature in the range $[-\sqrt{2}, \sqrt{2}]^2$.

For $X \in \text{skew}(3)$ as above, let $x := \text{vec}_3(X) = (a, b_1, b_2) \in \mathbb{R}^3$ be the vector that features the independent entries of X as components. Let $X, Y \in \text{skew}(3)$ and $x = \text{vec}_3(X)$, $y = \text{vec}_3(Y)$. The following facts are well-known and may be verified by elementary means:

- $\text{vec}_3([X, Y]) = x \times y$.
- $\langle X, Y \rangle_F = 2\langle x, y \rangle_2$. In particular, $\|X\|_F^2 = 2\|x\|_2^2$ and $(X \perp_F Y) \Leftrightarrow (x \perp_F y)$.

Let $X, Y \in \text{skew}(3)$ be an ONB w.r.t. the Frobenius inner product. In this case, $\|x\|_2 = \frac{1}{\sqrt{2}} = \|y\|_2$ for the associated vector representations. The curvature formula of Remark 1 applies and gives

$$K(X, Y) = \frac{1}{2} \|[X, Y]\|_F^2 = \frac{1}{2} 2\|x \times y\|_2^2 = \|x\|_2^2 \|y\|_2^2 \left\| \frac{x}{\|x\|_2} \times \frac{y}{\|y\|_2} \right\|_2^2 = \frac{1}{4}.$$

The same reasoning applies to $St(3, 3)$. $\square$

Note that with the tools at hand, (21) is a one-liner,

$$\|[X, Y]\|_F^2 = 2\|x \times y\|_2^2 \leq 2\|x\|_2^2 \|y\|_2^2 = 2 \frac{1}{2} \|X\|_F^2 \frac{1}{2} \|Y\|_F^2 = \frac{1}{2} \|X\|_F^2 \|Y\|_F^2.$$

Equality holds if and only if $X \perp_F Y$.

Appendix C. Lie group essentials. This section recaps the basics of Lie groups and quotients of Lie groups. It is mainly collected from the textbooks [10], [15] and [23].

C.1. Matrix Lie groups. A *Lie group* is a differentiable manifold $\mathcal{G}$ which also has a group structure, such that the group operations ‘multiplication’ and ‘inversion’,

$$\mathcal{G} \times \mathcal{G} \ni (g, \tilde{g}) \mapsto g \cdot \tilde{g} \in \mathcal{G} \quad \text{and} \quad \mathcal{G} \ni g \mapsto g^{-1} \in \mathcal{G}$$

are both smooth. By definition, a *matrix Lie group* $\mathcal{G}$ is a subgroup of $GL(n, \mathbb{C})$ that is closed relative to $GL(n, \mathbb{C})$. It is then also a Lie group in the above sense.

Let $\mathcal{G}$ be a real matrix Lie group. When endowed with the bracket operator or *matrix commutator* $[X, Y] = XY - YX$, the tangent space $T_I \mathcal{G}$ at the identity is the *Lie algebra* associated with the Lie group $\mathcal{G}$ and is denoted by $\mathfrak{g} = T_I \mathcal{G}$. The linear, skew-symmetric bracket operation is called *Lie bracket* and satisfies the *Jacobi identity*

$$[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0.$$

For any $A \in \mathcal{G}$, the function “left-multiplication with A ” is a diffeomorphism $L_A : \mathcal{G} \rightarrow \mathcal{G}$, $L_A(B) = AB$; its differential at a point $B \in \mathcal{G}$ is the isomorphism

$$d(L_A)_B : T_B \mathcal{G} \rightarrow T_{L_A(B)} \mathcal{G}, \quad d(L_A)_B(X) = AX.$$

(Analogous for “right-multiplication with A ”, $R_A(B) = BA$, $d(R_A)_B(X) = XA$.) The tangent space at an arbitrary location $A \in \mathcal{G}$ is given by the translates (by left-multiplication) of the tangent space at the identity:

$$(36) \quad T_A \mathcal{G} = T_{L_A(I)} \mathcal{G} = A\mathfrak{g} = \{ \Delta = AX \in \mathbb{R}^{n \times n} \mid X \in \mathfrak{g} \},$$

