The Classical and Symplectic Stiefel and Grassmann Manifolds

Geometry and Applications

Bendokat, Thomas

DOI:
10.21996/mmer-zp86

Publication date:
2021

Document version:
Final published version

SDU

The Classical and Symplectic Stiefel and Grassmann Manifolds

Geometry and Applications

PhD thesis by

Thomas Bendokat

Department of Mathematics and Computer Science (IMADA)
University of Southern Denmark (SDU)
Odense
Denmark

Supervisor: Associate Professor Ralf Zimmermann

September 30, 2021

Abstract

Abstract Geodesics, the related exponential map, and its inverse, or approximations of the latter two, lie at the heart of most data processing operations on matrix manifolds. When (efficient) formulas are available, it becomes possible to tackle tasks such as interpolation or optimization problems. This thesis solves related problems on the classical and symplectic Stiefel and Grassmann manifolds, i.e., the manifolds of linear or symplectic bases or subspaces of vector spaces, respectively, where the latter—the symplectic Grassmann manifold—is introduced as a new object of study in this thesis.

For the classical Grassmann manifold, an algorithm to compute the Riemannian logarithm is derived, which is numerically efficient and, as a new result, allows the computation of geodesics between any two given points. Furthermore, research tracks treating the Grassmann manifold as a set of projectors and as a quotient of the Stiefel manifold are combined and formulas for an easy transition between the two are stated. At last, a full description of the conjugate locus is found.

For the classical Stiefel manifold, an efficient method to compute a known kind of quasi-geodesics between two given points is found and an alternative kind of connecting quasi-geodesics, much closer to the Riemannian geodesics, is introduced.

For the symplectic Stiefel manifold, a pseudo-Riemannian and Riemannian framework are introduced, which allow for the computation of (for any metric hitherto unknown) geodesics. For the newly introduced symplectic Grassmann manifold, pseudo-Riemannian and Riemannian metrics and corresponding geodesics, with efficient formulas via horizontal lifts, are found in a similar fashion. On both the symplectic Stiefel and Grassmann manifold, efficient formulas to compute and invert the Cayley retraction are introduced.

Resumé Geodæter, den relaterede eksponentialafbildning og dens inverse, eller tilnærmelser til de to sidstnævnte, er centrale i de fleste databehandlingsoperationer på matrixmangfoldigheder. Når (effektive) formler er tilgængelige bliver det muligt at løse opgaver som interpolations- eller optimeringsproblemer. Denne afhandling løser problemer relateret hertil på de klassiske og symplektiske Stiefel- og Grassmannmangfoldigheder, dvs. mangfoldigheder af henholdsvis lineære eller symplektiske baser eller underrum af vectorrum, hvor sidstnævnte — den symplektiske Grassmannmangfoldighed — introduceres som et nyt undersøgelsesobjekt i denne afhandling.

For den klassiske Grassmannmangfoldighed afledes an algoritme til beregning af den Riemannske logaritme, som er numerisk effektiv og som et nyt resultat muliggør beregning af geodæter mellem ethvert par af to givne punkter. Desuden kombineres forskningsspor, der behandler Grassmannmangfoldigheden som et sæt projektorer og som en kvotient af Stiefelmangfoldigheden, og formler til en let overgang mellem de to angives. Endelig findes en fuldstændig beskrivelse af det konjugerede locus.

For den klassiske Stiefelmangfoldighed findes en effektiv metode til at beregne en kendt slags kvasi-geodæter mellem to givne punkter, og der indføres en alternativ form for forbindende kvasi-geodæter, som er meget tættere på de Riemannske geodæter.

For den symplektiske Stiefelmangfoldighed introduceres et pseudo-Riemannsk og Riemannsk framework, som muliggør beregning af (for enhver metrik hidtil ukendte) geodæter. For den nyindførte symplektiske Grassmannmangfoldighed findes pseudo-Riemannske og Riemannske metrikker og tilsvarende geodæter med effektive formler via horisontale løft på en lignende måde. På både den symplektiske Stiefelmangfoldighed og den symplektiske Grassmannmangfoldighed introduceres effektive formler til beregning og invertering af Cayleyretraktionen.

Preface

This thesis is the result of my work as a PhD student at the Department of Mathematics and Computer Science (IMADA) at the University of Southern Denmark in Odense between October 2018 and September 2021. It has been completed under the supervision of Associate Professor Ralf Zimmermann.

The main objects of interest of this thesis are so called *Stiefel* and *Grassmann* manifolds. Stiefel manifolds, for our purpose, consist of tall skinny matrices containing a fixed number of basis vectors as their columns, which span subspaces of a given vector space. Grassmann manifolds, in turn, are manifolds consisting of subspaces of a fixed dimension. As every subspace can be spanned by a whole range of different bases, every element of a Grassmann manifold intuitively corresponds to a whole equivalence class of matrices in a Stiefel manifold. This thesis treats problems on the classical Stiefel and Grassmann manifolds, corresponding to linear subspaces, and on the symplectic Stiefel and Grassmann manifolds, corresponding to symplectic subspaces.

After an introduction to the necessary Lie group theory and a motivation from parametric model order reduction, the thesis consists of two parts.

The first part treats the classical Stiefel and Grassmann manifolds and contains two papers. The first paper, titled “A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects” originated in my research visit of Professor Pierre-Antoine Absil at UCLouvain from March to June 2019. It provides an overview over some of the most relevant aspects of the Grassmann manifold, i.e., the manifold of fixed dimensional linear subspaces of the real vector space $\mathbb{R}^n$, with regards to applications. The paper combines the research track treating the Grassmann manifold as a set of projectors, represented by $n \times n$ matrices, with the research track treating it as a quotient of the Stiefel manifold, i.e., with tall skinny matrix representatives, and derives formulas for transitioning between the two approaches. It furthermore introduces a modified algorithm to compute the Riemannian logarithm, i.e., the tangent direction of the shortest curve between two given points. As a novelty, this algorithm also allows to connect points in the cut locus, i.e., in the case where there are multiple shortest curves, which is important for so-called “almost gradients”. The paper ends with explicit formulas for Jacobi fields vanishing at one point and a complete treatment of the conjugate locus of the Grassmann manifold. It has been submitted to *Advances in Computational Mathematics (ACOM)* and is currently under review. The current arXiv version can be found under [BZA20].

The second paper, titled “Efficient Quasi-Geodesics on the Stiefel Manifold”, treats so called quasi-geodesics between two given points on the Stiefel manifold. Quasi-geodesics are curves that feature constant speed and constant-norm covariant acceleration. A special case are geodesics, which feature vanishing covariant acceleration and are locally shortest curves. For a known type of quasi-geodesics, the paper introduces an efficient representation, making use of the above-mentioned Grassmann logarithm. This makes it possible to apply the quasi-geodesics in computations for high-dimensional problems. The paper furthermore introduces a new type of quasi-geodesics connecting two given points. In numerical experiments, these turn out to be much closer to the Riemannian geodesics and can also be applied to high-dimensional problems at a slightly higher computational cost. This paper was awarded the best paper award at the Geometric Science of Information conference (GSI21) and is published in [BZ21a].

The second part, containing the third paper, treats the symplectic Stiefel and Grassmann manifolds. The difference to the classical case is that the subspaces treated here are *symplectic*, i.e., subspaces for which the symplectic form of the surrounding symplectic space restricts to a symplectic form on the subspace itself. Symplectic spaces are fundamental in Hamiltonian mechanics. Symplectic subspaces feature therefore in a number of applications, most notably in model order reduction of Hamiltonian systems. For the known symplectic Stiefel manifold of symplectic basis matrices, the paper introduces a novel pseudo-Riemannian and Riemannian metric via a quotient manifold approach with the real symplectic group as the total space. It derives (hitherto for any metric unknown) geodesics for both metrics, efficient ways to compute them and an efficient, invertible formulation of Cayley retractions. The paper furthermore introduces the symplectic Grassmann manifold of fixed dimensional symplectic subspaces and derives geodesics for a pseudo-Riemannian and Riemannian metric on this manifold as well. The pseudo-Riemannian geodesics turn out to be invertible in closed form. Via the method of a horizontal lift, efficient formulas to compute the geodesics and invertible Cayley retractions are introduced as well. The derived Riemannian methods are compared numerically to the state of the art on gradient descent problems. In these experiments, the retraction-based methods turn out to converge either at a speed comparable to the state of the art methods, or faster, depending on the problem. This paper has been submitted to the *SIAM Journal on Matrix Analysis and Applications (SIMAX)* and is currently under review. The current arXiv version can be found under [BZ21b].

A bibliography containing all used references concludes the thesis.

The layout of the papers has been adapted to fit the thesis. Notational conventions may vary slightly between the papers.

Acknowledgements

First and foremost, I would like to thank my supervisor Associate Professor Ralf Zimmermann for guiding me through the PhD process, for introducing me to the numerical aspects of differential geometry, and for a productive collaboration. Thank you also for supporting me outside of the university, such as when I moved.

Furthermore, I would like to thank Professor Pierre-Antoine Absil for hosting my research visit at UCLouvain in Belgium and for enabling a fruitful collaboration.

I would also like to thank my colleagues at IMADA, especially those that I shared an office with (when circumstances permitted), for making the days at the university more enjoyable and for introducing me to the very danish hobby of badminton. Thank you also to my fellow PhD students at UCLouvain for many interesting conversations and board game nights.

Last but not least, I would like to thank my friends and family, and especially Katharina, for their constant support and for keeping me sane during the pandemic. I am not sure how this thesis would have turned out without you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Riemannian and Pseudo-Riemannian Homogeneous Spaces

This section recaps the basics from quotient geometry with Riemannian and pseudo-Riemannian metrics that are used in this thesis. Quotient geometry is to be understood here as differential geometry of homogeneous spaces, i.e., quotients of Lie groups. The main sources used in this section are [Lee12, Chapters 7, 8 & 21] and [ONe83, Chapter 11], unless otherwise noted. A brief introduction to this topic is also given in Appendix 2.A of Chapter 2. This section can be safely skipped by any reader familiar with the topic.

Smooth Manifolds The central objects in differential geometry are smooth manifolds. A proper introduction of smooth manifolds would lead too far here, but can for example be found in [Lee12; ONe83; KN96].

Intuitively, smooth manifolds are topological spaces which locally "look like" $\mathbb{R}^n$, i.e., are locally mapped to $\mathbb{R}^n$ by so called charts, where n is called the *dimension* of the manifold. To be more precise, a smooth manifold is a *topological manifold* together with a *smooth structure*. By definition, a topological manifold is a second countable Hausdorff space of which each point has an open neighborhood that is homeomorphic to an open subset of $\mathbb{R}^n$. To make sense of the term "smooth structure", the local homeomorphisms, called charts, are used. As differentiability and smoothness are well defined for maps from $\mathbb{R}^n$ to $\mathbb{R}^n$, smoothness of the manifold is inferred from the transition maps $\phi \circ \psi^{-1} \colon \psi(U \cap V) \subset \mathbb{R}^n \to \mathbb{R}^n$, where ϕ and ψ are two charts defined on open neighborhoods U and V with non-empty intersection. A smooth structure is now a maximal smooth atlas, i.e., a set of charts whose domains cover the entire manifold, for which all transition maps of the charts are diffeomorphism, and which contains every possible chart for which the transition map with respect to all other charts is a diffeomorphism, if the intersection of domains is non-empty. The charts in a smooth structure are called *smooth charts*. For a detailed introduction, see for example [Lee12, Chapter 1].

For a smooth manifold M, a function $f \colon M \to \mathbb{R}$ is called smooth if $f \circ \psi^{-1}$ is smooth for all smooth charts ψ, and the set of all such functions is denoted by $C^\infty(M)$. In general, a map between two smooth manifolds $f \colon M \to N$ is called smooth if it is smooth under all smooth charts, i.e., for all points $p \in M$ there is a smooth chart ψ around p and a smooth chart ϕ around $F(p)$ such that $\phi \circ f \circ \psi^{-1}$ is smooth as a map between two real vector spaces. This property is independent of the chosen charts.

Associated with each point p on a smooth manifold M is a vector space T_pM called the *tangent space*. The tangent space is of the same dimension as the manifold and can be thought of as containing all directions in which it is possible to move from the point p in M. Technically, the elements of the tangent space T_pM, called *tangent vectors*, are linear maps $v\colon C^\infty(M) \to \mathbb{R}$ which fulfill the product rule

$$v(fg) = f(p)v(g) + g(p)v(f)$$

for all $f, g \in C^\infty(M)$ [Lee12, Chapter 3, p. 54], where the product of functions is defined point-wise. In this way, a tangent vector $v \in T_pM$ acts as the directional derivative of a function $f \in C^\infty(M)$ at a point $p \in M$ in the direction v. The disjoint union of all tangent spaces of M is called the *tangent bundle*, denoted by TM. Every point in the tangent bundle TM is therefore a tuple of a point $p \in M$ and a tangent vector $v \in T_pM$.

The notion of tangent spaces makes it possible to define the *differential* of a smooth map $f\colon M \to N$ between two smooth manifolds M and N. At each point $p \in M$, it is defined as the map

$$\mathrm{d}f_p\colon T_pM \to T_{f(p)}N$$

fulfilling

$$\mathrm{d}f_p(v)(\tilde{f}) = v(\tilde{f} \circ f)$$

for all $\tilde{f} \in C^\infty(N)$ [Lee12, Chapter 3, p. 55]. For a smooth function $f \in C^\infty(M)$, this reduces to $\mathrm{d}f_p(v) = v(f)$ for $v \in T_pM$.

A *vector field* on M is a continuous map $X\colon M \to TM$ which assigns a tangent vector to every point of the manifold, i.e., $X_p := X(p) \in T_pM$ for all $p \in M$. When TM is equipped with the canonical smooth manifold structure from [Lee12, Proposition 3.18], a vector field is called a *smooth vector field* if it is smooth as a map, and the set of all smooth vector fields on M is denoted by $\mathfrak{X}(M)$. For details of these constructions, see for example [Lee12, Chapter 8]. An *integral curve* of a vector field $X \in \mathfrak{X}(M)$ is a curve $\gamma\colon I \to M$ on an open interval I, such that the *velocity* of $\gamma(t)$ is equal to vector field X at that point, i.e., $\dot{\gamma}(t) = X_{\gamma(t)}$ for all $t \in I$ [Lee12, Chapter 9, p. 206]. In coordinates, $\dot{\gamma}(t)$ is given by the usual derivative of the components with respect to t. It can be shown that every tangent vector $v \in T_p$ can be identified with a derivative of a smooth curve through the base point p [Lee12, Proposition 3.23].

Lie Groups A group G which is also a smooth manifold and for which the multiplication and inversion maps are both smooth is called a *Lie group*.

One of the most important properties of Lie groups with respect to quotient manifolds is the fact that Lie groups can *act* on manifolds. In the following, let G be a Lie group and M a smooth manifold. A *left action* of G on M is defined as a map $\theta\colon G \times M \to M$ which fulfills for all $g_1, g_2 \in G$ and $p \in M$ that

$$\theta(g_1, \theta(g_2, p)) = \theta(g_1 g_2, p),$$
$$\theta(e, p) = p,$$

where e denotes the identity element of the group. A *right action* is analogously a map $\theta\colon M \times G \to M$ with

$$\begin{aligned}\theta(\theta(p, g_1), g_2) &= \theta(p, g_1 g_2),\\ \theta(p, e) &= p,\end{aligned}$$

for all $g_1, g_2 \in G$ and $p \in M$, with identity element e [Lee12, p. 161]. Let $\theta\colon G \times M \to M$ be a left action. We denote the *orbit* of $p \in M$, i.e., the set of all images of p under the group action, by

$$\theta(G, p) = \{\theta(g, p) \mid g \in G\}$$

and the *stabilizer* of $p \in M$, i.e., the group elements that leave p invariant, by

$$\mathrm{stab}_p = \{g \in G \mid \theta(g, p) = p\}.$$

These definitions hold analogously for right actions. If the orbit $\theta(G, p)$ is equal to M for all $p \in M$, the action is called *transitive*. Equivalently, a transitive action is an action such that for all $p, q \in M$ there is $g \in G$ which fulfills $\theta(g, p) = q$ [Lee12, p. 162].

A Lie group acts transitively on itself by left and right translation, i.e., by $L_g(h) = gh$ and $R_g(h) = hg$ for all $g, h \in G$. A vector field X on G is called *left-invariant* if he equation $\mathrm{d}(L_g)_h(X_h) = X_{gh}$ holds for all $g, h \in G$ [Lee12, p. 189], and *right-invariant* accordingly for the right translation. As left translation is transitive, a left-invariant vector field is fully determined by its value at the identity element $X_e \in T_eG$.

Associated with a Lie group is a vector space called the *Lie algebra*. In general, a Lie algebra over the real numbers $\mathbb{R}$ is a real vector space $\mathfrak{g}$ with a map $[\cdot,\cdot]\colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called *bracket*, which fulfills for all $X, Y, Z \in \mathfrak{g}$ and for all $\alpha \in \mathbb{R}$ [Lee12, p. 190]:

1. $[X, Y] = -[Y, X]$ (antisymmetry),
2. $[\alpha X + Y, Z] = \alpha[X, Z] + [Y, Z]$ (linearity, and by antisymmetry bilinearity) and
3. $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ (Jacobi identity).

By [Lee12, Theorem 8.37], the vector space of left invariant vector fields on a Lie group G is a Lie algebra with the same dimension as G, called the *Lie algebra of* G, and can be identified with the tangent space at the identity element $T_eG = \mathfrak{g}$.

A subgroup of a Lie group is called *Lie subgroup*, if it is a so-called immersed submanifold, see [Lee12, Chapters 5 and 7] for definitions and details. The closed-subgroup theorem [Lee12, Theorem 20.12] states that any closed subgroup of a Lie group is a Lie subgroup.

The Exponential Map of a Lie Group The most important connection between a Lie group and its Lie algebra is given via the *exponential map*. An introduction can be found in [Lee12, Chapter 20], from which the following definitions are adapted.

Let G be a Lie group and $\mathfrak{g}$ the corresponding Lie algebra. The exponential map $\exp\colon \mathfrak{g} \to G$ maps any $X \in \mathfrak{g}$ to the evaluation at $t = 1$ of the integral curve of the left-invariant vector field defined by X, starting at the identity. In other words, $t \mapsto \exp(tX)$ solves the differential equation $\dot{\gamma}(t) = \mathrm{d}(L_{\gamma(t)})_e(X)$ with initial value $\gamma(0) = e$. The map $\gamma_X\colon \mathbb{R} \to G$, defined by $\gamma_X(t) = \exp(tX)$, is called the *one-parameter subgroup of G generated by X*. In case of the general linear group $\mathrm{GL}(n, \mathbb{R})$ of invertible matrices, the exponential map is given by the matrix exponential

$$\exp_{\mathrm{m}}(X) := \sum_{k=0}^{\infty} \frac{X^k}{k!},$$

which is shown in [Lee12, Proposition 20.2]. By [Lee12, Proposition 20.9], this also holds for all (real) matrix Lie groups, as they are Lie subgroups of $\mathrm{GL}(n, \mathbb{R})$.

