Stiefel and Grassmann Manifolds: A Geometric Foundation for Optimisation with Orthogonality Constraints

Yuvan Marimuthu

October 18, 2026

Abstract

We develop the geometry of the Stiefel and Grassmann manifolds as a foundation for optimisation with orthogonality constraints. The Stiefel manifold is realised as a compact embedded submanifold of a matrix space, with explicit formulas for its tangent projection and Riemannian gradient under the Frobenius metric. We also verify a polar retraction. The Grassmannian is constructed through coordinate charts, and the column-space map is shown to be a smooth surjective submersion inducing a homeomorphism $\mathrm{St}(n,k)/\mathrm{O}(k) \cong \mathrm{Gr}(k,n)$. This identifies the passage from an orthonormal frame to its span with forgetting the choice of basis. The exposition concludes with a brief application to a planned machine learning project involving orthogonal recurrent neural networks.

Contents

1 Introduction **3**

2 Preliminaries **3**
2.1 Basic Definitions 3
2.2 Examples 8

3 Motivation **9**

4 The Stiefel Manifold **10**

5 The Grassmann Manifold **14**

6 From Stiefel to Grassmann **18**

7 Applications **20**

8 Discussion **20**

9 Conclusion **21**

1 Introduction

In the Stiefel and Grassmann constructions, n and k are integers with $1 \leq k \leq n$.

Orthogonality constraints provide a geometric setting for problems involving matrices with orthonormal columns. The Stiefel manifold describes these matrices, while the Grassmann manifold describes the subspaces they span. Understanding their relationship makes precise the distinction between choosing an orthonormal basis and choosing a subspace.

This paper develops these manifolds as a mathematical foundation for a subsequent machine learning project involving orthogonal recurrent neural networks. We prove that the Stiefel manifold is a compact embedded submanifold of a matrix space and derive its tangent projection and Riemannian gradient. We also construct a polar retraction, which converts tangent vectors into feasible steps. For the Grassmannian, we construct a smooth structure through explicit charts and establish its relationship with Stiefel through the column-space map. This gives a smooth surjective submersion and a homeomorphism

$$\mathrm{St}(n,k)/\mathrm{O}(k) \cong \mathrm{Gr}(k,n),$$

where $\mathrm{O}(k)$ acts by changing the orthonormal basis.

The paper is an exposition, following Segal [2] for smooth manifolds and differential calculus, with Boumal [1] supplying the Riemannian and optimisation framework. We assume familiarity with basic topology and finite-dimensional linear algebra, together with multivariable differential calculus. The intended implementation and its experimental evaluation are future work.

Section 2 introduces the required definitions and foundational results. Section 3 explains the distinction between orthonormal frames and subspaces. Sections 4 and 5 develop the two manifolds, and Section 6 establishes their relationship. Section 7 connects the Stiefel results to the planned orthogonal recurrent-network project.

2 Preliminaries

This section introduces the smooth structures and differential calculus used to construct the Stiefel and Grassmann manifolds.

2.1 Basic Definitions

We follow Segal [2, Sections 2–5] for smooth structures and differential calculus.

Definition 2.1. Let X be a topological space. A **co-ordinate chart** on X is the pair (U, φ) where $U \subset X$ is open, and φ is a homeomorphism

$$\varphi : U \xrightarrow{\sim} \varphi(U)$$

where $\varphi(U)$ is open in $\mathbb{R}^n$, for some nonnegative integer n.

Definition 2.2. Let X be a topological space, and fix a nonnegative integer n. X is an **n-dimensional topological manifold** if it is Hausdorff and second countable, and every point lies in a co-ordinate chart with image open in $\mathbb{R}^n$. This last condition says that X is locally Euclidean.

We require that X is Hausdorff, and second countable, as to exclude certain pathological examples.

Definition 2.3. Let X be a topological manifold, and let (U, φ_1) and (V, φ_2) be two co-ordinate charts on X. The **transition function** between these two co-ordinate charts is the function:

$$\phi_{21} = \varphi_2 \circ \varphi_1^{-1} : \varphi_1(U \cap V) \xrightarrow{\sim} \varphi_2(U \cap V)$$

It is a homeomorphism, with inverse $\phi_{21}^{-1} = \phi_{12} = \varphi_1 \circ \varphi_2^{-1}$.

Definition 2.4. Let X be a topological manifold. An **atlas** on X is a collection of co-ordinate charts on X:

$$\varphi_i : U_i \xrightarrow{\sim} \widetilde{U}_i \subset \mathbb{R}^n$$

indexed by some, possibly infinite, set I, such that:

$$\bigcup_{i \in I} U_i = X$$

An atlas is a **smooth atlas** if all its transition functions are smooth, i.e. infinitely differentiable.

Definition 2.5. Let X be a topological manifold, and let $\mathcal{A} = \{(U_i, \varphi_i); i \in I\}$ be a smooth atlas on X. Let (U, φ) be any co-ordinate chart on X. The chart (U, φ) is **compatible** with the atlas $\mathcal{A}$ if the transition functions between (U, φ) and any chart in $\mathcal{A}$ are diffeomorphisms, i.e. they are smooth functions with smooth inverses.

In other words, the new chart (U, φ) is compatible with the atlas $\mathcal{A}$ if the union $\mathcal{A} \cup \{(U, \varphi)\}$ is still a smooth atlas.

Let $\mathcal{A}, \mathcal{B}$ be smooth atlases on X. The atlases $\mathcal{A}$ and $\mathcal{B}$ are **compatible** if every chart in $\mathcal{A}$ is compatible with $\mathcal{B}$, and vice versa.

Corollary 2.6 ([2, Corollary 2.20])**.** *Compatibility is an equivalence relation on smooth atlases.*

Definition 2.7. A **smooth manifold** is a topological manifold X together with an equivalence class $[\mathcal{A}]$ of compatible smooth atlases on X.

The equivalence class $[\mathcal{A}]$ of compatible smooth atlases is called a **smooth structure** on X.

Definition 2.8. Let X, Y be smooth manifolds, of dimension m and n respectively, and let $H : X \longrightarrow Y$ be a continuous function. Fix a point $x \in X$. H is **smooth at** x if there exists a chart (U, φ) compatible with the smooth structure on X, with $x \in U$, and a chart (V, ψ) compatible with the smooth structure on Y, with $H(U) \subset V$, such that the composition

$$\psi \circ H \circ \varphi^{-1} : \varphi(U) \longrightarrow \psi(V)$$

is smooth at $\varphi(x)$.

The function H is **smooth** if it is smooth at every point $x \in X$.

Definition 2.9. Let X, Y be smooth manifolds. A function $H : X \longrightarrow Y$ is a **diffeomorphism** if it is a bijection and both H and H^{-1} are smooth. If such a function exists, X and Y are **diffeomorphic**.

Definition 2.10. Let X be a set. A **pseudo-chart** on X is the pair (U, φ), where $U \subset X$ and

$$\varphi : U \longrightarrow \varphi(U)$$

is a bijection onto an open subset $\varphi(U) \subset \mathbb{R}^n$.