[13, §5.6, p. 160]. The Lie algebra $\mathfrak{g} = T_I \mathcal{G}$ of $\mathcal{G}$ can equivalently be characterized as the set of all matrices Δ such that $\exp_m(t\Delta) \in \mathcal{G}$ for all $t \in \mathbb{R}$, see [15, Def. 3.18 & Cor. 3.46] for the details. The exponential map for a matrix Lie group is the matrix exponential restricted to the corresponding Lie algebra [15, §3.7],

$$\exp_m|_{\mathfrak{g}} : \mathfrak{g} \rightarrow \mathcal{G}.$$

A Riemannian metric $\langle \cdot, \cdot \rangle_{\mathcal{G}}$ on $\mathcal{G}$ is called *left-invariant* if

$$\langle d(L_A)_B X, d(L_A)_B Y \rangle_{AB}^{\mathcal{G}} (= \langle AX, AY \rangle_{AB}^{\mathcal{G}}) = \langle X, Y \rangle_B^{\mathcal{G}}$$

for all $X, Y \in T_B \mathcal{G}$. It is called *right-invariant*, if

$$\langle d(R_A)_B X, d(R_A)_B Y \rangle_{BA}^{\mathcal{G}} (= \langle XA, YA \rangle_{BA}^{\mathcal{G}}) = \langle X, Y \rangle_B^{\mathcal{G}}$$

for all $X, Y \in T_B \mathcal{G}$. If a metric is left- and right-invariant, it is called *bi-invariant*.

C.2. Quotients of Lie groups by closed subgroups. Let $\mathcal{G}$ be a Lie group and $\mathcal{H} \leq \mathcal{G}$ be a Lie subgroup. For $A \in \mathcal{G}$, a subset of $\mathcal{G}$ of the form $[A] := A\mathcal{H} = \{A \cdot Q \mid Q \in \mathcal{H}\}$ is called a *left coset of $\mathcal{H}$* . The left coset $[I]$ is the subgroup itself. The left cosets form a partition of $\mathcal{G}$, and the quotient space $\mathcal{G}/\mathcal{H}$ determined by this partition is called the *left coset space of $\mathcal{G}$ modulo $\mathcal{H}$* , see [18, §21, p. 551]. The next is the central theorem for quotients of Lie groups.

THEOREM 12. (cf. [18, Thm 21.17, p. 551]) *Let $\mathcal{G}$ be a Lie group and let $\mathcal{H}$ be a closed subgroup of $\mathcal{G}$. The left coset space $\mathcal{G}/\mathcal{H}$ is a manifold of dimension $\dim \mathcal{G} - \dim \mathcal{H}$ with a unique differentiable structure such that the quotient map $\pi : \mathcal{G} \rightarrow \mathcal{G}/\mathcal{H}, A \mapsto [A]$ is a smooth surjective submersion. The left action of $\mathcal{G}$ on $\mathcal{G}/\mathcal{H}$ given by $A(B\mathcal{H}) = (AB)\mathcal{H}$ turns $\mathcal{G}/\mathcal{H}$ into a homogeneous $\mathcal{G}$ -space.*

Each preimage $\mathcal{G}_{\pi(A)} := \pi^{-1}([A]) \subset \mathcal{G}$, called *fiber*, is a closed, embedded submanifold. Under the Riemannian metric $\langle \cdot, \cdot \rangle_A^{\mathcal{G}}$, at each point $A \in \mathcal{G}$, the tangent space $T_A \mathcal{G}$ decomposes into an *orthogonal direct sum* $T_A \mathcal{G} = T_A \mathcal{G}_{\pi(A)} \oplus (T_A \mathcal{G}_{\pi(A)})^{\perp}$ with respect to the metric. The tangent space of the fiber $T_A \mathcal{G}_{\pi(A)} =: V_A$ is the kernel of the differential $d\pi_A : T_A \mathcal{G} \rightarrow T_{\pi(A)} \mathcal{G}/\mathcal{H}$ and is called the *vertical space*. Its orthogonal complement $H_A := V_A^{\perp}$ is the *horizontal space*. The key issue is that the tangent space of the quotient at $[A] = \pi(A)$ may be identified with the horizontal space at A .