Homogeneous Spaces A *homogeneous space* is a smooth manifold with a transitive (smooth) Lie group action. Trivial examples are Lie groups, as they act transitively on themselves. But not only Lie groups, all manifolds treated in this thesis are homogeneous spaces.

For a Lie group G and a closed subgroup H of G, a *left coset* of H is a set

$$gH := \{gh \mid h \in H\} \subset G,$$

where $g \in G$. The *left coset space G/H*, read G modulo H, is defined as the set of all left cosets together with the quotient topology [Lee12, Chapter 21, p. 551].

One central theorem of homogeneous space theory is the *Homogeneous Space Characterization Theorem* [Lee12, Theorem 21.18]. It states that any homogeneous space M is diffeomorphic to the left coset space of the Lie group G, acting transitively on M, modulo the stabilizer stab_p for any $p \in M$. The other central theorem is the *Homogeneous Space Construction Theorem* [Lee12, Theorem 21.17]. This theorem states that any left coset space G/H is a topological manifold of dimension $\dim G/H = \dim G - \dim H$ with a unique smooth structure such that the quotient map $\pi\colon G \to G/H$ is a smooth submersion.

On a Lie group G with Lie algebra $\mathfrak{g}$, for every $g \in G$ we define a map $C_g\colon G \to G$ by $C_g(h) = ghg^{-1}$, called *conjugation map*. Its differential $\mathrm{Ad}_g = \mathrm{d}(C_g)_e\colon \mathfrak{g} \to \mathfrak{g}$ is called the *adjoint representation* of G at g [ONe83, p. 301]. If G is a matrix Lie group, the adjoint representation is given by $\mathrm{Ad}_g(X) = gXg^{-1}$. Let H be a Lie subgroup of G. A subset $\mathfrak{m} \subset \mathfrak{g}$ is called $\mathrm{Ad}(H)$*-invariant* if $\mathrm{Ad}_h(m) \in \mathfrak{m}$ for all $h \in H$ and all $m \in \mathfrak{m}$ [ONe83, p. 301].

A homogeneous space G/H is called *reductive*, if the Lie algebra $\mathfrak{g}$ of G decomposes into the direct sum

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

where $\mathfrak{h}$ is the Lie algebra of H and $\mathfrak{m}$ is an $\mathrm{Ad}(H)$-invariant subspace of $\mathfrak{g}$ [ONe83, Definition 11.21]. For a reductive homogeneous space, the subspace $\mathfrak{m}$ can be identified with the tangent space $\mathfrak{m} \cong T_{\pi(e)}G/H$ [ONe83, Chapter 11, p. 311], similar to how a Lie algebra can be identified with the tangent space at the identity of the Lie group.

Homogeneous spaces are of particular interest in applications, since any set with a transitive Lie group action has a unique manifold structure for which it is a homogeneous space, if the stabilizer at some point is closed in the Lie group [Lee12, Theorem 21.20]. This implies that any set which is the orbit of a Lie group action is a homogeneous space if the stabilizer is closed in the Lie group. In practical applications, the Lie group will usually be a matrix Lie group acting on matrices by some form of multiplication, and the orbit will be given by the action of all Lie group elements on a fixed matrix.

Pseudo-Riemannian and Riemannian Metrics An important structure with which a smooth manifold can be equipped is a *(pseudo-)Riemannian metric*. It is defined as a (not necessarily positive definite) scalar product on the tangent spaces that is smoothly changing with the base point, and in case of *Riemannian metrics* is an inner product, i.e., positive definite. An introduction to pseudo-Riemannian (also called semi-Riemannian) geometry can be found in [ONe83], while the special but important case of Riemannian geometry is for example treated in [Lee18; KN96].

Let M be a smooth manifold. A *pseudo-Riemannian metric* is a smooth mapping h that at every $p \in M$ is defined as a mapping $h_p \colon T_pM \times T_pM \to \mathbb{R}$ with the properties

- $h_p(u, v) = h_p(v, u)$ (symmetric)
- $h_p(\alpha u + v, w) = \alpha h_p(u, w) + h_p(v, w)$ (linear, and by symmetry bilinear)
- $h_p(u, v) = 0$ for all $u \in T_pM$ implies $v = 0$ (non-degenerate)

for all $u, v, w \in T_pM$ and $\alpha \in \mathbb{R}$.

For a *Riemannian metric*, which we denote by g, the property of being non-degenerate is strengthened to

- $g_p(u, u) \geq 0$ for all $u \in T_pM$ with equality if and only if $u = 0$ (positive definite).

For a function $f \in C^\infty(M)$, a (pseudo-)Riemannian metric h defines a *gradient* $\mathrm{grad}_f \in \mathfrak{X}(M)$. This gradient is defined at every $p \in M$ as the unique tangent vector $\mathrm{grad}_f(p)$ fulfilling

$$h_p(\mathrm{grad}_f(p), v) = \mathrm{d}f_p(v) = v(f)$$

for all $v \in T_pM$ [ONe83, Definition 3.47]. If the metric is Riemannian, then the gradient is equal to the steepest-ascent direction of f at p [AMS08, Chapter 3, p. 46], and the negative gradient is accordingly the steepest-descent direction. This comes from the fact that for an inner product such as a Riemannian metric g_p at a point p, it is possible to define an *angle* between two nonzero tangent vectors $x, y \in T_pM$ as the unique $\phi \in [0, \pi]$ such that

$$\cos(\phi) = \frac{g_p(x, y)}{||x||||y||},$$

see [Lee18, Chapter 2, p. 10]. Here the norm $||\cdot||$ is defined as $\sqrt{g_p(\cdot,\cdot)}$. This makes Riemannian metrics attractive for optimization problems in which a function is to be minimized.

A diffeomorphism $\phi\colon M \to N$ between to pseudo-Riemannian manifolds (M, h) and $(N, \tilde{h})$ that preserves the pseudo-Riemannian metrics in the sense that

$$h_p(x, y) = \tilde{h}_{\phi(p)}(\mathrm{d}\phi_p(x), \mathrm{d}\phi_p(y))$$

for all $x, y \in T_pM$ and all $p \in M$ is called an *isometry* [ONe83, Definition 3.6].

With the help of (pseudo-)Riemannian metrics it is possible to define the *arc length* of a smooth curve $\gamma\colon [a, b] \to M$ on a manifold M by

$$L(\gamma) := \int_a^b |h_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))|^{\frac{1}{2}}\,\mathrm{d}t,$$

where h denotes the (pseudo-)Riemannian metric and $\dot{\gamma}(t)$ the tangent vector of γ at t [ONe83, Definition 5.11]. The arc-length is a length in the intuitive sense only for Riemannian metrics, since a tangent vector $v \in T_pM$ can have a norm of $h_p(v, v) = 0$ for a pseudo-Riemannian metric. In the Riemannian case it is therefore possible to define the Riemannian distance between two points as the length of the shortest curve between them, or, to be precise, the greatest lower bound of the length of all curves between the points [ONe83, Definition 5.15]. The latter is also defined when a shortest curve does not exist.

Another important concept in (pseudo-)Riemannian geometry is the concept of *connections*. As the name suggests, they connect different tangent spaces of a manifold by defining a way in which one vector field changes with respect to another vector field. While there are many connections that can be defined on the same (pseudo-)Riemannian manifold, it turns out that there is exactly one that features so called *compatibility with the metric* and is *free of torsion*. This connection is called the *Levi-Civita connection* and is characterized by the following theorem.

Theorem 1.1.1 ([ONe83, Theorem 3.11]). On a semi-Riemannian manifold (M, h), there is a unique connection ∇, i.e., a function $\nabla\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ with

1. $\nabla_V W$ is $C^\infty(M)$-linear in the V-argument,
2. $\nabla_V W$ is $\mathbb{R}$-linear in the W-argument,
3. $\nabla_V(fW) = (Vf)W + f\nabla_V W$ for all $f \in C^\infty(M)$ (product rule),

which additionally fulfills

4. $[V, W] = \nabla_V W - \nabla_W V$ (vanishing torsion) and
5. $Xh(V, W) = h(\nabla_X V, W) + h(V, \nabla_X W)$ (compatibility with the metric),

where $[\cdot, \cdot]$ denotes the Lie bracket of two vector fields. This connection is called the *Levi-Civita connection.*

The differentiation of a function $f \in C^\infty(M)$ in direction of a vector field $X \in \mathfrak{X}(M)$ is also sometimes denoted by $\nabla_X f := Xf$, motivated by the extension of connections to tensor bundles [Lee18, Proposition 4.15].

The Levi-Civita connection ∇ induces the so called *Riemannian curvature tensor* $R\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$,

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

Note that R is sometimes defined with the opposite sign, such as in [ONe83, Lemma 3.25]. If the Riemannian curvature tensor is equal to zero at every point, the manifold is called *flat*. It turns out that R is fully determined by a simpler function, called the *sectional curvature*. Let $x, y \in T_pM$ be two linearly independent tangent vectors at $p \in M$. Then x and y span a plane in T_pM, and the sectional curvature defined by

$$K_p(x, y) = \frac{h_p(R(x, y)y, x)}{h_p(x, x)h_p(y, y) - h_p(x, y)^2}$$

depends only on the plane, not on the chosen vectors, see [ONe83, Lemma 3.39]. Here $R(x, y)z$ denotes the evaluation of the Riemannian curvature tensor at the tangent vectors $x, y, z \in T_pM$, see [ONe83, Chapter 2, p. 38] for details. The fact that the sectional curvature fully determines the Riemannian curvature tensor is shown in [ONe83, Corollary 3.42].

A connection on a smooth manifold induces a *covariant derivative along curves*. With this, it is possible to define the *acceleration* of curves. To this end, for an interval I and a smooth manifold M, a *smooth vector field along a curve* $\gamma\colon I \to M$ is a smooth map $X\colon I \to TM$ fulfilling $X(t) \in T_{\gamma(t)}M$. An example is the curve's own velocity field $\dot\gamma\colon I \to TM$. The set of all smooth vector field along γ is denoted by $\mathfrak{X}(\gamma)$. If there is a smooth vector field $\tilde X$ on a neighborhood of the image of γ whose restriction to γ fulfills $\tilde X_{\gamma(t)} = X(t)$ for $X \in \mathfrak{X}(\gamma)$, X is called *extendible* [Lee18, p. 100]. For any connection ∇, and therefore also for the Levi-Civita connection, [Lee18, Theorem 4.24] states that there is a unique *covariant derivative along* γ, denoted by $D_t\colon \mathfrak{X}(\gamma) \to \mathfrak{X}(\gamma)$, which is linear over $\mathbb{R}$, fulfills the product rule $D_t(fX) = \frac{\mathrm{d}f}{\mathrm{d}t}X + fD_tX$ for all $f \in C^\infty(I)$ and $X \in \mathfrak{X}(\gamma)$, and fulfills for any extendible $X \in \mathfrak{X}(\gamma)$ and every extension $\tilde X$

$$D_tX(t) = \nabla_{\dot\gamma(t)}\tilde X,$$

where $\dot\gamma(t) \in T_{\gamma(t)}M$ denotes the tangent vector of γ at $t \in I$. The *covariant acceleration* of γ is now defined as $D_t\dot\gamma$.

Geodesics Connections make it possible to define generalizations of straight lines, so called geodesics. A *geodesic* is therefore defined as a curve $\gamma\colon I \to M$ with vanishing covariant acceleration $D_t\dot{\gamma} = 0$. Any connection defines geodesics, but geodesics with respect to the Levi-Civita connection of a (pseudo-)Riemannian metric are called *(pseudo-)Riemannian geodesics*. One important result of Riemannian geometry is the fact that, on a Riemannian manifold, the Riemannian geodesics are locally shortest connecting curves [Lee18, Theorem 6.15] for the Riemannian distance metric. This makes Riemannian geodesics attractive for optimization problems.

By [ONe83, Proposition 3.24], for any tangent vector $v \in T_pM$ there is unique geodesic $\gamma_v\colon I \to M$ such that $\dot{\gamma}_v(0) = v$, where the domain I around 0 is as large as possible. This motivates the definition of the *exponential map* of a connection (not to be confused with the exponential map of a Lie group), which maps tangent vectors to the associated geodesics. To be precise, let $\mathcal{E}_p \subset T_pM$ denote the set of all tangent vectors v for which the geodesic γ_v is defined at least on the interval $[0, 1]$. Then the exponential map of M at p is defined as the map $\operatorname{Exp}_p\colon \mathcal{E}_p \to M$, $\operatorname{Exp}_p(v) = \gamma_v(1)$ [ONe83, Definition 3.29]. By reparametrization of the geodesic [ONe83, p. 71], it holds that

$$\operatorname{Exp}_p(tv) = \gamma_v(t).$$

As the exponential map defines a curve from a point in a given direction in this way, it can be used in so called *gradient descent methods*, or to be more precise, *line-search schemes* on a Riemannian manifold. The goal is to minimize a function $f \in C^\infty(M)$ by repeatedly stepping in the direction of the negative gradient, since the exact solution $\gamma\colon I \to M$ to the gradient descent problem $\dot{\gamma}(t) = -\operatorname{grad}_f(\gamma(t))$ is typically unknown. For an introduction to such techniques, see for example [AMS08, Chapter 4].

The exponential map allows to map tangent vectors to the manifold. Since the exponential map is a diffeomorphism on a neighborhood around the zero tangent vector in every tangent space [ONe83, Proposition 3.30], it is locally invertible. It therefore allows to map "close-by" points on the manifold to a chosen tangent space. This local inverse is sometimes called *(pseudo-)Riemannian logarithm*, in analogy to the exponential and logarithm on $\mathbb{R}$ or $\mathbb{C}$. In applications, knowledge of the logarithm aids in multiple aspects: As it is a map from the (curved) manifold to a (linear) tangent space, it can be useful in interpolation problems. Furthermore, in the case of Riemannian metrics, it facilitates the computation of the Riemannian distance between two points, as the tangent vector it produces is exactly the initial tangent vector of the geodesic, i.e., shortest curve, between the points. As geodesics are free of acceleration, the length of the initial tangent vector is equal to the Riemannian distance in this case. By mapping from the manifold to the tangent space, the (pseudo-)Riemannian logarithm moreover defines local coordinates, the so called *normal coordinates*.

(Pseudo-)Riemannian Metrics on Homogeneous Spaces The interaction of homogenous spaces and (pseudo-)Riemannian metrics leads to a lot of structure, if the Lie group action and metric are compatible in a certain way [ONe83, Chapter 11]. In practical computations, this structure makes it possible to find explicit expressions for (pseudo-)Riemannian metrics, geodesics, curvature, etc. on quotients of matrix Lie groups.

A *left-invariant metric* on a Lie group G is defined as a pseudo-Riemannian metric h for which the left translation $L_g \colon G \to G$ is an isometry for every $g \in G$, i.e.,

$$h_g(\mathrm{d}(L_g)_{\tilde{g}}x, \mathrm{d}(L_g)_{\tilde{g}}y) = h_{\tilde{g}}(x, y)$$

holds for all $x, y \in T_{\tilde{g}}G$. Right-invariant metrics are defined in an analog fashion. Metrics that are both left- and right-invariant are called *bi-invariant*. One important property of bi-invariant metrics is the fact that in this and only this case, the Lie group exponential map and the geodesic exponential for the Levi-Civita connection coincide [ONe83, Proposition 9.11 (6)], when one identifies the tangent space at the identity with the Lie algebra of the Lie group. On a Lie group G with a bi-invariant metric h and associated scalar product $\langle \cdot, \cdot \rangle$ on the Lie algebra $\mathfrak{g}$, the Riemannian curvature tensor and the sectional curvature can be calculated by straightforward formulas. Let $X, Y, Z \in \mathfrak{g}$ and X, Y linearly independent. Then by [ONe83, Corollary 11.10] it holds that

1. $R(X, Y)Z = \frac{1}{4}[Z, [X, Y]]$ and
2. $K(X, Y) = \frac{1}{4}\frac{\langle [X,Y],[X,Y] \rangle}{\langle X,X \rangle \langle Y,Y \rangle - \langle X,Y \rangle^2}$.

In the study of homogeneous spaces, the concept of (pseudo-)Riemannian submersions is necessary in order to connect metrics on the group with metrics on the homogeneous space. Let M and N be two smooth manifolds and $\pi \colon M \to N$ be a smooth map with surjective differential at every point. Then π is called a *submersion*. For all of the following, we assume all submersions to be surjective themselves, a property that is sometimes even included in the definition of submersions, for example in [ONe83, Definition 1.38]. With respect to a submersion $\pi \colon M \to N$ and a (pseudo-)Riemannian metric h on M, it is possible to split up every tangent space T_pM into a vertical and horizontal part. The vertical space is defined as

$$\mathsf{Ver}_p^{\pi} M = \ker \mathrm{d}\pi_p$$

and is independent of the metric h, while the horizontal space is given by

$$\mathsf{Hor}_p^{\pi,h} M = (\mathsf{Ver}_p^{\pi})^{\perp} = \{u \in T_pM \mid h(u, v) = 0 \text{ for all } v \in T_pM\}.$$

By the non-degeneracy of pseudo-Riemannian metrics, it holds that the tangent space at p is given by the direct sum $T_pM = \mathsf{Ver}_p^{\pi} M \oplus \mathsf{Hor}_p^{\pi,h} M$.