A collection $\mathcal{A} = \{(U_i, \varphi_i) : i \in I\}$ of pseudo-charts, with the same dimension n, is called a pseudo-atlas on X if

$$\bigcup_{i \in I} U_i = X.$$

Proposition 2.11 ([2, Proposition 2.26])**.** *Let X be a set, and let $\mathcal{A} = \{(U_i, \varphi_i) : i \in I\}$ be a pseudo-atlas for X. Suppose that for any pair $(U_1, \varphi_1), (U_2, \varphi_2)$ of pseudo-charts in $\mathcal{A}$, we have that:*

1. *both $\varphi_1(U_1 \cap U_2) \subset \varphi_1(U_1)$ and $\varphi_2(U_1 \cap U_2) \subset \varphi_2(U_2)$ are open subsets, and*
2. *the transition function*

$$\phi_{21} = \varphi_2 \circ \varphi_1^{-1} : \varphi_1(U_1 \cap U_2) \longrightarrow \varphi_2(U_1 \cap U_2)$$

is continuous.

Then there exists a unique topology on X such that each pseudo-chart (U_i, φ_i) is a co-ordinate chart on X.

Corollary 2.12 ([2, Corollary 2.28])**.** *If we further assume that each transition function ϕ_{21} is smooth, then $[\mathcal{A}]$ is a smooth structure on the topological manifold X, provided the resulting topology is Hausdorff and second-countable.*

Definition 2.13. Let X be a smooth manifold of dimension n, let m be an integer satisfying $0 \leq m \leq n$, and let $\mathcal{M} \subset X$. The subset $\mathcal{M}$ is an **m-dimensional submanifold** of X if, for every $x \in \mathcal{M}$, there exists a compatible co-ordinate chart

$$\varphi : U \xrightarrow{\sim} \varphi(U) \subset \mathbb{R}^n$$

with $x \in U$, and an m-dimensional affine subspace $A \subset \mathbb{R}^n$, such that

$$\varphi(U \cap \mathcal{M}) = \varphi(U) \cap A.$$

This is the convention of [2, Definition 3.2]. We also use the term **embedded submanifold** for a submanifold in this sense.

By [2, Proposition 3.7], $\mathcal{M}$, with the subspace topology, carries a smooth structure induced from X. We always equip a submanifold with this structure. By [2, Lemmas 4.7–4.8], the inclusion $\iota : \mathcal{M} \hookrightarrow X$ is smooth, and a map $F : Y \longrightarrow \mathcal{M}$ from a smooth manifold is smooth if and only if $\iota \circ F$ is smooth. Smooth maps on an ambient manifold therefore restrict to smooth maps on its submanifolds; in particular, a map admitting smooth local ambient extensions is smooth.

Definition 2.14. Let $\mathcal{M}$ be a smooth manifold, and let $x \in \mathcal{M}$. A curve through x is a smooth function $c : I \longrightarrow \mathcal{M}$, where I is an open interval containing 0 and $c(0) = x$. Two curves c_1, c_2 through x are **tangent at** x if, for any compatible co-ordinate chart (U, φ) containing x,

$$\frac{d}{dt}(\varphi \circ c_1)(t)\bigg|_{t=0} = \frac{d}{dt}(\varphi \circ c_2)(t)\bigg|_{t=0}.$$

The compositions are taken on sufficiently small intervals around 0. This condition is independent of the choice of chart.

A **tangent vector** at x is an equivalence class of curves through x under this relation. The **tangent space** at x is the set

$$T_x\mathcal{M} = \{\text{smooth curves through } x\}/\sim,$$

where $c_1 \sim c_2$ means that the curves are tangent at x, following [2, Definitions 5.4–5.5].

By [2, Proposition 5.8], $T_x\mathcal{M}$ is a vector space of dimension $\dim \mathcal{M}$. Its vector space structure is the one for which the chart identification

$$[c] \longmapsto \frac{d}{dt}(\varphi \circ c)(t)\bigg|_{t=0}$$

is a linear isomorphism with $\mathbb{R}^{\dim \mathcal{M}}$.

For an embedded submanifold $\mathcal{M}$ of a finite-dimensional real vector space $\mathcal{E}$, we identify $[c]$ with the ambient velocity $c'(0)$, as in [2, Section 5.2]. Under this identification, $T_x\mathcal{M}$ is a subspace of $\mathcal{E}$, and

$$T_x\mathcal{M} = \{c'(0) : c : I \longrightarrow \mathcal{M} \text{ is smooth and } c(0) = x\}.$$

Here, derivatives are taken in $\mathcal{E}$. If $\mathcal{M}$ is open in $\mathcal{E}$, then $T_x\mathcal{M} = \mathcal{E}$.

Definition 2.15. Let X, Y be smooth manifolds, and let $F : X \longrightarrow Y$ be smooth. The **derivative** of F at $x \in X$ is the linear map

$$DF(x) : T_xX \longrightarrow T_{F(x)}Y, \qquad [c] \longmapsto [F \circ c].$$

This map is well-defined and linear by [2, Proposition 5.10].

For smooth maps $F : X \longrightarrow Y$ and $G : Y \longrightarrow Z$, this definition gives the chain rule

$$D(G \circ F)(x) = DG(F(x)) \circ DF(x).$$

Theorem 2.16 (Regular level sets [2, Proposition 3.16 and Lemma 5.13])**.** *Let $\mathcal{E}$ and $\mathcal{E}'$ be finite-dimensional real vector spaces, and let $F : \mathcal{E} \to \mathcal{E}'$ be smooth. A point $c \in \mathcal{E}'$ is called a regular value of F if $DF(x)$ is surjective for every $x \in F^{-1}(c)$.*

If c is a regular value and $\mathcal{M} = F^{-1}(c)$ is nonempty, then $\mathcal{M}$ is a smooth embedded submanifold of $\mathcal{E}$, with

$$\dim \mathcal{M} = \dim \mathcal{E} - \dim \mathcal{E}', \qquad T_x\mathcal{M} = \ker DF(x) \quad \textit{for every } x \in \mathcal{M}.$$

To define gradients on these manifolds, we equip their tangent spaces with inner products. For the following metric and gradient definitions, we follow [1, Definitions 3.51–3.53, Proposition 3.54, and Definitions 3.55, 3.58 and 3.60].

Definition 2.17. Let $\mathcal{M}$ be a smooth manifold. A **Riemannian metric** on $\mathcal{M}$ assigns an inner product $\langle \cdot, \cdot \rangle_x$ to each tangent space $T_x\mathcal{M}$, varying smoothly with x. A smooth manifold equipped with a Riemannian metric is a **Riemannian manifold**.

Definition 2.18. Let $\mathcal{M}$ be an embedded submanifold of a finite-dimensional real vector space $\mathcal{E}$ equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$. The induced metric on $\mathcal{M}$ is defined by

$$\langle u, v \rangle_x = \langle u, v \rangle_{\mathcal{E}}, \qquad u, v \in T_x\mathcal{M}.$$

This is a Riemannian metric by [1, Proposition 3.54]. Equipped with this metric, $\mathcal{M}$ is a **Riemannian submanifold** of $\mathcal{E}$.

Definition 2.19. Let $\mathcal{M}$ be a Riemannian manifold, and let $f : \mathcal{M} \to \mathbb{R}$ be smooth. For each $x \in \mathcal{M}$, the **Riemannian gradient** of f at x is the unique vector $\operatorname{grad} f(x) \in T_x\mathcal{M}$ satisfying

$$Df(x)[v] = \langle \operatorname{grad} f(x), v \rangle_x \qquad \text{for every } v \in T_x\mathcal{M}.$$

Here,

$$Df(x)[v] = \frac{d}{dt} f(c(t)) \bigg|_{t=0},$$

where c is any smooth curve through x representing v. This value is independent of the choice of such a curve.