$$H_A \cong T_{[A]} \mathcal{G}/\mathcal{H}.$$

- For every tangent vector $Y \in T_{[A]}(\mathcal{G}/\mathcal{H})$ there is $\bar{Z} = \bar{X} + \bar{Y} \in V_A \oplus H_A = T_A \mathcal{G}$ such that $d\pi_A(\bar{Z}) = Y$. The horizontal component $\bar{Y} \in H_A$ is unique and is called the *horizontal lift* of $Y \in T_{\pi(A)}(\mathcal{G}/\mathcal{H})$. By going to the horizontal lifts, a Riemannian metric on the quotient can be defined by

$$(37) \quad \langle Y_1, Y_2 \rangle_{[A]}^{\mathcal{G}/\mathcal{H}} := \langle \bar{Y}_1, \bar{Y}_2 \rangle_A^{\mathcal{G}}$$

for $Y_1, Y_2 \in T_{[A]}(\mathcal{G}/\mathcal{H})$.

- W.r.t. this (and only this) metric, by construction, $d\pi_A$ preserves inner products: $d\pi_A$ is a linear isometry between H_A and $T_{[A]}(\mathcal{G}/\mathcal{H})$.
- Horizontal geodesics in $\mathcal{G}$ are mapped to geodesics in $\mathcal{G}/\mathcal{H}$ under π . Horizontal geodesics are geodesics with velocity field staying in the horizontal space for all time t .

At the special point $A = I$, the vertical space is the Lie algebra of $\mathcal{H}$, $V_I = \ker d\pi_I = \mathfrak{h}$. This is because $\mathcal{H} = \pi^{-1}(I)$. For any curve $C(t) \subset \pi^{-1}(I) = \mathcal{H}$ starting from $C(0) = I$ with image in the fiber, it holds $\dot{C}(0) \in T_I\mathcal{H} = \mathfrak{h}$. On the other hand, π is constant along the fiber so that $0 = \frac{d}{dt}\big|_{t=0} \pi(C(t)) = d\pi_I[\dot{C}(0)]$. Hence, any vector tangent to $\mathcal{H}$ at I is in the kernel of $d\pi_I$. At the identity I , the splitting into vertical and horizontal space is

$$T_I\mathcal{G} = \mathfrak{h} \oplus \mathfrak{m} = V_I \oplus H_I = T_I\mathcal{H} \oplus (T_I\mathcal{H})^\perp.$$

Hence, the tangent space of the quotient at $\pi(I)$ is $\mathfrak{m} \cong T_{\pi(I)}\mathcal{G}/\mathcal{H}$. This choice of symbols is common in the literature, but the reader should be aware that $\mathfrak{h}$ is the vertical space, with the choice of letter referring to the subgroup name $\mathcal{H}$ and not to "h for horizontal". The choice of the symbol $\mathfrak{m}$ is motivated by the fact that if the quotient space is called $\mathcal{M} := \mathcal{G}/\mathcal{H}$, then $\mathfrak{m}$ is the tangent space at $\pi(I)$. It is not the associated Lie algebra though, because in general $\mathcal{G}/\mathcal{H}$ is not a Lie group.

REFERENCES

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80(2):199–220, 2004.
- [2] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, New Jersey, 2008.
- [3] P. A. Absil and S. Maitaine. The ultimate upper bound on the injectivity radius of the Stiefel manifold, 2024.
- [4] D. Attali, H. Edelsbrunner, and Y. Mileyko. Weak witnesses for Delaunay triangulations of submanifolds. In B. Lévy and D. Manocha, editors, *Proceedings of the 2007 ACM Symposium on Solid and Physical Modeling, Beijing, China, June 4-6, 2007*, pages 143–150. ACM, 2007.
- [5] E. Begelfor and M. Werman. Affine invariance revisited. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:2087–2094, 2006.
- [6] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.
- [7] E. Celledoni, S. Eidnes, B. Owren, and T. Ringholm. Energy preserving methods on Riemannian manifolds. *Mathematics of Computation*, 89:699–716, 2020.
- [8] M. P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, 1992.
- [9] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- [10] J. Gallier and J. Quaintance. *Differential Geometry and Lie Groups: A Computational Perspective*. Geometry and Computing. Springer International Publishing, 2020.
- [11] K. A. Gallivan, A. Srivastava, X. Liu, and P. Van Dooren. Efficient algorithms for inferences on Grassmann manifolds. In *IEEE Workshop on Statistical Signal Processing*, pages 315–318, 2003.
- [12] J. Q. Ge. DDVV-type inequality for skew-symmetric matrices and Simons-type inequality for Riemannian submersions. *Advances in Mathematics*, 251:62–86, 2014.
- [13] R. Godement and U. Ray. *Introduction to the Theory of Lie Groups*. Universitext. Springer International Publishing, 2017.
- [14] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration: Structure-preserving algorithms for ordinary differential equations.*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2nd edition, 2006.
- [15] B. C. Hall. *Lie Groups, Lie Algebras, and representations: An elementary introduction*. Springer Graduate texts in Mathematics. Springer-Verlag, New York – Berlin – Heidelberg, 2nd edition, 2015.