A surjective submersion π between two (pseudo-)Riemannian manifolds M and N is called a *pseudo-Riemannian submersion*, if the fibers $\pi^{-1}(q)$ are pseudo-Riemannian submanifolds of M for all $q \in N$ and the differential $\mathrm{d}\pi_p \colon T_pM \to T_{\pi(p)}N$ restricts to an isometry on the horizontal space $\mathsf{Hor}_p^{\pi,h} M$ for all $p \in M$, where h denotes the pseudo-Riemannian metric of M [ONe83, Definition 7.44]. The first property ensures that h is also non-degenerate when restricted to the fibers. If h is Riemannian, this holds automatically by positive-definiteness. In this case, π is called a *Riemannian submersion*. As $d\pi_p$ is an isometry between $\mathsf{Hor}_p^{\pi,h} M$ and $T_{\pi(p)}N$, tangent vectors of N can be identified with horizontal vectors of M, the so called *horizontal lifts*. This means that for every $v \in T_{\pi(p)}N$, there is a unique $v_p^{\mathsf{hor}} \in \mathsf{Hor}_p^{\pi,h} M$ such that $d\pi_p\left(v_p^{\mathsf{hor}}\right) = v$. The (pseudo-)Riemannian metric $\tilde{h}$ on N can then be calculated via the (pseudo-)Riemannian metric on M, applied to horizontal lifts, i.e.,

$$\tilde{h}_{\pi(p)}(v_1, v_2) = h_p\left((v_1)_p^{\mathsf{hor}}, (v_2)_p^{\mathsf{hor}}\right).$$

More generally, every smooth vector field $X: N \to TN$ on N can be identified with a vector field X^{hor} on M, called the horizontal lift of X, which lies in the horizontal space at every point and is mapped to X via $d\pi$.

The relation of two (pseudo-)Riemannian manifolds (M, h) and $(N, \tilde{h})$ via a (pseudo-)Riemannian submersion $\pi: M \to N$ allows to relate geometric properties such as geodesics and curvature between the two manifolds. By [ONe83, Corollary 7.46], any horizontal geodesic in M, i.e., any geodesic for which the tangent vector is in the horizontal space at every point, is mapped to a geodesic in N. Furthermore, by [ONe83, Theorem 11.47], for $X, Y \in \mathfrak{X}(N)$ linearly independent at every point, the sectional curvature at $\pi(p) \in N$ is given by

$$K_{\pi(p)}\left(X_{\pi(p)}, Y_{\pi(p)}\right) = K_p(X_p^{\mathsf{hor}}, Y_p^{\mathsf{hor}}) + \frac{3}{4}\frac{h([X^{\mathsf{hor}}, Y^{\mathsf{hor}}]_{\mathsf{ver}}, [X^{\mathsf{hor}}, Y^{\mathsf{hor}}]_{\mathsf{ver}})}{(h(X^{\mathsf{hor}}, X^{\mathsf{hor}})h(Y^{\mathsf{hor}}, Y^{\mathsf{hor}}) - h(X^{\mathsf{hor}}, Y^{\mathsf{hor}})^2)}\bigg|_p,$$

where X^{hor} is the horizontal lift of X and $[X^{\mathsf{hor}}, Y^{\mathsf{hor}}]_{\mathsf{ver}}$ denotes the vertical part of the vector field $[X^{\mathsf{hor}}, Y^{\mathsf{hor}}]$.

The left and right multiplication are the natural Lie group actions of a Lie group on itself. When a Lie group G acts on a general smooth manifold M, a (pseudo-)Riemannian metric on M is called G-invariant if the Lie group action is an isometry. For reductive homogeneous spaces, the concept of bi-invariant metrics on Lie groups is generalized through the notion of a *naturally reductive homogeneous* space [ONe83, Definition 11.23]. This is a reductive space G/H with a G-invariant metric h for which the equation

$$\langle [X, Y]_{\mathfrak{m}}, Z\rangle = \langle X, [Y, Z]_{\mathfrak{m}}\rangle$$

holds for all $X, Y, Z \in \mathfrak{m}$, where $[\cdot,\cdot]_{\mathfrak{m}}$ denotes the projection of the lie bracket into $\mathfrak{m}$ and $\langle\cdot,\cdot\rangle$ is the scalar product associated with h on $\mathfrak{m} \cong T_{\pi(e)}G/H$. In [ONe83, Lemma 11.24] it is shown that the projection π from a Lie group G to a naturally reductive space G/H is a pseudo-Riemannian submersion. In case of a Riemannian metric, the submersion is also a Riemannian submersion. It follows that the (pseudo-)Riemannian geodesics on a naturally reductive space are the projections of the (pseudo-)Riemannian geodesics in the Lie group with initial horizontal tangent vector, as can be seen from [ONe83, Proposition 11.24]. The sectional curvature of a naturally reductive homogeneous space is according to [ONe83, Proposition 11.26] given by

$$K(\mathrm{d}\pi X, \mathrm{d}\pi Y) = \frac{\frac{1}{4}\left\langle [X,Y]_{\mathfrak{m}}, [X,Y]_{\mathfrak{m}} \right\rangle + \left\langle [[X,Y]_{\mathfrak{h}}, X], Y \right\rangle}{\langle X,X\rangle \langle Y,Y\rangle - \langle X,Y\rangle^2},$$

where $X, Y \in \mathfrak{m}$ are linearly independent and $[\cdot,\cdot]_{\mathfrak{h}}$ denotes the projection of $[\cdot,\cdot]$ into $\mathfrak{h}$.

A *symmetric space* has even more structure than a naturally reductive space. It is defined as a connected smooth manifold M with (pseudo-)Riemannian metric h, with a unique isometry $\phi_p\colon M \to M$ at every $p \in M$ such that $\mathrm{d}\phi_p = -\,\mathrm{id}$, where id denotes the identity mapping [ONe83, Definition 8.18]. A symmetric space is in particular a naturally reductive space G/H [ONe83, Chapter 11, p. 317], and for $\mathfrak{h}$ and $\mathfrak{m}$ it holds by [ONe83, Lemma 11.30] that

$$[\mathfrak{h},\mathfrak{h}] \subset \mathfrak{h}, \quad [\mathfrak{h},\mathfrak{m}] \subset \mathfrak{m}, \quad [\mathfrak{m},\mathfrak{m}] \subset \mathfrak{h},$$

which is to be understood as the lie bracket of two elements in the respective spaces being an element in the given space. On a symmetric space, the sectional curvature is shown in [ONe83, Proposition 11.31] to be given by

$$K(X,Y) = \frac{\left\langle [[X,Y],X], Y \right\rangle}{\langle X,X\rangle \langle Y,Y\rangle - \langle X,Y\rangle^2},$$

where $X, Y \in \mathfrak{m}$ are linearly independent.

1.2 Subspaces in Parametric Model Order Reduction

This thesis treats classical and symplectic Stiefel and Grassmann manifolds, i.e., manifolds of linear or symplectic bases of subspaces and of the subspaces themselves. One motivation where a geometric treatment of subspaces is important in applications comes from *parametric model order reduction*. An introduction to the case where linear subspaces are used can for example be found in [BGW15].

Model Order Reduction via Proper Orthogonal Decomposition The basic idea is to take a mathematical model, i.e., a parameter dependent ordinary differential equation

$$\dot{y}(t,\rho) = f(t, y(t,\rho)),$$

which may come from the spatial discretization of a partial differential equation. Here $\rho \in D$ denotes the parameter in some parameter space D and $y(\cdot,\rho)\colon \mathbb{R} \to \mathbb{R}^n$ maps to a typically high-dimensional space $\mathbb{R}^n$. The reduction of the model comes by projection into a (much) lower dimensional subspace $\mathcal{U} \subset \mathbb{R}^n$ of dimension $k \ll n$. The optimal subspace for this reduction is defined by the optimization problem

$$\underset{U\in\mathbb{R}^{n\times k},\ U^TU=I_k}{\arg\min} \|S - UU^TS\|_F,$$

where I_k denotes the $k \times k$ identity matrix, $||\cdot||_F$ denotes the Frobenius norm, and $S \in \mathbb{R}^{n\times s}$ denotes a snapshot matrix of solutions $y_1, \ldots, y_s \in \mathbb{R}^n$ of the full order problem for different values of t and ρ. Here, the columns of U span the optimal subspace $\mathcal{U}$. The problem can be solved via *proper orthogonal decomposition*, i.e., by a singular value decomposition (SVD) of S.

Creating snapshot matrices S_ρ for different fixed parameter values ρ leads to different optimal subspaces for different parameter values. Here, differential geometry comes into play: Geometric methods can be used to approximate the optimal subspace for an unsampled parameter value by interpolating between subspaces for sampled parameter values [AF08; DVW10; Zim14; ZD17]. This facilitates the solution of the reduced order model for an unsampled parameter value without having to re-solve the full order model, thereby saving computation cost. Paramount to that approach are efficient algorithms to interpolate subspaces geometrically.

Symplectic Model Order Reduction While parametric model order reduction for general systems is well known, there has been recent interest in structure-preserving model order reduction for *Hamiltonian systems*, which are central in classical mechanics [AG01]. Symplectic model order reduction, studied in [PM16; AH17; Afk+18; BBH19], retains the symplectic structure of Hamiltonian systems and delivers therefore more accurate results than general model order reduction via proper orthogonal decomposition.

Let $\mathcal{V}$ be a vector space and $\omega\colon \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ a *symplectic form*, i.e., ω is skew-symmetric, non-degenerate and bilinear. The tupel $(\mathcal{V}, \omega)$ is called a *symplectic vector space*, and a *Hamiltonian system* is a symplectic vector space together with a smooth function $H\colon \mathcal{V} \to \mathbb{R}$, called *Hamiltonian function*. A *symplectic subspace* of $(\mathcal{V}, \omega)$ is a subspace $\mathcal{U} \subset \mathcal{V}$ such that ω restricts to a symplectic form on $\mathcal{U}$. An example of a symplectic vector space is the *standard symplectic space* $\mathbb{R}^{2n}$ with symplectic form

$$\omega_0(x, y) = x^T J_{2n} y,$$

where

$$J_{2n} = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} \in \mathbb{R}^{2n\times 2n}.$$

The linear Darboux theorem [AG01, p. 5] shows that every real symplectic vector space $(\mathcal{V}, \omega)$ is equivalent to the standard symplectic space $(\mathbb{R}^{2n}, \omega_0)$, in the sense that there is an isomorphism between $\mathcal{V}$ and $\mathbb{R}^{2n}$, preserving the symplectic form. All considerations can therefore be restricted to the standard case.

Let $H\colon \mathbb{R}^{2n} \to \mathbb{R}$ be a Hamiltonian function. Hamilton's equations, which describe the dynamics of the Hamiltonian system, are defined as

$$\dot{x}(t) = J_{2n} \nabla H(x(t)),$$

where $x\colon \mathbb{R} \to \mathbb{R}^{2n}$ and ∇H denotes the Euclidean gradient of H [PM16, Equation (2.2)].

The starting point of symplectic model order reduction is to find an optimal symplectic subspace of dimension $2k \ll 2n$ to approximate the Hamiltonian system. The related optimization problem is of the form

$$\underset{U \in \mathbb{R}^{2n \times 2k},\ U^T J_{2n} U = J_{2k}}{\arg\min} ||S - U J_{2k}^T U^T J_{2n} S||_F, \tag{1.2.1}$$

where $S \in \mathbb{R}^{2n,s}$ is again a snapshot matrix of the full order model [AH17, Equation (42)]. The constraint $U^T J_{2n} U = J_{2k}$ ensures that the columns of U span a symplectic subspace of $(\mathbb{R}^{2n}, \omega_0)$. This problem is called *proper symplectic decomposition*. While there are different approaches to approximate a solution, there is no known direct solution to the proper symplectic decomposition.

As a new approach, the methods introduced in Chapter 4 could lead to a geometric solution of (1.2.1). Furthermore, when the Hamiltonian function H depends on a parameter ρ, an analogue approach to the general subspace interpolation methods could be used in future projects, facilitated by the interpolation methods introduced in Chapter 4.

Part I

The Classical Stiefel and Grassmann Manifolds

Chapter 2

A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects

Thomas Bendokat[a1], Ralf Zimmermann[a], P.-A. Absil[b]

[a] Department of Mathematics and Computer Science, University of Southern Denmark (SDU), Odense, Denmark, {bendokat,zimmermann}@imada.sdu.dk
[b] ICTEAM Institute, UCLouvain, 1348 Louvain-la-Neuve, Belgium

Abstract The Grassmann manifold of linear subspaces is important for the mathematical modelling of a multitude of applications, ranging from problems in machine learning, computer vision and image processing to low-rank matrix optimization problems, dynamic low-rank decompositions and model reduction. With this work, we aim to provide a collection of the essential facts and formulae on the geometry of the Grassmann manifold in a fashion that is fit for tackling the aforementioned problems with matrix-based algorithms. Moreover, we expose the Grassmann geometry both from the approach of representing subspaces with orthogonal projectors and when viewed as a quotient space of the orthogonal group, where subspaces are identified as equivalence classes of (orthogonal) bases. This bridges the associated research tracks and allows for an easy transition between these two approaches.

Original contributions include a modified algorithm for computing the Riemannian logarithm map on the Grassmannian that is advantageous numerically but also allows for a more elementary, yet more complete description of the cut locus and the conjugate points. We also derive a formula for parallel transport along geodesics in the orthogonal projector perspective, formulae for the derivative of the exponential map, as well as a formula for Jacobi fields vanishing at one point.

Keywords Grassmann manifold, Stiefel manifold, orthogonal group, Riemannian exponential, geodesic, Riemannian logarithm, cut locus, conjugate locus, curvature, parallel transport, quotient manifold, horizontal lift, subspace, singular value decomposition

AMS subject classifications 15-02, 15A16, 15A18, 15B10, 22E70, 51F25, 53C80, 53Z99

[1] corresponding author
The arXiv version of this paper can be found at [BZA20]. The paper has been submitted to ACOM.

Acknowledgments This work was initiated when the first author was at UCLouvain for a research visit, hosted by the third author. The third author was supported by the Fonds de la Recherche Scientifique – FNRS and the Fonds Wetenschappelijk Onderzoek – Vlaanderen under EOS Project no 30468160.

Notation

Symbol	Matrix Definition	Name
I_p	$\operatorname{diag}(1,\dots,1) \in \mathbb{R}^{p\times p}$	Identity matrix
$I_{n,p}$	$\begin{pmatrix} I_p \\ 0 \end{pmatrix} \in \mathbb{R}^{n\times p}$	
Sym_n	$\{X \in \mathbb{R}^{n\times n} \mid X = X^T\}$	Space of symmetric matrices
$\mathrm{O}(n)$	$\{Q \in \mathbb{R}^{n\times n} \mid Q^TQ = I_n = QQ^T\}$	Orthogonal group
$T_Q\mathrm{O}(n)$	$\{Q\Omega \in \mathbb{R}^{n\times n} \mid \Omega^T = -\Omega\}$	Tangent space of $\mathrm{O}(n)$ at Q
$\mathrm{St}(n,p)$	$\{U \in \mathbb{R}^{n\times p} \mid U^TU = I_p\}$	Stiefel manifold
$T_U\mathrm{St}(n,p)$	$\{D \in \mathbb{R}^{n\times p} \mid U^TD = -D^TU\}$	Tangent space of $\mathrm{St}(n,p)$ at U
$\mathrm{Gr}(n,p)$	$\{P \in \mathrm{Sym}_n \mid P^2 = P,\ \operatorname{rank}(P) = p\}$	Grassmann manifold
$T_P\mathrm{Gr}(n,p)$	$\{\Delta \in \mathrm{Sym}_n \mid \Delta = P\Delta + \Delta P\}$	Tangent space of $\mathrm{Gr}(n,p)$ at P
$U_\perp$	$(U\ U_\perp) \in \mathrm{O}(n)$	Orthogonal completion of $U \in \mathrm{St}(n,p)$
$g_U^{\mathrm{St}}(D_1,D_2)$	$\operatorname{tr}(D_1^T(I_n - \frac{1}{2}UU^T)D_2)$	(Quotient) metric in $T_U\mathrm{St}(n,p)$
$g_P^{\mathrm{Gr}}(\Delta_1,\Delta_2)$	$\frac{1}{2}\operatorname{tr}(\Delta_1^T\Delta_2)$	Riemannian metric in $T_P\mathrm{Gr}(n,p)$
$\pi^{\mathrm{OS}}(Q)$	$QI_{n,p}$	Projection from $\mathrm{O}(n)$ to $\mathrm{St}(n,p)$
$\pi^{\mathrm{SG}}(U)$	UU^T	Projection from $\mathrm{St}(n,p)$ to $\mathrm{Gr}(n,p)$
$\pi^{\mathrm{OG}}(Q)$	$QI_{n,p}I_{n,p}^TQ^T$	Projection from $\mathrm{O}(n)$ to $\mathrm{Gr}(n,p)$
$\mathsf{Ver}_U\,\mathrm{St}(n,p)$	$\{UA \in \mathbb{R}^{n\times p} \mid A^T = -A \in \mathbb{R}^{p\times p}\}$	Vertical space w.r.t. π^{SG}
$\mathsf{Hor}_U\,\mathrm{St}(n,p)$	$\{U_\perp B \in \mathbb{R}^{n\times p} \mid B \in \mathbb{R}^{(n-p)\times p}\}$	Horizontal space w.r.t. π^{SG}
Δ_U^{hor}	$\Delta U \in \mathbb{R}^{n\times p}$	Horizontal lift of $\Delta \in T_P\mathrm{Gr}(n,p)$ to $\mathsf{Hor}_U\,\mathrm{St}(n,p)$
$[U]$	$\{UR \in \mathrm{St}(n,p) \mid R \in \mathrm{O}(p)\}$	Equivalence class representing a point in $\mathrm{Gr}(n,p)$
$\mathrm{Exp}_P^{\mathrm{Gr}}(t\Delta)$	$\pi^{\mathrm{SG}}(UV\cos(t\Sigma) + \hat{Q}\sin(t\Sigma))$	Riemannian exponential for $\Delta_U^{\mathsf{hor}} \overset{\mathrm{SVD}}{=} \hat{Q}\Sigma V^T \in \mathsf{Hor}_U\,\mathrm{St}(n,p)$
$\mathrm{Log}_P^{\mathrm{Gr}}(F)$	$\Delta \in T_P\mathrm{Gr}(n,p)$ s.t. $\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta) = F$	Riemannian logarithm in $\mathrm{Gr}(n,p)$
$K_P(\Delta_1,\Delta_2)$	$4\frac{\operatorname{tr}(\Delta_1^2\Delta_2^2) - \operatorname{tr}((\Delta_1\Delta_2)^2)}{\operatorname{tr}(\Delta_1^2)\operatorname{tr}(\Delta_2^2) - (\operatorname{tr}(\Delta_1\Delta_2))^2}$	Sectional curvature of $\mathrm{Gr}(n,p)$

2.1 Introduction

The collection of all linear subspaces of fixed dimension p of the Euclidean space $\mathbb{R}^n$ forms the Grassmann manifold $\mathrm{Gr}(n,p)$, also termed the Grassmannian. Subspaces and thus Grassmann manifolds play an important role in a large variety of applications. These include, but are not limited to, data analysis and signal processing [Gal+03; Rah+05; Ren13], subspace estimation and subspace tracking [BA15; BW15; ZB16], structured matrix optimization problems [EAS98; AMS04; AMS08], dynamic low-rank decompositions [HLW06; KL07], projection-based parametric model reduction [AF08; Ngu13; Zim14; NS15; ZPW18] and computer vision [Man12], see also the collections [MM16; ST15]. Moreover, Grassmannians are extensively studied for their purely mathematical aspects [Lei61; Won67; Won68a; Won68b; Sak77; MS85; Koz01] and often serve as illustrating examples in the differential geometry literature [KN96; Hel01].