Definition 2.20. Let $\mathcal{M} \subset \mathcal{E}$ be an embedded submanifold, where $\mathcal{E}$ is a finite-dimensional real inner product space, and let $x \in \mathcal{M}$. The **orthogonal projector** to $T_x\mathcal{M}$ is the linear map $\operatorname{Proj}_x : \mathcal{E} \to \mathcal{E}$ characterized by the following properties:

1. Range: $\operatorname{im}(\operatorname{Proj}_x) = T_x\mathcal{M}$.
2. Projector: $\operatorname{Proj}_x \circ \operatorname{Proj}_x = \operatorname{Proj}_x$.
3. Orthogonal: $\langle u - \operatorname{Proj}_x(u), v \rangle = 0 \quad$ for every $v \in T_x\mathcal{M}$ and $u \in \mathcal{E}$.

Proposition 2.21 ([1, Proposition 3.61])**.** *Let $\mathcal{M} \subset \mathcal{E}$ be a Riemannian submanifold, and let $f : \mathcal{M} \to \mathbb{R}$ be smooth. Then the Riemannian gradient of f is given by*

$$\operatorname{grad} f(x) = \operatorname{Proj}_x\Big(\nabla \bar{f}(x)\Big), \qquad x \in \mathcal{M},$$

where $\bar{f}$ is any smooth extension of f to an open neighbourhood of $\mathcal{M}$ in $\mathcal{E}$, and $\nabla \bar{f}(x)$ is its Euclidean gradient.

The Euclidean gradient is characterized by

$$D\bar{f}(x)[v] = \langle \nabla \bar{f}(x), v \rangle \qquad \textit{for every } v \in \mathcal{E}.$$

A gradient gives a tangent direction at a point. Retractions allow us to use this direction to move on the manifold.

Definition 2.22. Let $\mathcal{M}$ be a smooth manifold. The **tangent bundle** of $\mathcal{M}$ is the set

$$T\mathcal{M} = \{(x, v) : x \in \mathcal{M},\ v \in T_x\mathcal{M}\}.$$

It carries a natural smooth structure by [2, Proposition E.1]. Following [1, Definition 3.47], a **retraction** on $\mathcal{M}$ is a smooth map

$$R : T\mathcal{M} \to \mathcal{M} : (x, v) \mapsto R_x(v)$$

such that, for every $x \in \mathcal{M}$ and $v \in T_x\mathcal{M}$, the curve $c(t) = R_x(tv)$ satisfies $c(0) = x$ and represents v.

For an embedded submanifold $\mathcal{M} \subset \mathcal{E}$, we regard R as a map into $\mathcal{E}$. Under the ambient velocity identification, $T\mathcal{M}$ is an embedded submanifold of $\mathcal{E} \times \mathcal{E}$ with its natural tangent-bundle structure. In particular, an ambient smooth extension of $(x, v) \mapsto R_x(v)$ suffices to establish smoothness of R.

We will also need to determine when the derivative of a smooth map is surjective.

Definition 2.23. Let X, Y be smooth manifolds. A smooth function $F : X \longrightarrow Y$ is a **submersion** if

$$\operatorname{rank} DF(x) = \dim Y \qquad \text{for every } x \in X.$$

Equivalently, $DF(x) : T_xX \longrightarrow T_{F(x)}Y$ is surjective at every point, as in [2, Definition 4.16 and Section 5.1].

2.2 Examples

Example 2.24 (Sphere)**.** Following [2, Examples 3.19 and 5.14], consider the smooth function

$$h : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}, \qquad (x_0, \ldots, x_n) \mapsto x_0^2 + \cdots + x_n^2.$$

The unit sphere is $S^n = h^{-1}(1)$. The derivative of h at $x = (x_0, \ldots, x_n)$ is the 1-by-$(n+1)$ matrix

$$Dh(x) = (2x_0, \ldots, 2x_n) : \mathbb{R}^{n+1} \longrightarrow \mathbb{R},$$

which is surjective if at least one of the x_i is nonzero. Every point of S^n is nonzero, so 1 is a regular value of h. By definition 2.16, S^n is an n-dimensional embedded submanifold of $\mathbb{R}^{n+1}$.

The same theorem gives its tangent space at $x \in S^n$:

$$T_xS^n = \ker Dh(x) = \{v \in \mathbb{R}^{n+1} : x \cdot v = 0\} = x^\perp.$$

Example 2.25 (Real projective space)**.** Following [2, Example 2.30], let $\mathbb{RP}^n$ denote the set of lines through the origin in $\mathbb{R}^{n+1}$. For a nonzero vector $x = (x_0, \ldots, x_n)$, we denote the corresponding line by $[x] = x_0 : \cdots : x_n$. For each $i \in \{0, \ldots, n\}$, set

$$U_i = \{x_0 : \cdots : x_n : x_i \neq 0\}.$$

These sets cover $\mathbb{RP}^n$, since every nonzero vector has a nonzero component. Define

$$f_i : U_i \longrightarrow \mathbb{R}^n, \qquad (x_0 : \cdots : x_n) \mapsto \left(\frac{x_0}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_n}{x_i}\right).$$

The ratios are unchanged when x is multiplied by a nonzero scalar, so f_i is well-defined. It is a bijection, with inverse

$$f_i^{-1} : (\tilde{x}_1, \ldots, \tilde{x}_n) \mapsto \tilde{x}_1 : \cdots : \tilde{x}_i : 1 : \tilde{x}_{i+1} : \cdots : \tilde{x}_n.$$

Write $z_i(\tilde{x}) \in \mathbb{R}^{n+1}$ for the vector obtained by inserting 1 into coordinate i, with coordinates indexed from 0 to n. Thus $f_i^{-1}(\tilde{x}) = [z_i(\tilde{x})]$.

For any i, j, the overlap image is

$$f_i(U_i \cap U_j) = \{\tilde{x} \in \mathbb{R}^n : (z_i(\tilde{x}))_j \neq 0\},$$

which is open. On this set, the transition function is

$$(f_j \circ f_i^{-1})(\tilde{x}) = \left(\frac{(z_i(\tilde{x}))_r}{(z_i(\tilde{x}))_j} \right)_{r \neq j},$$

with components ordered by $r \in \{0, \ldots, n\}$. Each transition is smooth because its denominator is nonzero on the overlap. By definition 2.11, these pseudo-charts determine a unique topology in which they are co-ordinate charts.

To check Hausdorffness, consider

$$P : \mathbb{RP}^n \longrightarrow \mathrm{Sym}(n+1), \qquad [x] \longmapsto \frac{xx^\top}{\|x\|^2},$$

where $\mathrm{Sym}(n+1)$ is the vector space of real symmetric $(n+1)$-by-$(n+1)$ matrices. This map is well-defined because scaling x does not change the quotient. It is injective because the image of the matrix $P([x])$ is the line $[x]$. In the chart U_i,

$$(P \circ f_i^{-1})(\tilde{x}) = \frac{z_i(\tilde{x}) z_i(\tilde{x})^\top}{1 + \|\tilde{x}\|^2},$$

which is continuous. Thus P is a continuous injection into the Hausdorff space $\mathrm{Sym}(n+1)$, so $\mathbb{RP}^n$ is Hausdorff.

Each U_i has a countable basis transported from $\mathbb{R}^n$. Since the U_i form a finite open cover, the union of these bases is a countable basis for $\mathbb{RP}^n$. By definition 2.12, the charts therefore give $\mathbb{RP}^n$ the structure of a smooth manifold of dimension n.