- [16] K. Hüper, I. Markina, and F. Silva Leite. A Lagrangian approach to extremal curves on Stiefel manifolds. *Journal of Geometrical Mechanics*, 13(1):55–72, 2021.
- [17] A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna. Lie-group methods. *Acta Numerica*, 9:215–365, 2000.
- [18] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York, 2012.
- [19] J. M. Lee. *Introduction to Riemannian Manifolds*. Graduate Texts in Mathematics. Springer International Publishing, Cham, 2nd edition, 2018.
- [20] Y. Man Lui. Advances in matrix manifolds for computer vision. *Image and Vision Computing*, 30(6–7):380–388, 2012.
- [21] H. Q. Minh and V. Murino. *Algorithmic Advances in Riemannian Geometry and Applications: For Machine Learning, Computer Vision, Statistics, and Optimization*. Advances in Computer Vision and Pattern Recognition. Springer International Publishing, Cham, 2016.
- [22] D. Nguyen. Curvatures of Stiefel manifolds with deformation metrics. *Journal of Lie Theory*, 32(2):563–600, 2022.
- [23] B. O’Neill. *Semi-Riemannian geometry - With applications to relativity*, volume 103 of *Pure and Applied Mathematics*. Academic Press, New York, 1983.
- [24] V. Patrangenaru and L. Ellingson. *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2015.
- [25] P. Petersen. *Riemannian Geometry*. Graduate Texts in Mathematics. Springer International Publishing, 2016.
- [26] I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schröder. Multiscale representations for manifold-valued data. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4):1201–1232, 2005.
- [27] Q. Rentmeesters. *Algorithms for data fitting on some common homogeneous spaces*. PhD thesis, Université Catholique de Louvain, Louvain, Belgium, 2013.
- [28] A. Srivastava and P. K. Turaga. *Riemannian computing in computer vision*. Springer International Publishing, 2015.
- [29] Ja. Stoye and R. Zimmermann. On the injectivity radius of the Stiefel manifold: Numerical investigations and an explicit construction of a cut point at short distance, 2024.
- [30] S.-D. Wang, T.-S. Kuo, and C.-F. Hsu. Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation. *IEEE Transactions on Automatic Control*, AC-31(7):654–656, 1986.
- [31] Y.-C. Wong. Differential geometry of Grassmann manifolds. *Proceedings of the National Academy of Sciences of the United States of America*, 57:589–594, 1967.
- [32] Y.-C. Wong. Sectional curvatures of Grassmann manifolds. *Proceedings of the National Academy of Sciences of the United States of America*, 60(1):75–79, 1968.
- [33] G. L. Wu and W. H. Chen. A matrix inequality and its geometric applications. *Acta Math. Sinica*, 31(3):348–355, 1988.
- [34] R. Zimmermann. Manifold interpolation. In P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. Schilders, and L. M. Silveira, editors, *System- and Data-Driven Methods and Algorithms*, volume 1 of *Model Order Reduction*, pages 229–274. De Gruyter, Boston, 2021.
- [35] R. Zimmermann and K. Hüper. Computing the Riemannian logarithm on the Stiefel manifold: Metrics, methods, and performance. *SIAM Journal on Matrix Analysis and Applications*, 43(2):953–980, 2022.