In this work, we approach the Grassmannian from a matrix-analytic perspective. The focus is on the computational aspects as well as on geometric concepts that directly or indirectly feature in matrix-based algorithmic applications. The most prominent approaches of representing points on Grassmann manifolds with matrices in computational algorithms are

- *the basis perspective*: A subspace $\mathcal{U} \in \mathrm{Gr}(n,p)$ is identified with a (non-unique) matrix $U \in \mathbb{R}^{n\times p}$ whose columns form a basis of $\mathcal{U}$. In this way, a subspace is identified with the equivalence class of all rank-p matrices whose columns span $\mathcal{U}$. For an overview of this approach, see for example the survey [AMS04]. A brief introduction is given in [HM94].

- *the ONB perspective*: In analogy to the basis perspective above, a subspace $\mathcal{U}$ may be identified with the equivalence class of matrices whose columns form an orthonormal basis (ONB) of $\mathcal{U}$. This is often advantageous in numerical computations. This approach is surveyed in [EAS98].

- *the projector perspective*: A subspace $\mathcal{U} \in \mathrm{Gr}(n,p)$ is identified with the (unique) orthogonal projector $P \in \mathbb{R}^{n\times n}$ onto $\mathcal{U}$, which in turn is uniquely represented by $P = UU^T$, with U from the ONB perspective above. For an approach without explicit matrices see [MS85], and for the approach with matrices see for example [HHT07; Bat+15; HM94].

- *the Lie group perspective*: A subspace $\mathcal{U} \in \mathrm{Gr}(n,p)$ is identified with an equivalence class of orthogonal $n\times n$ matrices. This perspective is for example taken in [Gal+03; SL05].

These approaches are closely related and all of them rely on Lie group theory to some extent. Yet, the research literature on the basis/ONB perspective and the projector perspective is rather disjoint. The recent preprint [LLY20] proposes yet another perspective, namely representing p-dimensional subspaces as symmetric orthogonal matrices of trace $2p-n$. This approach corresponds to a scaling and translation of the projector matrices in the vector space of symmetric matrices, hence it yields very similar formulae.

Raison d'être and original contributions We treat the Lie group approach, the ONB perspective and the projector perspective simultaneously. This may serve as a bridge between the corresponding research tracks. Moreover, we collect the essential facts and concepts that feature as generic tools and building blocks in Riemannian computing problems on the Grassmann manifold in terms of matrix formulae, fit for algorithmic calculations. This includes, among others, the Grassmannian's quotient space structure (Subsection 2.2.2), the Riemannian metric (Subsection 2.3.1) and distance, the Riemannian connection (Subsection 2.3.2), the Riemannian exponential (Subsection 2.3.4) and its inverse, the Riemannian logarithm (Subsection 2.5.2), as well as the associated Riemannian normal coordinates (Section 2.6), parallel transport of tangent vectors (Subsection 2.3.6) and the sectional curvature (Subsection 2.4.2). Wherever possible, we provide self-contained and elementary derivations of the sometimes classical results. Here, the term elementary is to be understood as "via tools from linear algebra and matrix analysis" rather than "via tools from abstract differential geometry". Care has been taken that the quantities that are most relevant for algorithmic applications are stated in a form that allows calculations that scale in $\mathcal{O}(np^2)$.

As novel research results, we provide a modified algorithm (Algorithm 2.5.3) for computing the Riemannian logarithm map on the Grassmannian that has favorable numerical features and additionally allows to (non-uniquely) map points from the cut locus of a point to its tangent space. Therefore any set of points on the Grassmannian can be mapped to a single tangent space (Theorem 2.5.4 and Theorem 2.5.5). In particular, we give explicit formulae for the (possibly multiple) shortest curves between any two points on the Grassmannian as well as the corresponding tangent vectors. Furthermore, we present a more elementary, yet more complete description of the conjugate locus of a point on the Grassmannian, which is derived in terms of principal angles between subspaces (Theorem 2.7.2). We also derive a formula for parallel transport along geodesics in the orthogonal projector perspective (Proposition 2.3.5), formulae for the derivative of the exponential map (Subsection 2.3.5), as well as a formula for Jacobi fields vanishing at one point (Proposition 2.7.1).

Organization Section 2.2 introduces the manifold structure of the Grassmann manifold and provides basic formulae for representing Grassmann points and tangent vectors via matrices. Section 2.3 recaps the essential Riemann-geometric aspects of the Grassmann manifold, including the Riemannian exponential, its derivative and parallel transport. In Section 2.4, the Grassmannian's symmetric space structure is established by elementary means and used to explore the sectional curvature and its bounds. In Section 2.5, the (tangent) cut locus is described and a new algorithm is proposed to calculate the pre-image of the exponential map, i.e. the Riemannian logarithm where the pre-image is unique. Section 2.6 addresses normal coordinates and local parameterizations for the Grassmannian. In Section 2.7, questions on Jacobi fields and the conjugate locus of a point are considered. Section 2.8 concludes the paper.

2.2 The Manifold Structure of the Grassmann Manifold

In this section, we recap the definition of the Grassmann manifold and connect results from [EAS98; Bat+15; MS85; HHT07]. Tools from Lie group theory establish the quotient space structure of the Grassmannian, which gives rise to efficient representations. The required Lie group background can be found in the appendix and in [Hal15; Lee12].

The *Grassmann manifold* (also called *Grassmannian*) is defined as the set of all p-dimensional subspaces of the Euclidean space $\mathbb{R}^n$. This set can be identified with the set of orthogonal rank-p projectors,

$$\mathrm{Gr}(n,p) := \left\{ P \in \mathbb{R}^{n\times n} \;\middle|\; P^T = P,\ P^2 = P,\ \operatorname{rank} P = p \right\}, \tag{2.2.1}$$

as is for example done in [HHT07; Bat+15]. Note that a projector P is symmetric as a matrix (namely, $P^T = P$) if and only if it is orthogonal as a projection operation (its range and null space are mutually orthogonal) [Saa92, §3]. The identification in (2.2.1) associates P with the subspace $\mathcal{U} := \operatorname{range}(P)$. Every $P \in \mathrm{Gr}(n,p)$ can in turn be identified with an equivalence class of orthonormal basis matrices spanning the same subspace; an approach that is for example chosen in [EAS98]. These ONB matrices are elements of the so called *Stiefel manifold*

$$\mathrm{St}(n,p) := \left\{ U \in \mathbb{R}^{n\times p} \;\middle|\; U^T U = I_p \right\}.$$

The link between these two sets is via the projection

$$\pi^{\mathrm{SG}} \colon \mathrm{St}(n,p) \to \mathrm{Gr}(n,p),\ U \mapsto UU^T.$$

To obtain a manifold structure on $\mathrm{Gr}(n,p)$ and $\mathrm{St}(n,p)$, we recognize these matrix sets as quotients of the *orthogonal group*

$$\mathrm{O}(n) := \left\{ Q \in \mathbb{R}^{n\times n} \;\middle|\; Q^T Q = I_n = QQ^T \right\}, \tag{2.2.2}$$

which is a compact Lie group. For a brief introduction to Lie groups and their actions, see Appendix 2.A. The link from $\mathrm{O}(n)$ to $\mathrm{St}(n,p)$ and $\mathrm{Gr}(n,p)$ is given by the projections

$$\pi^{\mathrm{OS}} \colon \mathrm{O}(n) \to \mathrm{St}(n,p),\ Q \mapsto Q(:,1:p),$$

where $Q(:,1:p)$ is the matrix formed by the p first columns of Q, and

$$\pi^{\mathrm{OG}} := \pi^{\mathrm{SG}} \circ \pi^{\mathrm{OS}} \colon \mathrm{O}(n) \to \mathrm{Gr}(n,p),\ Q \mapsto Q(:,1:p)Q(:,1:p)^T,$$

respectively. We can consider the following hierarchy of quotient structures:

> Two square orthogonal matrices $Q, \tilde{Q} \in \mathrm{O}(n)$ determine the same rectangular, column-orthonormal matrix $U \in \mathbb{R}^{n\times p}$, if both Q and $\tilde{Q}$ feature U as their first p columns. Two column-orthonormal matrices $U, \tilde{U} \in \mathbb{R}^{n\times p}$ determine the same subspace, if they differ by an orthogonal coordinate change.

This hierarchy is visualized in Figure 2.1. In anticipation of the upcoming discussion, the figure already indicates the lifting of tangent vectors according to the quotient hierarchy.

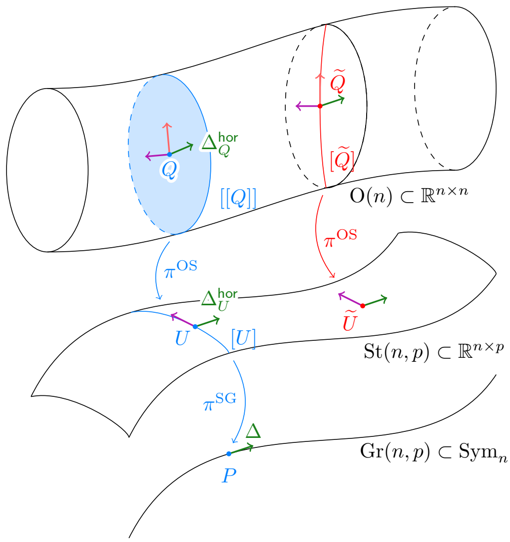

Figure 2.1: Conceptual visualization of the quotient structure of the Grassmann manifold. The double brackets $[[\cdot]]$ denote an equivalence class with respect to $\pi^{\mathrm{OG}} = \pi^{\mathrm{SG}} \circ \pi^{\mathrm{OS}}$, while the single brackets $[\cdot]$ denote an equivalence class with respect π^{OS} or π^{SG}, depending on the element inside the brackets. The tangent vectors along an equivalence class for a projection are vertical with respect to that projection, while the directions orthogonal to the vertical space are horizontal. Correspondingly, the horizontal lift of a tangent vector $\Delta \in T_P\mathrm{Gr}(n,p)$ to $T_U\mathrm{St}(n,p)$ or $T_Q\mathrm{O}(n)$ is orthogonal to all vertical tangent vectors at that point. With respect to the projection π^{SG} from the Stiefel to the Grassmann manifold, the green tangent vector Δ_U^{hor} is horizontal and the magenta tangent vector (along the equivalence class) is vertical. On the other hand, the magenta tangent vectors in $\mathrm{O}(n)$ (pointing to the left) are horizontal with respect to π^{OS} but vertical with respect to π^{OG}.

2.2.1 The Embedded Manifold Structure of the Grassmannian

In order to obtain a smooth manifold structure on the set of orthogonal projectors $\mathrm{Gr}(n,p)$, we can advance as in [HM94, Proposition 2.1.1]. Define an isometric group action of the orthogonal group $\mathrm{O}(n)$ on the symmetric $n \times n$ matrices Sym_n by

$$\Phi\colon \mathrm{O}(n) \times \mathrm{Sym}_n \to \mathrm{Sym}_n,\ (Q,S) \mapsto QSQ^T.$$

Introduce

$$P_0 := \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} \in \mathrm{Gr}(n,p),$$

which is the matrix representation of the canonical projection onto the first p coordinates with respect to the Cartesian standard basis. The set of orthogonal projectors $\mathrm{Gr}(n,p)$

is the orbit $\Phi(\mathrm{O}(n), P_0)$ of the element P_0 under the group action Φ: Any matrix QP_0Q^T obviously satisfies the defining properties of $\mathrm{Gr}(n,p)$ as stated in (2.2.1). Conversely, if $P \in \mathrm{Gr}(n,p)$, then P is real, symmetric and positive semidefinite with p eigenvalues equal to one and $n-p$ eigenvalues equal to zero. Hence, the eigenvalue decomposition (EVD) $P = Q\Lambda Q^T = QP_0Q^T$ establishes P as a point in the orbit of P_0. In other words, we have confirmed that

$$\pi^{\mathrm{OG}} = \Phi|_{O(n)\times\{P_0\}} \colon \mathrm{O}(n) \to \mathrm{Gr}(n,p),\ Q \mapsto QP_0Q^T, \tag{2.2.3}$$

maps into $\mathrm{Gr}(n,p)$ and is surjective. Since $\mathrm{O}(n)$ is compact, the first part of Proposition 2.A.2 in the appendix shows that $\mathrm{Gr}(n,p) = \Phi(\mathrm{O}(n), P_0)$ is an embedded submanifold of Sym_n.

2.2.2 The Quotient Structure of the Grassmannian

To formally introduce the quotient structure of the Grassmannian, we make use of the second part of Proposition 2.A.2. The objects of interest are the orthogonal group $\mathrm{O}(n)$, which is the domain of π^{OS} and π^{OG}, and the Cartesian product $\mathrm{O}(p)\times\mathrm{O}(n-p)$, which can be identified with a subgroup of $\mathrm{O}(n)$.

The stabilizer of Φ at P_0, i.e., of π^{OG}, is given by $H = \{\left(\begin{smallmatrix} R_1 & 0 \\ 0 & R_2 \end{smallmatrix}\right) \in \mathrm{O}(n) \mid R_1 \in \mathrm{O}(p),\ R_2 \in \mathrm{O}(n-p)\} \cong \mathrm{O}(p)\times\mathrm{O}(n-p)$. This is readily seen by noticing that $Q \in \mathrm{O}(n)$ fulfills $\pi^{\mathrm{OG}}(Q) = QP_0Q^T = P_0$ if and only if $Q = \left(\begin{smallmatrix} R_1 & 0 \\ 0 & R_2 \end{smallmatrix}\right)$. An equivalence relation on $\mathrm{O}(n)$ is defined by $\widetilde{Q} \sim Q$ if and only if $\pi^{\mathrm{OG}}(\widetilde{Q}) = \pi^{\mathrm{OG}}(Q)$. This equivalence relation collects all orthogonal matrices whose first p columns span the same subspace into an equivalence class. In other words, the equivalence classes of $\mathrm{O}(n)/H$ are

$$\begin{aligned} [[Q]] &= (\pi^{\mathrm{OG}})^{-1}(\pi^{\mathrm{OG}}(Q)) \\ &= \left\{ \widetilde{Q} \in \mathrm{O}(n) \;\middle|\; \widetilde{Q} = Q \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix},\ \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} \in H \right\}, \end{aligned} \tag{2.2.4}$$

which corresponds to [EAS98, Eq. (2.28)]. The manifold structure on $\mathrm{O}(n)/H \cong \mathrm{O}(n)/(\mathrm{O}(p)\times\mathrm{O}(n-p))$ is by definition the unique one that makes the quotient map

$$\mathrm{O}(n) \to \mathrm{O}(n)/(\mathrm{O}(p)\times\mathrm{O}(n-p)) \colon Q \mapsto [[Q]]$$

a smooth submersion, i.e., a smooth map with surjective differential at every point. The second part of Proposition 2.A.2 shows that, as $\mathrm{Gr}(n,p)$ is the orbit of P_0 under Φ, it holds that

$$\mathrm{Gr}(n,p) \cong \mathrm{O}(n)/(\mathrm{O}(p)\times\mathrm{O}(n-p)).$$

Therefore π^{OG} is also a smooth submersion. Furthermore, we have the well known result

$$\dim(\mathrm{Gr}(n,p)) = \dim(\mathrm{O}(n)) - \dim(\mathrm{O}(p)\times\mathrm{O}(n-p)) = (n-p)p.$$

2.2.3 The Tangent Spaces of the Grassmannian

The quotient structure of the Grassmannian allows to split every tangent space of $\mathrm{O}(n)$ into a vertical and (after choosing a Riemannian metric) horizontal part, and to identify every tangent space of $\mathrm{Gr}(n,p)$ with such a horizontal space as in [EAS98].