More generally, the set of k-dimensional subspaces of $\mathbb{R}^{n+1}$ is the Grassmannian $\mathrm{Gr}(k, n+1)$. Real projective space $\mathbb{RP}^n$ is the case $k = 1$.

3 Motivation

Orthogonality constraints arise when the columns of a matrix are required to form an orthonormal family. For a matrix $X \in \mathbb{R}^{n \times k}$, this condition is

$$X^\top X = I_k.$$

The admissible matrices form the Stiefel manifold $\mathrm{St}(n, k)$. An arbitrary change to X need not preserve this equation, so optimisation under the constraint requires a description of the allowable directions and a way to move while remaining on the manifold.

In some problems, the object of interest is the subspace spanned by the columns. If $Q \in \mathrm{O}(k)$, then XQ is another orthonormal basis of the same subspace, since

$$(XQ)^\top (XQ) = I_k, \qquad \mathrm{col}(XQ) = \mathrm{col}(X).$$

The Grassmannian $\mathrm{Gr}(k, n)$ records these k-dimensional subspaces without choosing a basis. This distinction leads to the column-space map

$$\pi : \mathrm{St}(n, k) \longrightarrow \mathrm{Gr}(k, n), \qquad \pi(X) = \mathrm{col}(X).$$

We will show that its fibres consist precisely of the orthonormal bases of each subspace, giving the identification

$$\mathrm{St}(n, k)/\mathrm{O}(k) \cong \mathrm{Gr}(k, n)$$

as topological spaces.

The following sections construct both manifolds and establish this relationship. For Stiefel, we also derive a tangent projection and a retraction, providing tools for optimisation with orthogonality constraints. These results will support a subsequent machine learning project involving orthogonal recurrent weight matrices, discussed briefly in section 7.

4 The Stiefel Manifold

We now apply the preceding results to the constraint $X^\top X = I_k$, deriving the gradient and a retraction for optimisation.

Convention 4.1 (Matrix conventions). Fix integers $1 \leq k \leq n$. Endow $\mathbb{R}^{n\times k}$ with the Frobenius inner product

$$\langle A, B\rangle = \mathrm{tr}(A^\top B), \qquad A, B \in \mathbb{R}^{n\times k}.$$

Let $\mathrm{Sym}(k) \subset \mathbb{R}^{k\times k}$ denote the space of symmetric $k \times k$ matrices, and for $B \in \mathbb{R}^{k\times k}$ define

$$\mathrm{sym}(B) = \frac{B + B^\top}{2}.$$

Throughout this section, the Stiefel manifold carries the Riemannian metric induced by the Frobenius inner product.

Definition 4.2. The **Stiefel manifold** $\mathrm{St}(n, k)$ is the set of matrices in $\mathbb{R}^{n\times k}$ with orthonormal columns in $\mathbb{R}^n$ with respect to the standard Euclidean inner product,

$$\mathrm{St}(n, k) = \{X \in \mathbb{R}^{n\times k} : X^\top X = I_k\},$$

where I_k denotes the identity matrix of size k. We call matrices in $\mathrm{St}(n, k)$ **orthonormal matrices** and we reserve the term **orthogonal matrices** for square orthonormal matrices.

Theorem 4.3. *The Stiefel manifold* $\mathrm{St}(n, k)$ *is an embedded submanifold of* $\mathbb{R}^{n\times k}$ *of dimension*

$$\dim \mathrm{St}(n, k) = \dim \mathbb{R}^{n\times k} - \dim \mathrm{Sym}(k) = nk - \frac{k(k+1)}{2}.$$

Proof. Following [1, Section 7.3], consider the smooth map

$$h : \mathbb{R}^{n\times k} \longrightarrow \mathrm{Sym}(k), \qquad h(X) = X^\top X - I_k.$$

Then $\mathrm{St}(n,k) = h^{-1}(0)$. For $X, V \in \mathbb{R}^{n\times k}$,

$$\begin{aligned} Dh(X)[V] &= \frac{d}{dt} h(X+tV)\Big|_{t=0} \\ &= X^\top V + V^\top X. \end{aligned}$$

To show that 0 is a regular value of h, let $X \in \mathrm{St}(n,k)$ and $S \in \mathrm{Sym}(k)$. Taking $V = \frac{1}{2}XS$, we obtain

$$Dh(X)[V] = \frac{1}{2}X^\top XS + \frac{1}{2}S^\top X^\top X = S,$$

since $X^\top X = I_k$ and $S^\top = S$. Hence $Dh(X)$ is surjective for every $X \in h^{-1}(0)$.

The set $\mathrm{St}(n,k)$ is nonempty, since the first k columns of I_n form an element of $\mathrm{St}(n,k)$. By definition 2.16, $\mathrm{St}(n,k)$ is an embedded submanifold of $\mathbb{R}^{n\times k}$, with

$$\dim \mathrm{St}(n,k) = \dim \mathbb{R}^{n\times k} - \dim \mathrm{Sym}(k) = nk - \frac{k(k+1)}{2}.$$

□

Proposition 4.4 (Compactness)**.** *The Stiefel manifold $\mathrm{St}(n,k)$ is compact. Consequently, every continuous function $f : \mathrm{St}(n,k) \longrightarrow \mathbb{R}$ attains its maximum and minimum.*

Proof. Since the map $h(X) = X^\top X - I_k$ is continuous, $\mathrm{St}(n,k) = h^{-1}(0)$ is closed in $\mathbb{R}^{n\times k}$. For every $X \in \mathrm{St}(n,k)$, its Frobenius norm satisfies

$$\|X\|_F^2 = \mathrm{tr}(X^\top X) = \mathrm{tr}(I_k) = k,$$

so $\mathrm{St}(n,k)$ is bounded. By the Heine–Borel theorem, $\mathrm{St}(n,k)$ is compact. Since it is nonempty, the extreme value theorem gives the stated conclusion. □

Proposition 4.5. *For $X \in \mathrm{St}(n,k)$, the tangent space is*

$$T_X\mathrm{St}(n,k) = \{V \in \mathbb{R}^{n\times k} : X^\top V + V^\top X = 0\}.$$

*The **normal space** is the orthogonal complement of $T_X\mathrm{St}(n,k)$ with respect to the Frobenius inner product, and is given by*

$$N_X\mathrm{St}(n,k) := (T_X\mathrm{St}(n,k))^\perp = \{XS : S \in \mathrm{Sym}(k)\}.$$

These are the descriptions in [1, Section 7.3, equations (7.17) and (7.25)].

Proof. For the map $h(X) = X^\top X - I_k$, we have already shown that 0 is a regular value and

$$Dh(X)[V] = X^\top V + V^\top X.$$

By definition 2.16,

$$T_X\mathrm{St}(n,k) = \ker Dh(X) = \{V \in \mathbb{R}^{n\times k} : X^\top V + V^\top X = 0\}.$$

Let $S \in \mathrm{Sym}(k)$ and $V \in T_X\mathrm{St}(n,k)$. The tangent-space equation implies that $X^\top V$ is skew-symmetric. Since symmetric and skew-symmetric matrices are orthogonal under the Frobenius inner product,

$$\langle XS, V\rangle = \mathrm{tr}(S^\top X^\top V) = 0.$$

Thus $\{XS : S \in \mathrm{Sym}(k)\} \subseteq N_X\mathrm{St}(n,k)$.