As the Lie algebra of $\mathrm{O}(n)$ is the set of skew-symmetric matrices

$$\mathfrak{so}(n) := T_I \mathrm{O}(n) = \left\{ \Omega \in \mathbb{R}^{n\times n} \;\middle|\; \Omega^T = -\Omega \right\},$$

the tangent space at an arbitrary $Q \in \mathrm{O}(n)$ is given by the left translates

$$T_Q \mathrm{O}(n) = \{ Q\Omega \mid \Omega \in \mathfrak{so}(n) \}.$$

Restricting the Euclidean matrix space metric $\langle A, B\rangle_0 = \mathrm{tr}(A^T B)$ to the tangent spaces turns the manifold $\mathrm{O}(n)$ into a Riemannian manifold. We include a factor of $\frac{1}{2}$ to obtain Riemannian metrics on the Stiefel and Grassmann manifold, in Subsections 2.2.4 and 2.3.1, respectively, that comply with common conventions. This yields the Riemannian metric (termed here *metric* for short) $g^{\mathrm{O}}_Q \colon T_Q\mathrm{O}(n) \times T_Q\mathrm{O}(n) \to \mathbb{R}$,

$$g^{\mathrm{O}}_Q(Q\Omega, Q\widetilde{\Omega}) := \left\langle Q\Omega, Q\widetilde{\Omega} \right\rangle_Q := \frac{1}{2}\,\mathrm{tr}\left((Q\Omega)^T Q\widetilde{\Omega} \right) = \frac{1}{2}\,\mathrm{tr}\left(\Omega^T \widetilde{\Omega} \right).$$

The differential of the projection π^{OG} at $Q \in \mathrm{O}(n)$ is a linear map $\mathrm{d}\pi^{\mathrm{OG}}_Q \colon T_Q\mathrm{O}(n) \to T_{\pi^{\mathrm{OG}}(Q)}\mathrm{Gr}(n,p)$, where $T_Q\mathrm{O}(n)$ and $T_{\pi^{\mathrm{OG}}(Q)}\mathrm{Gr}(n,p)$ are the tangent spaces of $\mathrm{O}(n)$ and $\mathrm{Gr}(n,p)$ at Q and $\pi^{\mathrm{OG}}(Q)$, respectively. The directional derivative of π^{OG} at $Q \in \mathrm{O}(n)$ in the tangent direction $Q\Omega = Q\left(\begin{smallmatrix} A & -B^T \\ B & C \end{smallmatrix}\right) \in T_Q\mathrm{O}(n)$ is given by

$$\mathrm{d}\pi^{\mathrm{OG}}_Q(Q\Omega) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\pi^{\mathrm{OG}}(\gamma(t))) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\gamma(t) P_0 \gamma(t)^T) = Q \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix} Q^T, \tag{2.2.5}$$

where $\gamma \colon t \mapsto \gamma(t) \in \mathrm{O}(n)$ is an arbitrary differentiable curve with $\gamma(0) = Q$, $\dot\gamma(0) = Q\Omega$. Since π^{OG} is a submersion, this spans the entire tangent space, i.e.,

$$T_{\pi^{\mathrm{OG}}(Q)}\mathrm{Gr}(n,p) = \left\{ Q \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix} Q^T \;\middle|\; B \in \mathbb{R}^{(n-p)\times p} \right\}.$$

In combination with the metric g^{O}_Q, the smooth submersion π^{OG} allows to decompose every tangent space $T_Q\mathrm{O}(n)$ into a vertical and horizontal part, c.f. [Lee18, Chapter 2]. The vertical part is the kernel of the differential $\mathrm{d}\pi^{\mathrm{OG}}_Q$, and the horizontal part is the orthogonal complement with respect to the metric g^{O}_Q. We therefore have

$$T_Q\mathrm{O}(n) = \mathsf{Ver}^{\pi^{\mathrm{OG}}}_Q \mathrm{O}(n) \oplus \mathsf{Hor}^{\pi^{\mathrm{OG}}}_Q \mathrm{O}(n),$$

where

$$\mathsf{Ver}^{\pi^{\mathrm{OG}}}_Q \mathrm{O}(n) = \left\{ Q \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} \;\middle|\; A \in \mathfrak{so}(p),\ C \in \mathfrak{so}(n-p) \right\}$$

and

$$\mathsf{Hor}_Q^{\pi^{\mathrm{OG}}} \mathrm{O}(n) = \left\{ Q \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix} \;\middle|\; B \in \mathbb{R}^{(n-p)\times p} \right\}, \tag{2.2.6}$$

c.f. [EAS98, Eq. (2.29) and (2.30)]. The tangent space of the Grassmann manifold at $P = \pi^{\mathrm{OG}}(Q)$ can be identified with the horizontal space at any representative $Q \in (\pi^{\mathrm{OG}})^{-1}(P) \subset \mathrm{O}(n)$,

$$T_P \mathrm{Gr}(n,p) \cong \mathsf{Hor}_Q^{\pi^{\mathrm{OG}}} \mathrm{O}(n).$$

In [Bat+15], the tangent space $T_P\mathrm{Gr}(n,p)$ is given by matrices of the form $[\Omega, P]$, where $[\cdot,\cdot]$ denotes the matrix commutator, and $\Omega \in \mathfrak{so}_P(n)$ fulfilling

$$\mathfrak{so}_P(n) := \{\Omega \in \mathfrak{so}(n) \mid \Omega = \Omega P + P\Omega\}. \tag{2.2.7}$$

Writing $P = QP_0Q^T$ and making use of (2.2.5) shows that every $\Delta \in T_P\mathrm{Gr}(n,p)$ is of the form

$$\Delta = Q \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix} Q^T = [Q \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix} Q^T, P]. \tag{2.2.8}$$

Since $\Omega \in \mathfrak{so}_P(n)$ is equivalent to $Q^T\Omega Q \in \mathfrak{so}_{Q^TPQ}(n)$ and $Q^TPQ = P_0$ it follows that every $\Omega \in \mathfrak{so}_P(n)$ is of the form

$$\Omega = Q \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix} Q^T.$$

Note that for $\Delta \in T_P\mathrm{Gr}(n,p)$, there is $\Omega \in \mathfrak{so}_P(n)$ such that $\Delta = [\Omega, P]$. This Ω can be calculated via $\Omega = [\Delta, P] \in \mathfrak{so}_P(n)$.

Proposition 2.2.1 (Tangent vector characterization)**.** Let $P \in \mathrm{Gr}(n,p)$ be the orthogonal projector onto the subspace $\mathcal{U}$. For every symmetric $\Delta = \Delta^T \in \mathbb{R}^{n\times n}$, the following conditions are equivalent:

a) $\Delta \in T_P\mathrm{Gr}(n,p)$,

b) $\Delta(\mathcal{U}) \subset \mathcal{U}^\perp$ and $\Delta(\mathcal{U}^\perp) \subset \mathcal{U}$,

c) $\Delta P + P\Delta = \Delta$,

d) $\Delta = [\Omega, P]$, where $\Omega := [\Delta, P] \in \mathfrak{so}_P(n)$.

Here, $\Delta(\mathcal{U}) := \{\Delta x \in \mathbb{R}^n \mid x \in \mathcal{U}\}$ and the orthogonal complement $\mathcal{U}^\perp$ is taken with respect to the Euclidean metric in $\mathbb{R}^n$.

Proof. The equivalence of a), b) and c) is from [MS85, Result 3.7]. To show c) implies d), note that $\Delta P + P\Delta = \Delta$ implies $P\Delta P = 0$ and therefore $[[\Delta, P], P] = \Delta P + P\Delta - 2P\Delta P = \Delta$. On the other hand, if d) holds then $\Delta = \Delta P + P\Delta - 2P\Delta P$, which also implies $P\Delta P = 0$ by multiplication with P from one side. Inserting $P\Delta P = 0$ into the equation shows that c) holds. The statement that $\Omega \in \mathfrak{so}_P(n)$ is automatically true. □

2.2.4 Horizontal Lift to the Stiefel Manifold

The elements of $\mathrm{Gr}(n,p)$ are $n \times n$ matrices (see the bottom level of Figure 2.1). The map π^{OG} makes it possible to (non uniquely) represent elements of $\mathrm{Gr}(n,p)$ as elements of $\mathrm{O}(n)$—the top level of Figure 2.1—which are also $n \times n$ matrices. In practical computations, however, it is often not feasible to work with $n \times n$ matrices, especially if n is large when compared to the subspace dimension p. A remedy is to resort to the middle level of Figure 2.1, namely the Stiefel manifold $\mathrm{St}(n,p)$ [EAS98]. By making use of the map π^{SG}, elements of $\mathrm{Gr}(n,p)$ can be (non uniquely) represented as elements of $\mathrm{St}(n,p)$, which are $n \times p$ matrices.

The Stiefel manifold can be obtained analogously to the Grassmann manifold by means of a group action of $\mathrm{O}(n)$ on $\mathbb{R}^{n\times p}$, defined by left multiplication. It is the orbit of

$$I_{n,p} := \begin{pmatrix} I_p \\ 0 \end{pmatrix} \in \mathbb{R}^{n\times p}$$

under this group action with stabilizer $\mathrm{O}(n-p) \cong \left\{ \left(\begin{smallmatrix} I_p & 0 \\ 0 & R \end{smallmatrix}\right) \;\middle|\; R \in \mathrm{O}(n-p) \right\}$. By Proposition 2.A.2, $\mathrm{St}(n,p) \cong \mathrm{O}(n)/\mathrm{O}(n-p)$ is an embedded submanifold of $\mathbb{R}^{n\times p}$ and the projection from the orthogonal group onto the Stiefel manifold is given by

$$\pi^{\mathrm{OS}} \colon \mathrm{O}(n) \to \mathrm{St}(n,p),\ Q \mapsto QI_{n,p},$$

the projection onto the first p columns. It defines an equivalence relation on $\mathrm{O}(n)$ by collecting all orthogonal matrices that share the same first p column vectors into an equivalence class. As above,

$$\dim(\mathrm{St}(n,p)) = \dim(\mathrm{O}(n)) - \dim(\mathrm{O}(n-p)) = np - \frac{1}{2}p(p+1),$$

and π^{OS} is a smooth submersion, which allows to decompose every tangent space $T_Q\mathrm{O}(n)$ into a vertical and horizontal part with respect to the metric g^{O}_Q. We therefore have

$$T_Q\mathrm{O}(n) = \mathsf{Ver}^{\pi^{\mathrm{OS}}}_Q \mathrm{O}(n) \oplus \mathsf{Hor}^{\pi^{\mathrm{OS}}}_Q \mathrm{O}(n),$$

where

$$\mathsf{Ver}^{\pi^{\mathrm{OS}}}_Q \mathrm{O}(n) = \left\{ Q \begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix} \;\middle|\; C \in \mathfrak{so}(n-p) \right\}$$

and

$$\mathsf{Hor}^{\pi^{\mathrm{OS}}}_Q \mathrm{O}(n) = \left\{ Q \begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix} \;\middle|\; A \in \mathfrak{so}(p),\ B \in \mathbb{R}^{(n-p)\times p} \right\}.$$

By the identification

$$T_U\mathrm{St}(n,p) \cong \mathsf{Hor}^{\pi^{\mathrm{OS}}}_Q \mathrm{O}(n),$$

see [EAS98], and orthogonal completion $U_\perp \in \mathbb{R}^{n\times(n-p)}$ of U, i.e., such that $\begin{pmatrix} U & U_\perp \end{pmatrix} \in \mathrm{O}(n)$, the tangent spaces of the Stiefel manifold are explicitly given by either of the following expressions

$$\begin{aligned} T_U\mathrm{St}(n,p) &= \left\{ UA + U_\perp B \in \mathbb{R}^{n\times p} \;\middle|\; A \in \mathfrak{so}(p),\ B \in \mathbb{R}^{(n-p)\times p} \right\} \\ &= \left\{ UA + (I_n - UU^T)T \mid A \in \mathfrak{so}(p),\ T \in \mathbb{R}^{n\times p} \right\} \\ &= \left\{ \Omega U \mid \Omega \in \mathfrak{so}(n) \right\} \\ &= \left\{ D \in \mathbb{R}^{n\times p} \mid U^T D = -D^T U \right\}. \end{aligned} \tag{2.2.9}$$

Note that $U^T U_\perp = 0$ and $U_\perp^T U_\perp = I_{n-p}$, as well as $I_n = UU^T + U_\perp U_\perp^T$.

The *canonical metric* $g_U^{\mathrm{St}}(\cdot,\cdot)$ on the Stiefel manifold is given via the horizontal lift. That means that for any two tangent vectors in $D_1 = UA_1 + U_\perp B_1$, $D_2 = UA_2 + U_\perp B_2 \in T_U\mathrm{St}(n,p)$, we take a total space representative $Q \in \mathrm{O}(n)$ of $U \in \mathrm{St}(n,p)$ and 'lift' the tangent vectors $D_1, D_2 \in T_U\mathrm{St}(n,p)$ to tangent vectors $D_{1,Q}^{\mathsf{hor}}, D_{2,Q}^{\mathsf{hor}} \in \mathsf{Hor}_Q^{\pi^{\mathrm{OS}}}\mathrm{O}(n) \subset T_Q\mathrm{O}(n)$, defined by $\mathrm{d}(\pi^{\mathrm{OS}})_Q(D_{i,Q}^{\mathsf{hor}}) = D_i$, $i = 1,2$. The inner product between $D_{1,Q}^{\mathsf{hor}}, D_{2,Q}^{\mathsf{hor}}$ is now computed according to the metric of $\mathrm{O}(n)$. In practice, this leads to

$$\begin{aligned} g_U^{\mathrm{St}}(D_1, D_2) := g_Q^{\mathrm{O}}(D_{1,Q}^{\mathsf{hor}}, D_{2,Q}^{\mathsf{hor}}) &= \frac{1}{2}\operatorname{tr}(A_1^T A_2) + \operatorname{tr}(B_1^T B_2) \\ &= \operatorname{tr}\left(D_1^T (I_n - \frac{1}{2} UU^T) D_2 \right), \end{aligned}$$

c.f. [EAS98]. The last equality shows that it does not matter which base point $Q \in (\pi^{\mathrm{OS}})^{-1}(U)$ is chosen for the lift.

In order to make the transition from column-orthogonal matrices U to the associated subspaces $\mathcal{U} = \operatorname{span}(U)$, another equivalence relation, this time on the Stiefel manifold, is required: Identify any matrices $U \in \mathrm{St}(n,p)$, whose column vectors span the same subspace $\mathcal{U}$. For any two Stiefel matrices $U, \tilde{U}$ that span the same subspace, it holds that $\tilde{U} = UU^T\tilde{U}$. As a consequence, $I_p = (\tilde{U}^T U)(U^T\tilde{U})$, so that $R = (U^T\tilde{U}) \in \mathrm{O}(p)$. Hence, any two such Stiefel matrices differ by a rotation/reflection $R \in \mathrm{O}(p)$. Define a smooth right action of $\mathrm{O}(p)$ on $\mathrm{St}(n,p)$ by multiplication from the right. Every equivalence class

$$\mathcal{U} \cong [U] := \left\{ \widetilde{U} \in \mathrm{St}(n,p) \;\middle|\; \widetilde{U} = UR,\ R \in \mathrm{O}(p) \right\} \tag{2.2.10}$$

under this group action can be identified with a projector UU^T and vice versa. Therefore, according to [Lee12, Thm 21.10, p. 544], the set of equivalence classes $[U]$, denoted by $\mathrm{St}(n,p)/\mathrm{O}(p)$, is a smooth manifold with a manifold structure for which the quotient map is a smooth submersion. To show that the manifold structure is indeed the same as the one on $\mathrm{Gr}(n,p)$ (which we can identify as a set with $\mathrm{St}(n,p)/\mathrm{O}(p)$), we show directly that the projection from $\mathrm{St}(n,p)$ to $\mathrm{Gr}(n,p)$,

$$\pi^{\mathrm{SG}} \colon \mathrm{St}(n,p) \to \mathrm{Gr}(n,p),\ U \mapsto UU^T,$$

is a smooth submersion. Indeed, the derivative $\mathrm{d}(\pi^{\mathrm{SG}})_U(D) = DU^T + UD^T$ is surjective, since every tangent vector $\Delta \in T_{\pi^{\mathrm{OG}}(Q)}\mathrm{Gr}(n,p)$ can be written as

$$\Delta = U_\perp B U^T + U B^T U_\perp^T, \tag{2.2.11}$$

by making use of (2.2.8). This shows surjectivity, since for every $\Delta \in T_{\pi^{\mathrm{OG}}(Q)}\mathrm{Gr}(n,p)$ we can choose $U_\perp B \in T_U\mathrm{St}(n,p)$, such that $\mathrm{d}(\pi^{\mathrm{SG}})_U(U_\perp B) = \Delta$.

Again, we split every tangent space $T_U\mathrm{St}(n,p)$ with respect to the projection π^{SG} and the metric $g_U^{\mathrm{St}}(\cdot,\cdot)$ on the Stiefel manifold. Defining the kernel of $\mathrm{d}(\pi^{\mathrm{SG}})_U$ as the vertical space and its orthogonal complement (with respect to the metric g_U^{St}) as the horizontal space leads to the direct sum decomposition

$$T_U\mathrm{St}(n,p) = \mathsf{Ver}_U\,\mathrm{St}(n,p) \oplus \mathsf{Hor}_U\,\mathrm{St}(n,p),$$

where

$$\mathsf{Ver}_U\,\mathrm{St}(n,p) = \{UA \mid A \in \mathfrak{so}(p)\}$$

and

$$\begin{aligned}\mathsf{Hor}_U\,\mathrm{St}(n,p) &= \left\{U_\perp B \;\middle|\; B \in \mathbb{R}^{(n-p)\times p}\right\} = \left\{(I_n - UU^T)T \mid T \in \mathbb{R}^{n\times p}\right\} \\ &= \left\{D \in \mathbb{R}^{n\times p} \;\middle|\; U^T D = 0\right\}.\end{aligned} \tag{2.2.12}$$

Since π^{SG} is the only projection that we use on the Stiefel manifold, the dependence of the splitting on the projection is omitted in the notation.

The tangent space $T_P\mathrm{Gr}(n,p)$ of the Grassmannian can be identified with the horizontal space $\mathsf{Hor}_U\,\mathrm{St}(n,p)$. Therefore, for every tangent vector $\Delta \in T_P\mathrm{Gr}(n,p)$, there is a unique $\Delta_U^{\mathsf{hor}} \in \mathsf{Hor}_U\,\mathrm{St}(n,p)$, called the *horizontal lift of* Δ *to* U. By (2.2.12), there are matrices $T \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{(n-p)\times p}$ such that

$$\Delta_U^{\mathsf{hor}} = U_\perp B = (I_n - UU^T)T \in \mathsf{Hor}_U\,\mathrm{St}(n,p).$$

Note that Δ_U^{hor} depends only on the chosen representative U of P, while B depends on the chosen orthogonal completion $U_\perp$ as well.

Multiplication of (2.2.11) from the right with U shows that the horizontal lift of $\Delta \in T_P\mathrm{Gr}(n,p)$ to $U \in \mathrm{St}(n,p)$ can be calculated by

$$\Delta_U^{\mathsf{hor}} = \Delta U. \tag{2.2.13}$$

Therefore, the horizontal lifts of Δ to two different representatives U and UR are connected by

$$\Delta_{UR}^{\mathsf{hor}} = \Delta_U^{\mathsf{hor}} R, \tag{2.2.14}$$

which relates to [AMS08, Prop. 3.6.1]. The lift of $\Delta \in T_P\mathrm{Gr}(n,p)$ to $Q = \begin{pmatrix} U & U_\perp \end{pmatrix} \in \mathrm{O}(n)$ can also be calculated explicitly. By (2.2.5), (2.2.6) and (2.2.8), it is given by

$$\Delta_Q^{\mathsf{hor}} = [\Delta, P]Q = Q\begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix} \in \mathsf{Hor}_Q\,\mathrm{O}(n).$$

In conclusion, the Grassmann manifold is placed at the end of the following quotient space hierarchy with equivalence classes $[\cdot]$ from (2.2.10) and $[[\cdot]]$ from (2.2.4):

$$\begin{aligned} \mathrm{Gr}(n,p) \cong & \quad \mathrm{St}(n,p)/\mathrm{O}(p) && = \{[U] \mid U \in \mathrm{St}(n,p)\} \\ \cong & \quad \mathrm{O}(n)/(\mathrm{O}(p) \times \mathrm{O}(n-p)) && = \{[[Q]] \mid Q \in \mathrm{O}(n)\}. \end{aligned}$$

Remark. It should be noted that there is yet another way of viewing the Grassmann manifold as a quotient. Instead of taking equivalence classes in $\mathrm{O}(n)$, one can take the quotient of the noncompact Stiefel manifold by the general linear group $\mathrm{GL}(p)$. This introduces a factor of the form $(Y^TY)^{-1}$ into many formulae, where $Y \in \mathbb{R}^{n\times p}$ is a rank p matrix with (not necessarily orthogonal) column vectors spanning the desired subspace. For this approach see for example [AMS04].

2.3 Riemannian Structure

In this section, we study the basic Riemannian structure of the Grassmannian. We introduce the canonical metric coming from the quotient structure—which coincides with the Euclidean metric—and the Riemannian connection. The Riemannian exponential mapping for geodesics is derived in the formulation as projectors as well as with Stiefel representatives. Lastly, we study the concept of parallel transport on the Grassmannian. Many of those results have been studied before for the projector or the ONB perspective. For the metric and the exponential see for example [EAS98; AMS04] (Stiefel perspective) and [Bat+15] (projector perspective). For the horizontal lift of the Riemannian connection see [AMS04]. A formula for parallel transport in the ONB perspective was given in [EAS98]. Here we combine the approaches and provide some modifications and additions. We derive formulae for all mentioned concepts in both perspectives and also study the derivative of the exponential mapping.

2.3.1 Riemannian Metric

The Riemannian metric on the Grassmann manifold that is induced by the quotient structure coincides with (one half times) the Euclidean metric. To see this, let $\Delta_1, \Delta_2 \in T_P\mathrm{Gr}(n,p)$ be two tangent vectors at $P \in \mathrm{Gr}(n,p)$ and let $Q = \begin{pmatrix} U & U_\perp \end{pmatrix} \in \mathrm{O}(n)$ such that $\pi^{\mathrm{OG}}(Q) = P$. The metric on the Grassmann manifold is then inherited from the metric on $\mathrm{O}(n)$ applied to the horizontal lifts, i.e.

$$g_P^{\mathrm{Gr}}(\Delta_1, \Delta_2) := g_Q^{\mathrm{O}}(\Delta_{1,Q}^{\mathsf{hor}}, \Delta_{2,Q}^{\mathsf{hor}}). \tag{2.3.1}$$

Let $\Delta_i = [\Omega_i, P]$, where $\Omega_i \in \mathfrak{so}_P(n)$, as well as $\Delta_{i,Q}^{\mathsf{hor}} = Q\left(\begin{smallmatrix} 0 & -B_i^T \\ B_i & 0 \end{smallmatrix}\right)$ and $\Delta_{i,U}^{\mathsf{hor}} = U_\perp B_i$. We immediately see that

$$\begin{aligned} g_P^{\mathrm{Gr}}(\Delta_1, \Delta_2) &= \frac{1}{2}\operatorname{tr}\left((\Delta_{1,Q}^{\mathsf{hor}})^T \Delta_{2,Q}^{\mathsf{hor}}\right) = \operatorname{tr}\left({\Delta_{1,U}^{\mathsf{hor}}}^T \Delta_{2,U}^{\mathsf{hor}}\right) = \operatorname{tr}(U^T\Delta_1\Delta_2 U) \\ &= \operatorname{tr}(B_1^T B_2) = \frac{1}{2}\operatorname{tr}(\Delta_1\Delta_2) = \frac{1}{2}\operatorname{tr}(\Omega_1^T\Omega_2). \end{aligned} \tag{2.3.2}$$

The last equality can be seen by noticing $[\Omega_i, P] = (I_n - 2P)\Omega_i$ for $\Omega_i \in \mathfrak{so}_P(n)$ and $(I_n - 2P)^2 = I_n$. Although the formulae in (2.3.2) all look similar, notice that $\Delta_i, \Omega_i, \Delta^{\mathsf{hor}}_{i,Q} \in \mathbb{R}^{n\times n}$, but $\Delta^{\mathsf{hor}}_{i,U} \in \mathbb{R}^{n\times p}$ and $B_i \in \mathbb{R}^{(n-p)\times p}$.

The metric does not depend on the point to which we lift: Lifting to a different $UR \in \mathrm{St}(n,p)$ results in a postmultiplication of $\Delta^{\mathsf{hor}}_{i,U}$ with R according to (2.2.14). By the invariance properties of the trace, this does not change the metric. An analogous argument holds for the lift to $\mathrm{O}(n)$.

With the Riemannian metric we can define the induced norm of a tangent vector $\Delta \in T_P\mathrm{Gr}(n,p)$ by

$$\|\Delta\| := \sqrt{g^{\mathrm{Gr}}_P(\Delta,\Delta)} = \frac{1}{\sqrt{2}}\sqrt{\mathrm{tr}(\Delta^2)}.$$

2.3.2 Riemannian Connection

The disjoint collection of all tangent spaces of a manifold M is called the *tangent bundle* $TM = \dot\cup_{p\in M} T_pM$. A *smooth vector field* on M is a smooth map X from M to TM that maps a point $p \in M$ to a tangent vector $X(p) \in T_pM$. The set of all smooth vector fields on M is denoted by $\mathfrak{X}(M)$. Plugging smooth vector fields $Y, Z \in \mathfrak{X}(M)$ into the metric of a Riemannian manifold gives a smooth function $g(Y,Z)\colon M \to \mathbb{R}$. It is not possible to calculate the differential of a vector field in the classical sense, since every tangent space is a separate vector space and the addition of $X(p) \in T_pM$ and $X(q) \in T_qM$ is not defined for $p \neq q$. To this end, the abstract machinery of differential geometry provides special tools called *connections*. A connection acts as the derivative of a vector field in the direction of another vector field. On a Riemannian manifold (M,g), the *Riemannian* or *Levi-Civita connection* is the unique connection $\nabla\colon \mathfrak{X}(M)\times\mathfrak{X}(M) \to \mathfrak{X}(M) : (X,Y) \mapsto \nabla_X Y$ that is

- *compatible with the metric*: for all vector fields $X, Y, Z \in \mathfrak{X}(M)$, we have the product rule
$$\nabla_X g(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$
- *torsion free*: for all $X, Y \in \mathfrak{X}(M)$, $\nabla_X Y - \nabla_Y X = [X,Y]$, where $[X,Y] = X(Y) - Y(X)$ denotes the Lie bracket of two vector fields.

The Riemannian connection can be explicitly calculated in the case of embedded submanifolds: It is the projection of the Levi-Civita connection of the ambient manifold onto the tangent space of the embedded submanifold. For details see for example [Lee18].

The Euclidean space $\mathbb{R}^{n\times p}$ is a vector space, which implies that every tangent space of $\mathbb{R}^{n\times p}$ can be identified with $\mathbb{R}^{n\times p}$ itself. Therefore, the Riemannian connection of the Euclidean space $\mathbb{R}^{n\times p}$ with the Euclidean metric (2.B.1) is the usual directional derivative: Let $F\colon \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$ and $X, Y \in \mathbb{R}^{n\times p}$. The directional derivative of F at X in direction Y is then

$$\mathrm{d}F_X(Y) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} F(X+tY).$$

The same holds for the space of symmetric matrices Sym_n. When considered as the set of orthogonal projectors, the Grassmann manifold $\mathrm{Gr}(n,p)$ is an embedded submanifold of Sym_n. In this case, the projection onto the tangent space is

$$\Pi_{T_P\mathrm{Gr}} \colon \mathrm{Sym}_n \to T_P\mathrm{Gr}(n,p), \quad S \mapsto (I_n - P)SP + PS(I_n - P), \tag{2.3.3}$$

see also [MS85]. In order to restrict calculations to $n \times p$ matrices, we can lift to the Stiefel manifold and use the projection onto the horizontal space, which is

$$\Pi_{\mathsf{Hor}_U \mathrm{St}} \colon \mathbb{R}^{n\times p} \to \mathsf{Hor}_U \mathrm{St}(n,p), \quad Z \mapsto (I_n - UU^T)Z, \tag{2.3.4}$$

see also [EAS98; AMS04]. Note that $\mathsf{Hor}_U \mathrm{St}(n,p) \cong T_{\pi^{\mathrm{SG}}(U)}\mathrm{Gr}(n,p)$ as described in Subsection 2.2.4. The Riemannian connection on $\mathrm{Gr}(n,p)$ is now obtained via the following proposition.

Proposition 2.3.1 (Riemannian Connection)**.** Let $X \in \mathfrak{X}(\mathrm{Gr}(n,p))$ be a smooth vector field on $\mathrm{Gr}(n,p)$, i.e., $X(P) \in T_P\mathrm{Gr}(n,p)$, with a smooth extension to an open set in the symmetric $n \times n$ matrices, again denoted by X. Let $Y \in T_P\mathrm{Gr}(n,p)$. The Riemannian connection on $\mathrm{Gr}(n,p)$ is then given by

$$\nabla_Y(X) = \Pi_{T_P\mathrm{Gr}}(\mathrm{d}X_P(Y)) = \Pi_{T_P\mathrm{Gr}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} X(P+tY)\right). \tag{2.3.5}$$

It can also be calculated via the horizontal lift,

$$(\nabla_Y(X))^{\mathsf{hor}}_U = \Pi_{\mathsf{Hor}_U \mathrm{St}}(\mathrm{d}(U \mapsto X^{\mathsf{hor}}_U)_U(Y^{\mathsf{hor}}_U)) = (I_n - UU^T)\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} X^{\mathsf{hor}}_{U+tY^{\mathsf{hor}}_U}. \tag{2.3.6}$$

Here, $\mathbb{R}^{n\times p} \ni U \mapsto X^{\mathsf{hor}}_U \in \mathbb{R}^{n\times p}$ is to be understood as a smooth extension to an open subset of $\mathbb{R}^{n\times p}$ of the actual horizontal lift $U \mapsto X^{\mathsf{hor}}_U := (X(UU^T))^{\mathsf{hor}}_U$ from (2.2.13), i.e., fulfilling $\mathrm{d}(\pi^{\mathrm{SG}})_U X^{\mathsf{hor}}_U = X(UU^T)$, where X is the vector field $P \mapsto X(P)$.

Proof. Equation (2.3.5) follows directly from the preceding discussion. It can be checked that (2.3.6) is the horizontal lift of (2.3.5). Alternatively, (2.3.6) can be deduced from [ONe83, Lemma 7.45] by noticing that the horizontal space of the Stiefel manifold is the same for the Euclidean and the canonical metric. Furthermore, (2.3.6) coincides with [AMS04, Theorem 3.4]. □

2.3.3 Gradient

The gradient of a real-valued function on the Grassmannian for the canonical metric was computed in [EAS98] for the Grassmannian with Stiefel representatives, in [HHT07] for the projector perspective and in [AMS04] for the Grassmannian as a quotient of the noncompact Stiefel manifold. For the sake of completeness, we introduce it here as well. The gradient is dual to the differential of a function in the following sense: For a function $f\colon \mathrm{Gr}(n,p) \to \mathbb{R}$, the gradient at P is defined as the unique tangent vector $(\operatorname{grad} f)_P \in T_P\mathrm{Gr}(n,p)$ fulfilling

$$\mathrm{d}f_P(\Delta) = g_P^{\mathrm{Gr}}((\operatorname{grad} f)_P, \Delta)$$

for all $\Delta \in T_P\mathrm{Gr}(n,p)$, where $\mathrm{d}f_P$ denotes the differential of f at P. It is well known that the gradient for the induced Euclidean metric on a manifold is the projection of the Euclidean gradient $\operatorname{grad}^{\mathrm{eucl}}$ to the tangent space. For the Euclidean gradient to be well-defined, f is to be understood as a smooth extension of the actual function f to an open subset of Sym_n. Therefore

$$(\operatorname{grad} f)_P = \Pi_{T_P\mathrm{Gr}}((\operatorname{grad}^{\mathrm{eucl}} f)_P).$$

The function f on $\mathrm{Gr}(n,p)$ can be lifted to the function $\bar{f} := f \circ \pi^{\mathrm{SG}}$ on the Stiefel manifold. Again, when necessary, we identify $\bar{f}$ with a suitable differentiable extension. These two functions are linked by

$$((\operatorname{grad} f)_P)_U^{\mathsf{hor}} = (\operatorname{grad} \bar{f})_U = \Pi_{T_U\mathrm{St}}((\operatorname{grad}^{\mathrm{eucl}} \bar{f})_U) = \Pi_{\mathsf{Hor}_U\mathrm{St}}((\operatorname{grad}^{\mathrm{eucl}} \bar{f})_U),$$

where $\Pi_{T_U\mathrm{St}}(X) = X - \frac{1}{2}U(X^TU + U^TX)$ is the projection of $X \in \mathbb{R}^{n\times p}$ to $T_U\mathrm{St}(n,p)$. The first equality is [AMS04, Equation (3.39)], while the second equality uses the same argument as above. The last equality is due to the fact that the gradient of $\bar{f}$ has no vertical component. For further details see [EAS98; HHT07; AMS04].

2.3.4 Exponential Map

The *exponential map* $\exp_p\colon T_pM \to M$ on a Riemannian manifold M maps a tangent vector $\Delta \in T_pM$ to the endpoint $\gamma(1) \in M$ of the unique geodesic γ that emanates from p in the direction Δ. Thus, geodesics and the Riemannian exponential are related by $\gamma(t) = \exp_p(t\Delta)$. Under a Riemannian submersion $\pi\colon M \to N$, geodesics with horizontal tangent vectors in M are mapped to geodesics in N, cf. [ONe83, Corollary 7.46]. Since the projection $\pi^{\mathrm{OG}}\colon \mathrm{O}(n) \to \mathrm{Gr}(n,p)$ defined in (2.2.3) is a Riemannian submersion by construction, this observation may be used to obtain the Grassmann geodesics.

We start with the geodesics of the orthogonal group. For any Lie group with bi-invariant metric, the geodesics are the one-parameter subgroups, [AB15, §2]. Therefore, the geodesic from $Q \in \mathrm{O}(n)$ in direction $Q\Omega \in T_Q\mathrm{O}(n)$ is calculated via

$$\mathrm{Exp}_Q^{\mathrm{O}}(tQ\Omega) = Q\exp_{\mathrm{m}}(t\Omega),$$

where $\exp_{\mathrm{m}}$ denotes the matrix exponential, see (2.B.2). If $\pi^{\mathrm{OG}}(Q) = P \in \mathrm{Gr}(n,p)$ and $\Delta \in T_P\mathrm{Gr}(n,p)$ with $\Delta_Q^{\mathsf{hor}} = Q\left(\begin{smallmatrix} 0 & -B^T \\ B & 0 \end{smallmatrix}\right) \in \mathsf{Hor}_Q^{\pi^{\mathrm{OG}}}\mathrm{O}(n)$, the geodesic in the Grassmannian is therefore

$$\mathrm{Exp}_P^{\mathrm{Gr}}(t\Delta) = \pi^{\mathrm{OG}}\left(Q\exp_{\mathrm{m}}\left(t\begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix}\right)\right). \tag{2.3.7}$$

This formula, while simple, is not useful for applications with large n, since it involves the matrix exponential of an $n \times n$ matrix. Evaluating the projection π^{OG} leads to the geodesic formula from [Bat+15]:

Proposition 2.3.2 (Grassmann Exponential: Projector Perspective). Let $P \in \mathrm{Gr}(n,p)$ be a point in the Grassmannian and $\Delta \in T_P\mathrm{Gr}(n,p)$. The exponential map is given by

$$\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta) = \exp_{\mathrm{m}}([\Delta, P])P\exp_{\mathrm{m}}(-[\Delta, P]).$$

Proof. With $\Omega = [\Delta, P] = Q\widetilde{\Omega}Q^T \in \mathfrak{so}_P(n)$ and $\widetilde{\Omega} = \left(\begin{smallmatrix} 0 & -B^T \\ B & 0 \end{smallmatrix}\right)$, the horizontal lift of the tangent vector $\Delta = [\Omega, P] \in T_P\mathrm{Gr}(n,p)$ is given by $\Omega Q \in \mathsf{Hor}_Q^{\pi^{\mathrm{OG}}} \mathrm{O}(n)$, see (2.2.7). Then

$$\begin{aligned}\mathrm{Exp}_P^{\mathrm{Gr}}([\Omega, P]) &= Q\exp_{\mathrm{m}}(\widetilde{\Omega})I_{n,p}I_{n,p}^T\exp_{\mathrm{m}}(\widetilde{\Omega}^T)Q^T \\ &= \exp_{\mathrm{m}}(Q\widetilde{\Omega}Q^T)QI_{n,p}I_{n,p}^TQ^T\exp_{\mathrm{m}}(Q\widetilde{\Omega}^TQ^T) = \exp_{\mathrm{m}}(\Omega)P\exp_{\mathrm{m}}(\Omega^T).\end{aligned}$$

□

If $n \gg p$, then working with Stiefel representatives reduces the computational effort immensely. The corresponding geodesic formula appears in [AMS04; EAS98] and is restated in the following proposition. The bracket $[\cdot]$ denotes the equivalence classes from (2.2.10).

Proposition 2.3.3 (Grassmann Exponential: ONB Perspective). For a point $P = UU^T \in \mathrm{Gr}(n,p)$ and a tangent vector $\Delta \in T_P\mathrm{Gr}(n,p)$, let $\Delta_U^{\mathsf{hor}} \in \mathsf{Hor}_U \mathrm{St}(n,p)$ be the horizontal lift of Δ to $\mathsf{Hor}_U \mathrm{St}(n,p)$. Let $r \leq \min(p, n-p)$ be the number of non-zero singular values of Δ_U^{hor}. Denote the thin singular value decomposition (SVD) of Δ_U^{hor} by

$$\Delta_U^{\mathsf{hor}} = \hat{Q}\Sigma V^T,$$

i.e., $\hat{Q} \in \mathrm{St}(n,r), \Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$ and $V \in \mathrm{St}(p,r)$. The Grassmann exponential for the geodesic from P in direction Δ is given by

$$\begin{aligned}\mathrm{Exp}_P^{\mathrm{Gr}}(t\Delta) &= [UV\cos(t\Sigma)V^T + \hat{Q}\sin(t\Sigma)V^T + UV_\perp V_\perp^T] \\ &= [\begin{pmatrix} UV\cos(t\Sigma) + \hat{Q}\sin(t\Sigma) & UV_\perp \end{pmatrix}],\end{aligned} \tag{2.3.8}$$

which does not depend on the chosen orthogonal completion $V_\perp$.