The linear map $S \mapsto XS$ is injective, since $X^\top XS = S$. Its image therefore has dimension $k(k+1)/2$. Also,

$$\dim N_X\mathrm{St}(n,k) = nk - \dim T_X\mathrm{St}(n,k) = \frac{k(k+1)}{2}.$$

The inclusion is therefore an equality. □

Proposition 4.6 (Tangent projection)**.** *For $X \in \mathrm{St}(n,k)$, the orthogonal projection onto $T_X\mathrm{St}(n,k)$ is*

$$\mathrm{Proj}_X(Z) = Z - X\,\mathrm{sym}(X^\top Z), \qquad Z \in \mathbb{R}^{n\times k},$$

as in [1, Section 7.3, equation (7.26)].

Proof. Let $Z \in \mathbb{R}^{n\times k}$. By definition 4.5, the difference between Z and its orthogonal projection has the form

$$Z - \mathrm{Proj}_X(Z) = XS$$

for some $S \in \mathrm{Sym}(k)$. Since $\mathrm{Proj}_X(Z) = Z - XS$ is tangent at X,

$$X^\top(Z - XS) + (Z - XS)^\top X = 0.$$

Using $X^\top X = I_k$ and $S^\top = S$, this gives

$$X^\top Z + Z^\top X - 2S = 0.$$

Hence

$$S = \frac{X^\top Z + Z^\top X}{2} = \mathrm{sym}(X^\top Z),$$

which yields the stated formula. □

Corollary 4.7 (Riemannian gradient)**.** *Let $f : \mathrm{St}(n,k) \longrightarrow \mathbb{R}$ be smooth, and let $\bar{f}$ be a smooth extension to an open neighbourhood of $\mathrm{St}(n,k)$ in $\mathbb{R}^{n\times k}$. With the induced Frobenius metric,*

$$\mathrm{grad}\, f(X) = \nabla\bar{f}(X) - X\,\mathrm{sym}\Big(X^\top\nabla\bar{f}(X)\Big),$$

as in [1, Section 7.3, equation (7.28)].

In particular, for $A \in \mathrm{Sym}(n)$ and $f(X) = \mathrm{tr}(X^\top AX)$,

$$\mathrm{grad}\, f(X) = 2(I_n - XX^\top)AX.$$

Proof. The general formula follows from definitions 2.21 and 4.6.

For the particular objective, take the extension $\bar{f}(X) = \mathrm{tr}(X^\top AX)$ on $\mathbb{R}^{n\times k}$. Since $A^\top = A$, for every $V \in \mathbb{R}^{n\times k}$,

$$\begin{aligned} D\bar{f}(X)[V] &= \mathrm{tr}(V^\top AX + X^\top AV) \\ &= 2\,\mathrm{tr}(V^\top AX) = \langle 2AX, V\rangle. \end{aligned}$$

Hence $\nabla \bar{f}(X) = 2AX$. The matrix $X^\top AX$ is symmetric, so

$$\operatorname{grad} f(X) = 2AX - 2XX^\top AX = 2(I_n - XX^\top)AX.$$

□

To construct a retraction, we first establish smoothness of the matrix inverse square root.

Lemma 4.8 (Smooth matrix square roots)**.** *Let* $\mathrm{Sym}^+(k)$ *denote the open subset of* $\mathrm{Sym}(k)$ *consisting of positive-definite matrices. Every* $A \in \mathrm{Sym}^+(k)$ *has a unique positive-definite symmetric square root* $A^{1/2}$*. The maps*

$$A \longmapsto A^{1/2}, \qquad A \longmapsto A^{-1/2} := (A^{1/2})^{-1}$$

are smooth on $\mathrm{Sym}^+(k)$*.*

Proof. By the spectral theorem, write

$$A = Q \operatorname{diag}(\lambda_1, \ldots, \lambda_k) Q^\top, \qquad \lambda_i > 0,$$

where Q is orthogonal. Then

$$S = Q \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k}) Q^\top$$

is positive-definite and satisfies $S^2 = A$.

Uniqueness follows from the spectral theorem: any positive-definite symmetric square root commutes with A and acts as $\sqrt{\lambda} I$ on each eigenspace of A with eigenvalue λ.

The positive-definite matrices form an open subset of $\mathrm{Sym}(k)$: if $u^\top Au \geq m\|u\|^2$ for some $m > 0$, every symmetric perturbation E with $\|E\|_F < m$ leaves $A + E$ positive-definite.

Consider the smooth map

$$q : \mathrm{Sym}^+(k) \longrightarrow \mathrm{Sym}^+(k), \qquad q(S) = S^2.$$

Its derivative is

$$Dq(S)[H] = SH + HS.$$

Choose an orthogonal matrix Q such that $S = Q \operatorname{diag}(\mu_1, \ldots, \mu_k) Q^\top$, with each $\mu_i > 0$. Writing $\widetilde{H} = Q^\top H Q$, we have

$$\left(Q^\top Dq(S)[H] Q\right)_{ij} = (\mu_i + \mu_j)\widetilde{H}_{ij}.$$

Since every $\mu_i + \mu_j$ is positive, $Dq(S)$ is an invertible linear map on $\mathrm{Sym}(k)$.

By the inverse function theorem [2, Theorem 3.9], q has a smooth local inverse near every $S \in \mathrm{Sym}^+(k)$. Uniqueness of the positive-definite square root identifies these local inverses with $A \mapsto A^{1/2}$, so this map is smooth. Since matrix inversion is smooth on invertible matrices, $A \mapsto A^{-1/2}$ is also smooth. □

Proposition 4.9 (Polar retraction)**.** *The map*

$$R : T\mathrm{St}(n,k) \longrightarrow \mathrm{St}(n,k), \qquad R_X(V) = (X+V)(I_k + V^\top V)^{-1/2},$$

is a retraction on $\mathrm{St}(n,k)$*, called the* ***polar retraction*** *[1, Section 7.3, equation (7.24)].*

Proof. For every $V \in \mathbb{R}^{n\times k}$, the matrix $I_k + V^\top V$ is positive-definite, since

$$u^\top (I_k + V^\top V)u = \|u\|^2 + \|Vu\|^2 > 0 \qquad \text{for every } u \neq 0.$$

By definition 4.8, the formula for $R_X(V)$ defines a smooth map into $\mathbb{R}^{n\times k}$ on the ambient space $\mathbb{R}^{n\times k} \times \mathbb{R}^{n\times k}$.

For $X \in \mathrm{St}(n,k)$ and $V \in T_X\mathrm{St}(n,k)$, we have $X^\top V + V^\top X = 0$, so

$$(X+V)^\top (X+V) = I_k + V^\top V.$$

Writing $B = I_k + V^\top V$, it follows that

$$R_X(V)^\top R_X(V) = B^{-1/2} B B^{-1/2} = I_k.$$

Thus $R_X(V) \in \mathrm{St}(n,k)$, and R is smooth.

Clearly, $R_X(0) = X$. To verify the first-order condition, write

$$C(t) = (I_k + t^2 V^\top V)^{-1/2}.$$

The chain rule gives $C'(0) = 0$, since the derivative of $I_k + t^2 V^\top V$ vanishes at $t = 0$. As $C(0) = I_k$, we obtain

$$\frac{d}{dt} R_X(tV)\Big|_{t=0} = \frac{d}{dt}\Big((X+tV)C(t)\Big)\Big|_{t=0} = V.$$

Hence R satisfies the defining properties of a retraction. □

Remark 4.10. The Stiefel manifold includes two familiar special cases. When $k = 1$, the condition $X^\top X = 1$ says that X is a unit vector, so

$$\mathrm{St}(n,1) = S^{n-1}.$$

When $k = n$, the condition $X^\top X = I_n$ defines the orthogonal group, so

$$\mathrm{St}(n,n) = \mathrm{O}(n).$$

5 The Grassmann Manifold

We now consider the subspaces spanned by matrices, independently of the chosen basis. We follow [2, Example 2.31], representing bases by columns.