Proof. This is essentially [EAS98, Theorem 2.3] with a reduced storage requirement for $\hat{Q}$ in case of rank-deficient tangent velocity vectors. The thin SVD of B is given by

$$B = U_\perp^T\Delta_U^{\mathsf{hor}} = U_\perp^T\hat{Q}\Sigma V^T$$

with $W := U_\perp^T \hat{Q} \in \mathrm{St}(n-p, r)$, $\Sigma \in \mathbb{R}^{r\times r}$, $V \in St(p,r)$. Let $W_\perp, V_\perp$ be suitable orthogonal completions. Then,

$$\exp_{\mathrm{m}}\left(\begin{smallmatrix} 0 & -B^T \\ B & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} V & V_\perp & 0 & 0 \\ 0 & 0 & W & W_\perp \end{smallmatrix}\right) \begin{pmatrix} \cos(\Sigma) & 0 & -\sin(\Sigma) & 0 \\ 0 & I_{p-r} & 0 & 0 \\ \sin(\Sigma) & 0 & \cos(\Sigma) & 0 \\ 0 & 0 & 0 & I_{n-p-r} \end{pmatrix} \begin{pmatrix} V^T & 0 \\ V_\perp^T & 0 \\ 0 & W^T \\ 0 & W_\perp^T \end{pmatrix},$$

which leads to the desired result when inserted into (2.3.7). The second equality in (2.3.8) is given by a postmultiplication by $\begin{pmatrix} V & V_\perp \end{pmatrix} \in \mathrm{O}(p)$, which does not change the equivalence class. This postmultiplication does however change the Stiefel representative, so $\begin{pmatrix} UV\cos(t\Sigma) + \hat{Q}\sin(t\Sigma) & UV_\perp \end{pmatrix}$ is the Stiefel geodesic from $\begin{pmatrix} UV & UV_\perp \end{pmatrix}$ in direction $\begin{pmatrix} \hat{Q}\Sigma & 0 \end{pmatrix}$. A different orthogonal completion of V does not change the second expression in (2.3.8) and results in a different representative of the same equivalence class in the third expression. □

The formula established in [EAS98] uses the compact SVD $\Delta_U^{\mathsf{hor}} = \tilde{Q}\tilde{\Sigma}\tilde{V}^T$ with $\tilde{Q} \in \mathrm{St}(n,p)$, $\tilde{\Sigma} = \mathrm{diag}(\sigma_1, ..., \sigma_p)$ and $\tilde{V} \in \mathrm{O}(p)$. Then

$$\mathrm{Exp}_P^{\mathrm{Gr}}(t\Delta) = [U\tilde{V}\cos(t\tilde{\Sigma})\tilde{V}^T + \tilde{Q}\sin(t\tilde{\Sigma})\tilde{V}^T]. \tag{2.3.9}$$

By a slight abuse of notation we also define

$$\mathrm{Exp}_U^{\mathrm{Gr}}(t\Delta_U^{\mathsf{hor}}) = U\tilde{V}\cos(t\tilde{\Sigma})\tilde{V}^T + \tilde{Q}\sin(t\tilde{\Sigma})\tilde{V}^T \tag{2.3.10}$$

to be the Grassmann exponential on the level of Stiefel representatives.

2.3.5 Differentiating the Grassmann Exponential

In this section, we compute explicit expressions for the differential $\mathrm{d}(\mathrm{Exp}_P^{\mathrm{Gr}})_\Delta$ of the Grassmann exponential at a tangent location $\Delta \in T_P\mathrm{Gr}(n,p)$. One possible motivation is the computation of Jacobi fields vanishing at a point in Subsection 2.7.1. Another motivation is, e.g., Hermite manifold interpolation as in [Zim20]. Formally, the differential at Δ is the linear map

$$\mathrm{d}(\mathrm{Exp}_P^{\mathrm{Gr}})_\Delta \colon T_\Delta(T_P\mathrm{Gr}(n,p)) \to T_{\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta)}\mathrm{Gr}(n,p). \tag{2.3.11}$$

The tangent space to a linear space can be identified with the linear space itself, so that $T_\Delta(T_P\mathrm{Gr}(n,p)) \cong T_P\mathrm{Gr}(n,p)$. We also exploit this principle in practical computations. We consider the exponential in the form of (2.3.9). The task boils down to computing the directional derivatives

$$\mathrm{d}(\mathrm{Exp}_P^{\mathrm{Gr}})_\Delta(\tilde{\Delta}) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathrm{Exp}_P^{\mathrm{Gr}}(\Delta + t\tilde{\Delta}), \tag{2.3.12}$$

where $\Delta, \tilde{\Delta} \in T_P\mathrm{Gr}(n,p)$. A classical result in Riemannian geometry [Lee18, Prop. 5.19] ensures that for $\Delta = 0 \in T_P\mathrm{Gr}(n,p)$ the derivative is the identity $\mathrm{d}(\mathrm{Exp}_P^{\mathrm{Gr}})_0(\tilde{\Delta}) = \tilde{\Delta}$. For $\Delta \neq 0$, we can proceed as follows:

Proposition 2.3.4 (Derivative of the Grassmann Exponential)**.** Let $P = UU^T \in \mathrm{Gr}(n,p)$ and $\Delta, \tilde{\Delta} \in T_P\mathrm{Gr}(n,p)$ such that Δ_U^{hor} has mutually distinct, non-zero singular values. Furthermore let $\Delta_U^{\mathsf{hor}} = Q\Sigma V^T$ and $(\Delta + t\tilde{\Delta})_U^{\mathsf{hor}} = Q(t)\Sigma(t)V(t)^T$ be the compact SVDs of the horizontal lifts of Δ and $\Delta + t\tilde{\Delta}$, respectively. Denote the derivative of $Q(t)$ evaluated at $t = 0$ by $\dot{Q} = \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} Q(t)$ and likewise for $\Sigma(t)$ and $V(t)$.[2] Let

$$Y := UV\cos(\Sigma) + Q\sin(\Sigma) \in \mathrm{St}(n,p)$$

and

$$\Gamma := U\dot{V}\cos(\Sigma) - UV\sin(\Sigma)\dot{\Sigma} + \dot{Q}\sin(\Sigma) + Q\cos(\Sigma)\dot{\Sigma} \in T_Y\mathrm{St}(n,p).$$

Then the derivative of the Grassmann exponential is given by

$$\mathrm{d}(\mathrm{Exp}_P^{\mathrm{Gr}})_\Delta(\tilde{\Delta}) = \Gamma Y^T + Y\Gamma^T \in T_{\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta)}\mathrm{Gr}(n,p) \subseteq \mathbb{R}^{n\times n}. \tag{2.3.13}$$

The horizontal lift to Y is accordingly

$$\left(\mathrm{d}(\mathrm{Exp}_P^{\mathrm{Gr}})_\Delta(\tilde{\Delta})\right)_Y^{\mathsf{hor}} = (I_n - YY^T)\Gamma = \Gamma + Y\Gamma^T Y \in \mathbb{R}^{n\times p}. \tag{2.3.14}$$

Proof. The curve $\gamma(t) := \mathrm{Exp}_P^{\mathrm{Gr}}(\Delta + t\widetilde{\Delta})$ on the Grassmannian is given by

$$\gamma(t) = \pi^{\mathrm{SG}}\left(UV(t)\cos(\Sigma(t))V(t)^T + Q(t)\sin(\Sigma(t))V(t)^T\right),$$

according to (2.3.9). Note that this is in general not a geodesic in $\mathrm{Gr}(n,p)$ but merely a curve through the endpoints of the geodesics from P in direction $\Delta + t\widetilde{\Delta}$. That is to say, it is the mapping of the (non-radial) straight line $\Delta + t\widetilde{\Delta}$ in $T_P\mathrm{Gr}(n,p)$ to $\mathrm{Gr}(n,p)$ via the exponential map. The projection π^{SG} is not affected by the postmultiplication of $V(t)^T \in \mathrm{O}(p)$, because of the nature of the equivalence classes in $\mathrm{St}(n,p)$. Therefore we set

$$\mu\colon [0,1] \to \mathrm{St}(n,p), \quad \mu(t) := UV(t)\cos(\Sigma(t)) + Q(t)\sin(\Sigma(t))$$

and have $\gamma(t) = \pi^{\mathrm{SG}}(\mu(t))$. The derivative of γ with respect to t evaluated at $t = 0$ is then given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\gamma(t) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\pi^{\mathrm{SG}}(\mu(t)) = \mathrm{d}\pi^{\mathrm{SG}}_{\mu(0)}(\dot{\mu}(0)) = \dot{\mu}(0)\mu(0)^T + \mu(0)\dot{\mu}(0)^T. \tag{2.3.15}$$

But with the definitions above, $Y = \mu(0)$ and $\Gamma = \dot{\mu}(0)$, so (2.3.15) is equivalent to (2.3.13). The horizontal lift of (2.3.15) to Y is according to (2.2.13) given by a post-multiplication of Y, which shows (2.3.14). Note however that $\Gamma \in T_Y\mathrm{St}(n,p)$ is not necessarily horizontal, so $0 \neq \Gamma^T Y \in \mathfrak{so}(p)$. □

[2] The matrices $\dot{Q}, \dot{\Sigma}$ and $\dot{V}$ can be calculated via Algorithm 2.B.2.

In order to remove the "mutually distinct singular values" assumption of Proposition 2.3.4 and to remedy the numerical instability of the SVD in the presence of clusters of singular values, we introduce an alternative computational approach that relies on the derivative of the QR-decomposition rather than that of the SVD. Yet in this case, the "non-zero singular values" assumption is retained, and instabilities may arise for matrices that are close to being rank-deficient.

Let $U, \Delta_U^{\mathsf{hor}}, \tilde{\Delta}_U^{\mathsf{hor}}$ be as introduced in Prop. 2.3.4 (now with possibly repeated singular values of Δ_U^{hor}) and consider the t-dependent QR-decomposition of the matrix curve $(\Delta + t\tilde{\Delta})_U^{\mathsf{hor}} = Q(t)R(t)$. The starting point is (2.3.7), which can be transformed to

$$\gamma(t) = \pi^{\mathrm{SG}}\left((U, Q(t)) \exp_m \begin{pmatrix} 0 & -R(t)^T \\ R(t) & 0 \end{pmatrix} \begin{pmatrix} I_p \\ 0 \end{pmatrix}\right) =: \pi^{\mathrm{SG}}(\tilde{\gamma}(t))$$

by means of elementary matrix operations. Write $M(t) = \begin{pmatrix} 0 & -R(t)^T \\ R(t) & 0 \end{pmatrix}$. By the product rule,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \tilde{\gamma}(t) = (0, \dot{Q}(0)) \exp_m(M(0)) \begin{pmatrix} I_p \\ 0 \end{pmatrix} + (U, Q(0)) \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \exp_m(M(t)) \begin{pmatrix} I_p \\ 0 \end{pmatrix}. \tag{2.3.16}$$

The derivative $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} \exp_m(M(t)) = \mathrm{d}(\exp_m)_{M(0)}(\dot{M}(0))$ can be computed according to Mathias' Theorem [Hig08, Thm 3.6, p. 58] from

$$\begin{aligned} \exp_m \begin{pmatrix} M(0) & \dot{M}(0) \\ 0 & M(0) \end{pmatrix} &= \begin{pmatrix} \exp_m(M(0)) & \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} \exp_m(M(0) + t\dot{M}(0)) \\ 0 & \exp_m(M(0)) \end{pmatrix} \\ &= \begin{pmatrix} \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} & \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \\ \mathbf{0} & \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} \end{pmatrix} \end{aligned}$$

which is a $(4p \times 4p)$-matrix exponential written in sub-blocks of size $(p \times p)$. Substituting in (2.3.16) gives the $\mathcal{O}(np^2)$-formula

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \tilde{\gamma}(t) = \dot{Q}(0)E_{21} + UD_{11} + Q(0)D_{21}. \tag{2.3.17}$$

This corresponds to [Zim20, Lemma 5], which addresses the Stiefel case. The derivative matrices $\dot{Q}(0), \dot{R}(0)$ can be obtained from Alg. 2.B.3 in Appendix 2.B. The final formula is obtained by taking the projection into account as in (2.3.15), where μ is to be replaced by $\tilde{\gamma}$. The horizontal lift is computed accordingly.

The derivative of the Grassmann exponential can also be computed directly in $\mathrm{Gr}(n,p)$ without using horizontal lifts, at the cost of a higher computational complexity, but without restrictions with regard to the singular values. The key is again to apply Mathias' Theorem to evaluate the derivative of the matrix exponential. Let $P \in \mathrm{Gr}(n,p)$ and $\Delta = [\Omega, P]$, $\widetilde{\Delta} = [\widetilde{\Omega}, P] \in T_P\mathrm{Gr}(n,p)$ with $\Omega = (I_n - 2P)\Delta$, $\widetilde{\Omega} = (I_n - 2P)\widetilde{\Delta} \in \mathfrak{so}_P(n)$.

Denote $Q := \exp_{\mathrm{m}}(\Omega) \in \mathrm{O}(n)$ and $\Psi Q = \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} \exp_{\mathrm{m}}(\Omega + t\widetilde{\Omega})$. Here, $\Psi \in \mathfrak{so}(n)$, since $\exp_{\mathrm{m}}(\Omega + t\widetilde{\Omega})$ is a curve in $\mathrm{O}(n)$ through Q at $t = 0$. Then a computation shows that the derivative of

$$\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta + t\widetilde{\Delta}) = \exp_{\mathrm{m}}(\Omega + t\widetilde{\Omega}) P \exp_{\mathrm{m}}(-\Omega - t\widetilde{\Omega})$$

is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathrm{Exp}_P^{\mathrm{Gr}}(\Delta + t\widetilde{\Delta}) = \Psi Q P Q^T + QP(\Psi Q)^T \in T_{QPQ^T}\mathrm{Gr}(n,p).$$

The matrices Q and ΨQ can be obtained in one calculation by evaluating the left side of

$$\exp_{\mathrm{m}}\begin{pmatrix} \Omega & \widetilde{\Omega} \\ 0 & \Omega \end{pmatrix} = \begin{pmatrix} \exp_{\mathrm{m}}(\Omega) & \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} \exp_{\mathrm{m}}(\Omega + t\widetilde{\Omega}) \\ 0 & \exp_{\mathrm{m}}(\Omega) \end{pmatrix} = \begin{pmatrix} Q & \Psi Q \\ 0 & Q \end{pmatrix}$$

according to Mathias' Theorem.

2.3.6 Parallel Transport

On a Riemannian manifold (M, g), *parallel transport* of a tangent vector $v \in T_pM$ along a smooth curve $\gamma\colon I \to M$ through p gives a smooth vector field $V \in \mathfrak{X}(\gamma)$ along γ that is parallel with respect to the Riemannian connection ∇ and fulfills the initial condition $V(p) = v$. A vector field $V \in \mathfrak{X}(\gamma)$ along a curve γ is a vector field that is defined on the range of the curve, i.e., $V\colon \gamma(I) \to TM$ and $V(\gamma(t)) \in T_{\gamma(t)}M$. The term "parallel" means that for all $t \in I$, the covariant derivative of V in direction of the tangent vector of γ vanishes, i.e.

$$\nabla_{\dot{\gamma}(t)} V = 0.$$

Parallel transport on the Grassmannian (ONB perspective) was studied in [EAS98], where an explicit formula for the horizontal lift of the parallel transport of a tangent vector along a geodesic was derived, and in [AMS04], where a differential equation for the horizontal lift of parallel transport along general curves was given. In the next proposition, we complete the picture by providing a formula for the parallel transport on the Grassmannian from the projector perspective. Note that this formula is similar to the parallel transport formula in the preprint [LLY20].

Proposition 2.3.5 (Parallel Transport: Projector Perspective)**.** Let $P \in \mathrm{Gr}(n,p)$ and $\Delta, \Gamma \in T_P\mathrm{Gr}(n,p)$. Then the parallel transport $\mathbb{P}_\Delta(\mathrm{Exp}_P^{\mathrm{Gr}}(t\Gamma))$ of Δ along the geodesic

$$\mathrm{Exp}_P^{\mathrm{Gr}}(t\Gamma) = \exp_{\mathrm{m}}(t[\Gamma, P]) P \exp_{\mathrm{m}}(-t[\Gamma, P])$$

is given by

$$\mathbb{P}_\Delta(\mathrm{Exp}_P^{\mathrm{Gr}}(t\Gamma)) = \exp_{\mathrm{m}}(t[\Gamma, P]) \Delta \exp_{\mathrm{m}}(-t[\Gamma, P]).$$

Proof. Denote $\gamma(t) := \operatorname{Exp}^{\mathrm{Gr}}_P(t\Gamma)$ and note that $\Omega := [\Gamma, P] \in \mathfrak{so}_P(n)$. The fact that $\mathbb{P}_\Delta(\operatorname{Exp}^{\mathrm{Gr}}_P(t\Gamma)) \in T_{\operatorname{Exp}^{\mathrm{Gr}}_P(t\Gamma)}\mathrm{Gr}(n,p)$ can be checked with Proposition 2.2.1 c). To show that $\mathbb{P}_\Delta$ gives parallel transport, we need to show that $\nabla_{\dot\gamma(t)}(\mathbb{P}_\Delta(\gamma(t))) = \Pi_{T_{\gamma(t)}\mathrm{Gr}}\left(\mathrm{d}(\mathbb{P}_\Delta)_{\gamma(t)}(\dot\gamma(t))\right) = 0$ as in (2.3.5). By making use of the chain rule, we have $\mathrm{d}(\mathbb{P}_\Delta)_{\gamma(t)}(\dot\gamma(t)) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_\Delta(\gamma(t)) = [\Omega, \mathbb{P}_\Delta(\gamma(t))]$, where $[\cdot,\cdot]$ denotes the matrix commutator. Applying the projection $\Pi_{T_{\gamma(t)}\mathrm{Gr}}$ from (2.3.3) and making use of the relation (2.2.7) and the tangent vector properties from Proposition 2.2.1 give the desired result. □

Applying the horizontal lift to the parallel transport equation leads to the formula also found in [EAS98]. Let $Q = \begin{pmatrix} U & U_\perp \end{pmatrix} \in (\pi^{\mathrm{OG}})^{-1}(P)$. Then $\Omega = Q\left(\begin{smallmatrix} 0 & -A^T \\ A & 0 \end{smallmatrix}\right)Q^T$ and $\Delta = Q\left(\begin{smallmatrix} 0 & B^T \\ B & 0 \end{smallmatrix}\right)Q^T$ for some $A, B \in \mathbb{R}^{(n-p)\times p}$. According to (2.2.13), the horizontal lift of $\mathbb{P}_\Delta(\operatorname{Exp}^{\mathrm{Gr}}_P(t\Gamma))$ to the Stiefel geodesic representative $U(t) = Q\exp_{\mathrm{m}}(tQ^T\Omega Q)I_{n,p}$ is given by a post-multiplication with $U(t)$,

$$\left(\mathbb{P}_\Delta(\operatorname{Exp}^{\mathrm{Gr}}_P(t\Gamma))\right)^{\mathsf{hor}}_{U(t)} = \mathbb{P}_\Delta(\operatorname{Exp}^{\mathrm{Gr}}_P(t\Gamma))U(t) = Q\exp_{\mathrm{m}}\left(t\begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix}\right)\begin{pmatrix} 0 \\ B \end{pmatrix}.$$

This formula can be simplified similarly to [EAS98, Theorem 2.4] by discarding all principal angles equal to zero. With notation as above, $\Gamma^{\mathsf{hor}}_U = U_\perp A$ and $\Delta^{\mathsf{hor}}_U = U_\perp B$. Let $r \leq \min(p, n-p)$ be the number of non-zero singular values of Γ^{hor}_U. Denote the thin SVD of Γ^{hor}_U by $\Gamma^{\mathsf{hor}}_U = \hat{Q}\Sigma V^T$, where $\hat{Q} \in \mathrm{St}(n,r), \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$ and $V \in \mathrm{St}(p,r)$, which means Σ has full rank. Then $A = U_\perp^T\hat{Q}\Sigma V^T$ with $W := U_\perp^T\hat{Q} \in \mathrm{St}(n-p,r)$. Similarly to the proof of Proposition 2.3.3, with $\gamma_\Gamma(t) := \operatorname{Exp}^{\mathrm{Gr}}_P(t\Gamma)$,

$$\begin{aligned}(\mathbb{P}_\Delta(\gamma_\Gamma(t)))^{\mathsf{hor}}_{U(t)} &= \left(-UV\sin(t\Sigma)W^T + U_\perp W\cos(t\Sigma)W^T + U_\perp(I_{n-p} - WW^T)\right)B \\ &= (-UV\sin(t\Sigma)\hat{Q}^T + \hat{Q}\cos(t\Sigma)\hat{Q}^T + I_n - \hat{Q}\hat{Q}^T)\Delta^{\mathsf{hor}}_U.\end{aligned}$$

The difference between this formula and the one found from [EAS98, Theorem 2.4] is in the usage of the thin SVD and the therefore smaller matrices $\hat{Q}, \Sigma$ and V, depending on the problem. But the first line also shows that if $r = n - p$, the term $I_{n-p} - WW^T$ vanishes, and therefore also the term $(I_n - \hat{Q}\hat{Q}^T)\Delta^{\mathsf{hor}}_U$. This can happen if $p \geq n/2$.