Definition 5.1 (Grassmann manifold as a set)**.** Fix integers $1 \le k \le n$. The **Grassmannian** $\mathrm{Gr}(k,n)$ is the set of k-dimensional subspaces of $\mathbb{R}^n$.

Every $S \in \mathrm{Gr}(k,n)$ can be represented as $S = \mathrm{col}(Y)$, where $Y \in \mathbb{R}^{n\times k}$ has rank k and $\mathrm{col}(Y)$ denotes its column space.

Lemma 5.2 (Change of basis)**.** *Let $Y, Y' \in \mathbb{R}^{n\times k}$ have rank k. Then*

$$\mathrm{col}(Y) = \mathrm{col}(Y') \quad \Longleftrightarrow \quad Y' = YG \text{ for some } G \in \mathrm{GL}_k(\mathbb{R}),$$

where $\mathrm{GL}_k(\mathbb{R})$ is the group of invertible $k \times k$ real matrices.

Proof. Suppose $\mathrm{col}(Y) = \mathrm{col}(Y')$. Each column of Y' is a linear combination of the columns of Y, so $Y' = YG$ for some $G \in \mathbb{R}^{k\times k}$. If $Gv = 0$, then $Y'v = YGv = 0$. Since Y' has rank k, this implies $v = 0$. Thus G is invertible.

Conversely, if $Y' = YG$ with G invertible, then $\mathrm{col}(Y') \subseteq \mathrm{col}(Y)$. Since $Y = Y'G^{-1}$, the reverse inclusion also holds. □

The charts normalise an invertible row minor to the identity. For $I = \{1, \ldots, k\}$, the normalised matrix is $\binom{I_k}{B}$, whose column space is the graph of $u \mapsto Bu$.

Definition 5.3 (Grassmann pseudo-charts)**.** Let $I \subseteq \{1, \ldots, n\}$ have k elements, and let I^c denote its complement. For $Y \in \mathbb{R}^{n\times k}$, write Y_I for the submatrix formed by the rows indexed by I, in increasing order. Define

$$U_I = \{S \in \mathrm{Gr}(k, n) : Y_I \text{ is invertible for a basis matrix } Y \text{ of } S\}.$$

This condition is independent of the choice of basis: by definition 5.2, another basis matrix has the form $Y' = YG$, with $G \in \mathrm{GL}_k(\mathbb{R})$, and

$$Y'_I = Y_I G.$$

Thus Y'_I is invertible if and only if Y_I is.

For $S \in U_I$, define

$$\varphi_I(S) = (YY_I^{-1})_{I^c} \in \mathbb{R}^{(n-k)\times k},$$

where the rows indexed by I^c are also taken in increasing order. This is well-defined, since

$$Y'(Y'_I)^{-1} = YG(Y_I G)^{-1} = YY_I^{-1}.$$

Lemma 5.4 (Chart inverses and covering)**.** *Each map*

$$\varphi_I : U_I \longrightarrow \mathbb{R}^{(n-k)\times k}$$

is a bijection, and the sets U_I cover $\mathrm{Gr}(k, n)$. Hence these maps form a pseudo-atlas on $\mathrm{Gr}(k, n)$, identifying $\mathbb{R}^{(n-k)\times k}$ with $\mathbb{R}^{k(n-k)}$. When $k = n$, this coordinate space consists of a single point.

Proof. For $B \in \mathbb{R}^{(n-k)\times k}$, let $Y_B \in \mathbb{R}^{n\times k}$ be the matrix whose rows indexed by I form I_k, and whose remaining rows form B, in the specified order. Since $(Y_B)_I = I_k$, the matrix Y_B has rank k, and $\mathrm{col}(Y_B) \in U_I$. Moreover,

$$\varphi_I(\mathrm{col}(Y_B)) = B.$$

Conversely, if $S = \mathrm{col}(Y) \in U_I$, the matrix YY_I^{-1} has rows I_k at the indices I and rows $\varphi_I(S)$ at the indices I^c. It is therefore $Y_{\varphi_I(S)}$, and its column space is S. Thus

$$\varphi_I^{-1}(B) = \mathrm{col}(Y_B).$$

Finally, every rank-k matrix $Y \in \mathbb{R}^{n\times k}$ has an invertible $k \times k$ row minor. Hence every $S \in \mathrm{Gr}(k, n)$ belongs to some U_I, proving that these sets cover $\mathrm{Gr}(k, n)$. □

Lemma 5.5 (Smooth transitions)**.** *Let $I, J \subseteq \{1, \dots, n\}$ each have k elements. For $B \in \mathbb{R}^{(n-k)\times k}$, let Y_B denote the basis matrix associated with the chart indexed by I, so that*

$$(Y_B)_I = I_k, \qquad (Y_B)_{I^c} = B.$$

Then

$$\varphi_I(U_I \cap U_J) = \{B \in \mathbb{R}^{(n-k)\times k} : \det((Y_B)_J) \neq 0\}$$

is open. On this set, the transition function is

$$(\varphi_J \circ \varphi_I^{-1})(B) = \left(Y_B((Y_B)_J)^{-1}\right)_{J^c},$$

and is smooth.

Proof. By definition 5.4, $\varphi_I^{-1}(B) = \mathrm{col}(Y_B)$. This subspace belongs to U_J precisely when $(Y_B)_J$ is invertible, giving the stated overlap image.

The entries of Y_B are constant or entries of B. Hence $B \mapsto \det((Y_B)_J)$ is polynomial, so its nonzero set is open.

Applying the definition of φ_J to the basis matrix Y_B gives

$$(\varphi_J \circ \varphi_I^{-1})(B) = \left(Y_B((Y_B)_J)^{-1}\right)_{J^c}.$$

Matrix inversion is smooth on invertible matrices. Matrix multiplication and row extraction are also smooth, so this transition function is smooth on the overlap image. Interchanging I and J gives the same conclusions for the opposite transition. □

Proposition 5.6 (Topology from the pseudo-atlas)**.** *There is a unique topology on $\mathrm{Gr}(k, n)$ for which each U_I is open and*

$$\varphi_I : U_I \longrightarrow \mathbb{R}^{(n-k)\times k}$$

is a homeomorphism.

Proof. By definition 5.4, the pairs (U_I, φ_I) form a pseudo-atlas on $\mathrm{Gr}(k, n)$. By definition 5.5, the overlap images are open and the transition functions are smooth, hence continuous. The conclusion follows from definition 2.11. □

Proposition 5.7 (Projector representation)**.** *For $S \in \mathrm{Gr}(k, n)$, let $Y \in \mathbb{R}^{n\times k}$ be a basis matrix of S. The map*

$$P : \mathrm{Gr}(k, n) \longrightarrow \mathrm{Sym}(n), \qquad P(S) = Y(Y^\top Y)^{-1}Y^\top$$

is well-defined, continuous and injective, and $P(S)$ is the orthogonal projector onto S. Consequently, $\mathrm{Gr}(k, n)$ is Hausdorff. If $X \in \mathrm{St}(n, k)$ is an orthonormal basis matrix of S, then $P(S) = XX^\top$.

Proof. Since Y has rank k, the matrix $Y^\top Y$ is positive-definite and hence invertible. If $Y' = YG$ is another basis matrix of S, with $G \in \mathrm{GL}_k(\mathbb{R})$, then

$$\begin{aligned} Y'(Y'^\top Y')^{-1}Y'^\top &= YG(G^\top Y^\top YG)^{-1}G^\top Y^\top \\ &= Y(Y^\top Y)^{-1}Y^\top. \end{aligned}$$

Thus P is well-defined.

The matrix $P(S)$ is symmetric, its image is contained in S, and

$$P(S)Y = Y.$$

Hence it acts as the identity on S, so its image is exactly S and $P(S)^2 = P(S)$. Therefore $P(S)$ is the orthogonal projector onto S. Since its image determines S, the map P is injective.

In the chart indexed by I, we have

$$(P \circ \varphi_I^{-1})(B) = Y_B(Y_B^\top Y_B)^{-1}Y_B^\top.$$

The entries of Y_B depend affinely on B, and $Y_B^\top Y_B$ is invertible for every B. Matrix multiplication and inversion are continuous, so this expression is continuous. Since the sets U_I form an open cover, P is continuous on $\mathrm{Gr}(k, n)$.

Distinct points of $\mathrm{Gr}(k, n)$ have distinct images under P, which admit disjoint open neighbourhoods in $\mathrm{Sym}(n)$. Their inverse images under P are disjoint open neighbourhoods in $\mathrm{Gr}(k, n)$, proving Hausdorffness.

Finally, if $X \in \mathrm{St}(n, k)$ represents S, then $X^\top X = I_k$, giving

$$P(S) = X(X^\top X)^{-1}X^\top = XX^\top.$$

□

Theorem 5.8 (Smooth structure and dimension)**.** *The pseudo-atlas* $\{(U_I, \varphi_I)\}$ *equips* $\mathrm{Gr}(k, n)$ *with a smooth structure of dimension* $k(n-k)$*. With this structure, the projector map* $P : \mathrm{Gr}(k, n) \longrightarrow \mathrm{Sym}(n)$ *is smooth.*

Proof. By definition 5.6, the sets U_I form an open cover of $\mathrm{Gr}(k, n)$, and each φ_I is a homeomorphism onto $\mathbb{R}^{(n-k)\times k}$. Moreover, $\mathrm{Gr}(k, n)$ is Hausdorff by definition 5.7.

Each U_I has a countable basis transported from $\mathbb{R}^{(n-k)\times k}$ through φ_I. There are only finitely many index sets I, so the union of these bases is a countable basis for $\mathrm{Gr}(k, n)$. Thus $\mathrm{Gr}(k, n)$ is second countable.

By definition 5.5, all transition functions are smooth. Hence definition 2.12 gives a smooth structure on $\mathrm{Gr}(k, n)$, with

$$\dim \mathrm{Gr}(k, n) = \dim \mathbb{R}^{(n-k)\times k} = k(n-k).$$

Finally, in each chart,

$$(P \circ \varphi_I^{-1})(B) = Y_B(Y_B^\top Y_B)^{-1}Y_B^\top.$$

The entries of Y_B depend affinely on B, and $Y_B^\top Y_B$ is invertible for every B. Since matrix multiplication and inversion are smooth, these coordinate expressions are smooth. Therefore P is smooth. □

6 From Stiefel to Grassmann

An element of the Stiefel manifold specifies an orthonormal basis, while an element of the Grassmannian records only the subspace it spans. We relate these manifolds by forgetting the choice of basis.

Definition 6.1 (Column-space map)**.** The **column-space map** is

$$\pi : \mathrm{St}(n,k) \longrightarrow \mathrm{Gr}(k,n), \qquad \pi(X) = \mathrm{col}(X).$$

Proposition 6.2 (Smoothness and fibres)**.** *The map π is smooth and surjective. For $X, X' \in \mathrm{St}(n,k)$,*

$$\pi(X) = \pi(X') \quad \Longleftrightarrow \quad X' = XQ \text{ for some } Q \in \mathrm{O}(k).$$

Thus the fibre over $\pi(X)$ is

$$\pi^{-1}(\pi(X)) = \{XQ : Q \in \mathrm{O}(k)\}.$$

Proof. Every k-dimensional subspace of $\mathbb{R}^n$ has an orthonormal basis, so π is surjective.

For each index set I, we have

$$\pi^{-1}(U_I) = \{X \in \mathrm{St}(n,k) : \det(X_I) \neq 0\},$$

which is open in $\mathrm{St}(n,k)$. On this set,

$$(\varphi_I \circ \pi)(X) = (XX_I^{-1})_{I^c}.$$

This formula extends smoothly to the open subset $\{Y \in \mathbb{R}^{n\times k} : \det(Y_I) \neq 0\}$ of the ambient matrix space. Hence each coordinate expression is smooth. Since the sets U_I cover $\mathrm{Gr}(k,n)$, the map π is smooth.

Suppose $\pi(X) = \pi(X')$. By definition 5.2, there is an invertible matrix Q such that $X' = XQ$. Since both matrices have orthonormal columns,

$$I_k = X'^{\top}X' = Q^{\top}X^{\top}XQ = Q^{\top}Q.$$

Thus $Q \in \mathrm{O}(k)$. Conversely, if $X' = XQ$ with $Q \in \mathrm{O}(k)$, then Q is invertible, so $\mathrm{col}(X') = \mathrm{col}(X)$. The description of the fibre follows. □

Lemma 6.3 (Local sections)**.** *Fix an index set I, and for $B \in \mathbb{R}^{(n-k)\times k}$, let Y_B be the matrix with $(Y_B)_I = I_k$ and $(Y_B)_{I^c} = B$. The map*

$$s_I : \mathbb{R}^{(n-k)\times k} \longrightarrow \mathrm{St}(n,k), \qquad s_I(B) = Y_B(Y_B^{\top}Y_B)^{-1/2}$$

is smooth and satisfies

$$\pi \circ s_I = \varphi_I^{-1}.$$

Consequently, $\sigma_I = s_I \circ \varphi_I : U_I \longrightarrow \mathrm{St}(n,k)$ is a smooth local section of π, meaning that $\pi \circ \sigma_I = \mathrm{id}_{U_I}$.

Proof. Since Y_B has rank k, the matrix $A_B = Y_B^\top Y_B$ is positive-definite. By definition 4.8, the map $B \mapsto A_B^{-1/2}$ is smooth. Hence s_I is smooth as a map into $\mathbb{R}^{n\times k}$. Moreover,

$$s_I(B)^\top s_I(B) = A_B^{-1/2} A_B A_B^{-1/2} = I_k,$$

so its image lies in $\mathrm{St}(n,k)$. By the smoothness criterion for maps into embedded submanifolds, s_I is smooth as a map into $\mathrm{St}(n,k)$.

Since $A_B^{-1/2}$ is invertible, multiplication by it does not change the column space. Therefore

$$\pi(s_I(B)) = \mathrm{col}(Y_B) = \varphi_I^{-1}(B).$$

Composing with φ_I gives $\pi \circ \sigma_I = \mathrm{id}_{U_I}$, and σ_I is smooth as a composition of smooth maps. □

Theorem 6.4 (Submersion and quotient)**.** *The column-space map $\pi : \mathrm{St}(n,k) \longrightarrow \mathrm{Gr}(k,n)$ is a surjective submersion, and $\mathrm{Gr}(k,n)$ is compact. The induced map*

$$\overline{\pi} : \mathrm{St}(n,k)/\mathrm{O}(k) \longrightarrow \mathrm{Gr}(k,n), \qquad [X] \longmapsto \mathrm{col}(X),$$

is a homeomorphism, where $\mathrm{St}(n,k)/\mathrm{O}(k)$ is the set of orbits under right multiplication by $\mathrm{O}(k)$, equipped with the quotient topology.

Proof. Smoothness and surjectivity follow from definition 6.2. Let $X \in \mathrm{St}(n,k)$, set $S = \pi(X)$, and choose I such that $S \in U_I$. By definition 6.3, there is a smooth local section $\sigma_I : U_I \longrightarrow \mathrm{St}(n,k)$ satisfying $\pi \circ \sigma_I = \mathrm{id}_{U_I}$. Differentiating at S gives

$$D\pi(\sigma_I(S)) \circ D\sigma_I(S) = \mathrm{id}_{T_S \mathrm{Gr}(k,n)}.$$

Hence $D\pi(\sigma_I(S))$ is surjective.

Since X and $\sigma_I(S)$ span the same subspace, definition 6.2 gives $X = \sigma_I(S)Q$ for some $Q \in \mathrm{O}(k)$. Right multiplication

$$\rho_Q : \mathrm{St}(n,k) \longrightarrow \mathrm{St}(n,k), \qquad \rho_Q(Y) = YQ,$$

is a diffeomorphism with inverse $\rho_{Q^\top}$, and $\pi \circ \rho_Q = \pi$. The chain rule therefore gives

$$D\pi(X) \circ D\rho_Q(\sigma_I(S)) = D\pi(\sigma_I(S)).$$

Since the right-hand side is surjective, $D\pi(X)$ is surjective. Thus π is a submersion.

By definition 4.4, $\mathrm{St}(n,k)$ is compact. Since π is continuous and surjective, $\mathrm{Gr}(k,n)$ is compact.

Let $q : \mathrm{St}(n,k) \longrightarrow \mathrm{St}(n,k)/\mathrm{O}(k)$ be the quotient map. By definition 6.2, the fibres of π are precisely the orbits, so $\overline{\pi}$ is well-defined and bijective. Since $\pi = \overline{\pi} \circ q$ is continuous, the definition of the quotient topology implies that $\overline{\pi}$ is continuous. The space $\mathrm{St}(n,k)/\mathrm{O}(k)$ is compact as the continuous image of $\mathrm{St}(n,k)$ under q, and $\mathrm{Gr}(k,n)$ is Hausdorff by definition 5.7. Therefore $\overline{\pi}$ is a homeomorphism. □

Each orbit $\{XQ : Q \in \mathrm{O}(k)\}$ consists of all orthonormal bases of one subspace. Thus passing from the Stiefel manifold to the Grassmannian forgets the choice of orthonormal basis.

7 Applications

The Stiefel manifold provides a constraint set for recurrent neural networks with orthogonal recurrent weights. Consider the recurrence

$$h_t = \sigma(Wh_{t-1} + Ux_t + b),$$

where $x_t \in \mathbb{R}^d$ is the input, $h_t \in \mathbb{R}^n$ is the hidden state, $U \in \mathbb{R}^{n\times d}$ is the input weight matrix, $b \in \mathbb{R}^n$ is a bias, and σ is a smooth activation function applied componentwise. Requiring the recurrent matrix $W \in \mathbb{R}^{n\times n}$ to satisfy $W^\top W = I_n$ places it in $\mathrm{St}(n, n) = \mathrm{O}(n)$. The linear transformation $h \mapsto Wh$ then preserves Euclidean norms, although the full recurrence need not do so because of the activation and input terms.

Let L be a smooth training loss, and let $G = \nabla\bar{L}(W)$ denote its ambient gradient with respect to W, holding the other parameters fixed. Backpropagation through time computes this gradient by differentiating through the recurrence. By definition 4.7, the corresponding Riemannian gradient is

$$\operatorname{grad} L(W) = G - W\operatorname{sym}(W^\top G),$$

following [1, Section 7.3, equation (7.28)]. Since $WW^\top = I_n$, this can also be written as

$$\operatorname{grad} L(W) = W\operatorname{skew}(W^\top G), \qquad \operatorname{skew}(B) := \frac{B - B^\top}{2}.$$

Thus the ambient gradient determines a tangent direction compatible with the orthogonality constraint. A retraction converts a step in this direction into another orthogonal matrix, as established for the polar retraction in definition 4.9.

These results provide the mathematical foundation for a planned NumPy implementation of Riemannian gradient descent for orthogonal recurrent networks. The implementation will use a Cayley retraction, whose construction and verification will be treated separately from the polar retraction proved here. Backpropagation through time will be derived explicitly, and experiments across multiple random seeds will compare hard orthogonality constraints with an unconstrained objective containing a soft penalty of the form

$$\bar{L}(W) + \lambda\|W^\top W - I_n\|_F^2, \qquad \lambda > 0.$$

Such a penalty discourages departures from orthogonality without requiring the constraint to hold exactly. The experimental comparison is future work; no claim of improved training performance follows from the geometric results alone.

The Grassmannian clarifies the distinction between a matrix and the subspace it spans. In this application, the recurrent transformation depends on W itself, while every $W \in \mathrm{O}(n)$ spans $\mathbb{R}^n$. Hence $\mathrm{Gr}(n, n)$, which consists of a single point, does not parametrise the recurrent weights.

8 Discussion

The relationship between the Stiefel and Grassmann manifolds expresses the distinction between an orthonormal frame and the subspace it spans. The column-space map identifies

precisely those frames that differ by right multiplication by an orthogonal matrix. By definition 6.4, this identification gives a homeomorphism

$$\mathrm{St}(n,k)/\mathrm{O}(k) \cong \mathrm{Gr}(k,n),$$

and the column-space map is a smooth surjective submersion. The explicit Grassmann charts provide the smooth structure used in this argument, while the projector representation describes each subspace independently of a chosen basis.

The Stiefel calculations also explain how the orthogonality constraint can be respected during optimisation. The tangent projection converts an ambient gradient into the Riemannian gradient for the induced Frobenius metric. The polar retraction then provides a way to return a tangent step to the manifold while preserving its direction to first order. These results establish feasible updates, but do not by themselves guarantee that a chosen step decreases the loss or that an iteration converges. They provide a mathematical starting point for the subsequent implementation, whose update rule and training behaviour require separate analysis and evaluation.

9 Conclusion

We have constructed the Stiefel and Grassmann manifolds and established their relationship through the column-space map. The Stiefel manifold records orthonormal frames, while the Grassmannian records their spans. The identification $\mathrm{St}(n,k)/\mathrm{O}(k) \cong \mathrm{Gr}(k,n)$ makes this distinction precise: passing to the quotient removes the choice of orthonormal basis.

The tangent projection, Riemannian gradient and polar retraction derived for Stiefel provide tools for optimisation under orthogonality constraints. Together with the geometric constructions, these results form a foundation for the planned machine learning project on orthogonal recurrent networks. The implementation and its experimental evaluation remain subsequent work.

References

[1] Nicolas Boumal, *An introduction to optimization on smooth manifolds*, Cambridge University Press, 2023.

[2] Ed Segal, *Manifolds*, 2016, Lecture notes, University College London.