2.4 Symmetry and Curvature

In this section, we establish the symmetric space structure of the Grassmann manifold by elementary means. The symmetric structure of the Grassmannian was for example shown in [KN96, Vol. II] and [BN91].

Exploiting the symmetric space structure, the curvature of the Grassmannian can be calculated explicitly. Curvature formulae for symmetric spaces can be found for example in [ONe83] and [KN96, Vol. II]. To the best of the authors' knowledge, a first formula for the sectional curvature of the Grassmannian was given in [Won68b], without making use of the symmetric structure. The bounds were studied in [WC88]. In

[Lei61], curvature formulae have been derived in local coordinates via differential forms. Explicit curvature formulae for a generalized version of the Grassmannian as the space of orthogonal projectors were given in [MS85].

Curvature bounds are required for the analysis of Riemannian optimization problems and, in particular, for studying the Riemannian centers of mass, see for example [Bou14] and [Li+19], and several references therein. The sectional curvature features also in statistical problems on Riemannian manifolds [CV19], and enables estimates for data processing errors on manifolds [Zim20].

2.4.1 Symmetric Space Structure

In differential geometry, a *metric symmetry at q* is an isometry $\sigma : M \to M$ of a manifold M that fixes a certain point $\sigma(q) = q$ with the additional property that $d\sigma_q = -\operatorname{id}|_{T_qM}$. This relates to the concept of a point reflection in Euclidean geometry. A (metric) *symmetric space* is a connected differentiable manifold that has a metric symmetry at every point, [ONe83, §8]. Below, we execute an explicit construction of symmetries for the Grassmannian, which compares to the abstract course of action in [ONe83, §11, p. 315ff].

Consider the orthogonal matrix $S_0 = \begin{pmatrix} I_p & 0 \\ 0 & -I_{n-p} \end{pmatrix} \in \mathrm{O}(n)$. Then S_0 induces a symmetry at P_0 via $\sigma^{P_0} : P \mapsto P^{S_0} := S_0 P S_0^T$, which is defined on all of $\mathrm{Gr}(n,p)$. Obviously, $\sigma^{P_0}(P_0) = P_0$. For any point $P \in \mathrm{Gr}(n,p)$ and any tangent vector $\Delta \in T_P\mathrm{Gr}(n,p)$, the differential in direction Δ can be computed as $\mathrm{d}\sigma_P^{P_0}(\Delta) = \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\sigma(P(t))$, where $P(t)$ is any curve on $\mathrm{Gr}(n,p)$ with $P(0) = P$ and $\dot{P}(0) = \Delta$. This gives

$$\mathrm{d}\sigma_{P_0}^{P_0} : T_{P_0}\mathrm{Gr}(n,p) \to T_{P_0}\mathrm{Gr}(n,p), \quad \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix} \mapsto S_0 \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix} S_0^T = -\begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix},$$

so that σ^{P_0} is indeed a symmetry of $\mathrm{Gr}(n,p)$ at P_0.

Given any other point $P \in \mathrm{Gr}(n,p)$, we can compute the EVD $P = QP_0Q^T$ and define $\sigma^P : \tilde{P} \mapsto (QS_0Q^T)\tilde{P}(QS_0Q^T)$. This isometry fixes P, $\sigma^P(P) = P$. Moreover, for any curve with $P(0) = P$, $\dot{P}(0) = \Delta \in T_P\mathrm{Gr}(n,p)$, it holds $\Delta = \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0} Q(t)P_0Q^T(t) = \dot{Q}P_0Q^T + QP_0\dot{Q}^T$ (evaluated at $t = 0$). Since $Q(t)$ is a curve on $O(n)$, it holds $Q^T\dot{Q} = -\dot{Q}^TQ$, so that $Q^T\dot{Q} = \begin{pmatrix} C_{11} & -C_{21}^T \\ C_{21} & C_{22} \end{pmatrix}$ is skew. As a consequence, we use the transformation $Q^T\Delta Q = \begin{pmatrix} 0 & C_{21}^T \\ C_{21} & 0 \end{pmatrix}$ to move Δ to the tangent space at P_0 and compute

$$\mathrm{d}\sigma_P^P(\Delta) = QS_0(Q^T\Delta Q)S_0Q^T = QS_0 \begin{pmatrix} 0 & C_{21}^T \\ C_{21} & 0 \end{pmatrix} S_0Q^T = -Q(Q^T\Delta Q)Q^T = -\Delta.$$

Hence, we have constructed metric symmetries at every point of $\mathrm{Gr}(n,p)$.

The symmetric space structure of $\mathrm{Gr}(n,p)$ implies a number of strong properties. First of all, it follows that $\mathrm{Gr}(n,p)$ is *geodesically complete* [ONe83, Chap. 8, Lemma 20]. This means that the maximal domain of definition for all Grassmann geodesics is the whole real line $\mathbb{R}$. As a consequence, all the statements of the Hopf-Rinow Theorem [Car92, Chap. 7, Thm 2.8], [AB15, Thm 2.9] hold for the Grassmannian:

1. The Riemannian exponential $\mathrm{Exp}^{\mathrm{Gr}}_P \colon T_P\mathrm{Gr}(n,p) \to \mathrm{Gr}(n,p)$ is globally defined.

2. $(\mathrm{Gr}(n,p), \mathrm{dist}(\cdot,\cdot))$ is a complete metric space, where $\mathrm{dist}(\cdot,\cdot)$ is the Riemannian distance function.

3. Every closed and bounded set in $\mathrm{Gr}(n,p)$ is compact.

These statements are equivalent. Any one of them additionally implies

4. For any two points $P_1, P_2 \in \mathrm{Gr}(n,p)$, there exists a geodesic γ of length $L(\gamma) = \mathrm{dist}(P_1,P_2)$ that joins P_1 to P_2; hence any two points can be joined by a *minimal* geodesic segment.

5. The exponential map $\mathrm{Exp}^{\mathrm{Gr}}_P \colon T_P\mathrm{Gr}(n,p) \to \mathrm{Gr}(n,p)$ is surjective for all $P \in \mathrm{Gr}(n,p)$.

2.4.2 Sectional Curvature

For $X, Y, Z \in \mathbb{R}^{(n-p)\times p}$, let $\hat{X} := \left(\begin{smallmatrix} 0 & -X^T \\ X & 0 \end{smallmatrix}\right) \in \mathsf{Hor}_I\, \mathrm{O}(n)$ and $\hat{Y}, \hat{Z} \in \mathsf{Hor}_I\, \mathrm{O}(n)$ accordingly. Denote the projections to $T_{P_0}\mathrm{Gr}(n,p)$ by $x := \mathrm{d}\pi^{\mathrm{OG}}_{P_0}(\hat{X}) = \left(\begin{smallmatrix} 0 & X^T \\ X & 0 \end{smallmatrix}\right) \in T_{P_0}\mathrm{Gr}(n,p)$, etc. Then, by [ONe83, Proposition 11.31], the curvature tensor at P_0 is given by $R_{xy}z = \mathrm{d}\pi^{\mathrm{OG}}_{P_0}([\hat{Z},[\hat{X},\hat{Y}]])$, since the Grassmannian is symmetric and therefore also reductive homogeneous. This formula coincides with the formula found in [MS85]. Explicitly, we can calculate

$$R_{xy}z = \begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix} \in T_{P_0}\mathrm{Gr}(n,p),$$

where $B = ZX^TY - ZY^TX - XY^TZ + YX^TZ \in \mathbb{R}^{(n-p)\times p}$.

The sectional curvature of the Grassmannian can be calculated by the following formulae. It depends only on the plane spanned by two given tangent vectors, not the spanning vectors themselves. For a Riemannian manifold, the sectional curvature completely determines the curvature tensor, see for example [Lee18, Proposition 8.31].

Proposition 2.4.1. Let $P \in \mathrm{Gr}(n,p)$ and let $\Delta_1, \Delta_2 \in T_P\mathrm{Gr}(n,p)$ span a non-degenerate plane in $T_P\mathrm{Gr}(n,p)$. The sectional curvature is then given by

$$K_P(\Delta_1,\Delta_2) = 4\frac{\mathrm{tr}\big(\Delta_1^2\Delta_2^2\big) - \mathrm{tr}\big((\Delta_1\Delta_2)^2\big)}{\mathrm{tr}(\Delta_1^2)\,\mathrm{tr}(\Delta_2^2) - (\mathrm{tr}(\Delta_1\Delta_2))^2} = 2\frac{\|[\Delta_1,\Delta_2]\|_F^2}{\|\Delta_1\|_F^2\|\Delta_2\|_F^2 - \langle\Delta_1,\Delta_2\rangle_0^2}. \tag{2.4.1}$$

Proof. This formula can be derived from the result in [MS85]. For a direct proof, we proceed as follows. The tangent vectors can be expressed as $\Delta_1 = [\Omega_1, P]$, $\Delta_2 = [\Omega_2, P] \in T_P\mathrm{Gr}(n,p)$ for some $\Omega_1, \Omega_2 \in \mathfrak{so}_P(n)$. Using the fact

$$[\Omega_1, P][\Omega_2, P] = -\Omega_1\Omega_2,$$

we see that

$$\mathrm{tr}\big([\Omega_2, P][\Omega_1, P]^2[\Omega_2, P]\big) - \mathrm{tr}\big(([\Omega_2, P][\Omega_1, P])^2\big) = \mathrm{tr}(\Omega_2\Omega_1[\Omega_1, \Omega_2])\,.$$

The property that for any two $X, Y \in \mathsf{Hor}_I\, \mathrm{O}(n)$ the equality

$$\mathrm{tr}(YX[X,Y]) = \langle [Y,[X,Y]], X\rangle$$

holds, shows the claim according to [ONe83, Proposition 11.31]. □

With (2.2.11) every $\Delta_i \in T_P\mathrm{Gr}(n,p)$ can be written as $\Delta_i = U_\perp B_i U^T + U B_i^T U_\perp^T$ for some $\begin{pmatrix} U & U_\perp \end{pmatrix} \in (\pi^{\mathrm{OG}})^{-1}(P)$ and $B_i \in \mathbb{R}^{(n-p)\times p}$. Since every tangent vector in $T_P\mathrm{Gr}(n,p)$ is uniquely determined by such a B for a chosen representative $\begin{pmatrix} U & U_\perp \end{pmatrix}$, we can insert this into (2.4.1) and get the simplified formula

$$\begin{aligned} K_P(B_1, B_2) &= \frac{\mathrm{tr}\big(B_1^T B_2 \left(B_2^T B_1 - 2B_1^T B_2\right) + B_1^T B_1 B_2^T B_2\big)}{\mathrm{tr}\big(B_1^T B_1\big)\,\mathrm{tr}\big(B_2^T B_2\big) - \big(\mathrm{tr}\big(B_1^T B_2\big)\big)^2} \\ &= \frac{\|B_2^T B_1\|_F^2 + \|B_1 B_2^T\|_F^2 - 2\langle B_2^T B_1, B_1^T B_2\rangle_0}{\|B_1\|_F^2\|B_2\|_F^2 - \langle B_1, B_2\rangle_0^2}. \end{aligned} \tag{2.4.2}$$

This formula is equivalent to the slightly more extended form in [Won68b] and depends only on the factors $B_1^T B_2$, $B_1^T B_1$ and $B_2^T B_2 \in \mathbb{R}^{p\times p}$. It also holds for the horizontal lifts of Δ_i by just replacing the symbols B_i by $(\Delta_i)_U^{\mathsf{hor}}$, which can also be shown by exploiting (2.2.11) and $(\Delta_i)_U^{\mathsf{hor}} = U_\perp B_i$.

In summary, for two orthonormal tangent vectors $\Delta_1 = [\Omega_1, P]$, $\Delta_2 = [\Omega_2, P] \in T_P\mathrm{Gr}(n,p)$ with $\Omega_1, \Omega_2 \in \mathfrak{so}_P(n)$, i.e.

$$1 = \langle \Delta_i, \Delta_i\rangle = \frac{1}{2}\,\mathrm{tr}(\Delta_i^T \Delta_i) \text{ and } 0 = \langle \Delta_1, \Delta_2\rangle$$

the sectional curvature is given by

$$\begin{aligned} K_P(\Delta_1, \Delta_2) &= \mathrm{tr}(\Omega_2\Omega_1[\Omega_1, \Omega_2]) \\ &= \mathrm{tr}\left(\Delta_{2,U}^{\mathsf{hor}^T} \Delta_{1,U}^{\mathsf{hor}} \left(\Delta_{1,U}^{\mathsf{hor}^T} \Delta_{2,U}^{\mathsf{hor}} - 2\Delta_{2,U}^{\mathsf{hor}^T} \Delta_{1,U}^{\mathsf{hor}}\right) + \Delta_{1,U}^{\mathsf{hor}^T} \Delta_{1,U}^{\mathsf{hor}} \Delta_{2,U}^{\mathsf{hor}^T} \Delta_{2,U}^{\mathsf{hor}}\right). \end{aligned}$$

Inserting any pair of orthonormal tangent vectors shows that for $n > 2$, the sectional curvature of the real projective space $\mathrm{Gr}(n,1) = \mathbb{RP}^{n-1}$ is constant $K_P \equiv 1$, as it is by the same calculation for $\mathrm{Gr}(n, n-1)$, see also [Won68b]. The same source also states a list of facts about the sectional curvature on $\mathrm{Gr}(n,p)$ without proof, especially that

$$0 \leq K_P(\Delta_1, \Delta_2) \leq 2 \tag{2.4.3}$$

for $\min(p, n-p) \geq 2$. Nonnegativity follows directly from (2.4.1). The upper bound was proven in [WC88], by proving that for any two matrices $A, B \in \mathbb{R}^{m\times n}$, with $m, n \geq 2$, the inequality

$$\|AB^T - BA^T\|_F^2 \leq 2\|A\|_F^2\|B\|_F^2 \tag{2.4.4}$$

holds. Note that (2.4.2) can be rewritten as

$$K_P(B_1, B_2) = \frac{\frac{1}{2}\left(\|B_1B_2^T - B_2B_1^T\|_F^2 + \|B_1^TB_2 - B_2^TB_1\|_F^2\right)}{\|B_1\|_F^2\|B_2\|_F^2 - (\mathrm{tr}(B_1^TB_2))^2}.$$

The bounds of the sectional curvature (2.4.3) are sharp for all cases except those mentioned in the next paragraph: The lower bound zero is attained whenever Δ_1, Δ_2 commute. The upper curvature bound is attained, e.g., for $B_1 = \left(\begin{smallmatrix} 1 & 1 \\ -1 & 1\end{smallmatrix}\right)$, $B_2 = \left(\begin{smallmatrix} -1 & 1 \\ -1 & -1\end{smallmatrix}\right)$, or matrices containing B_1 and B_2 as their top-left block and else only zeros, when $p > 2$.

In [Lei61] it was shown that a Grassmannian $\mathrm{Gr}(n,p)$ features a strictly positive sectional curvature K_P only if the sectional curvature is constant throughout. The sectional curvature is constant (and equal to $K_P \equiv 1$) only in the cases $p = 1,\ n > 2$ or $p = n-1,\ n > 2$. In the case of $n = 2,\ p = 1$, the sectional curvature is not defined, since $\dim(\mathrm{Gr}(2,1)) = 1$. Hence, in this case, there are no non-degenerate two-planes in the tangent space.

2.5 Cut Locus and Riemannian Logarithm

We have seen in Section 2.4.1 that $\mathrm{Gr}(n,p)$ is a complete Riemannian manifold. On such manifolds, the *cut locus* of a point P consists of those points F beyond which the geodesics starting at P cease to be length-minimizing. It is known [Sak96, Ch. III, Prop. 4.1] that P and F are in each other's cut locus if there is more than one shortest geodesic from P to F. We will see that, on the Grassmannian, this "if" is an "if and only if", and moreover "more than one" is always either two or infinitely many.

To get an intuitive idea of the cut locus, think of the earth as an ideal sphere. Then the cut locus of the north pole is the south pole, as it is the only point beyond which the geodesics starting at the north pole cease to be length-minimizing. In the case of the sphere, the "if and only if" statement that we just mentioned for the Grassmannian also holds; however, for the sphere, "more than one" is always infinitely many.

Given two points $P, F \in \mathrm{Gr}(n,p)$ that are not in each other's cut locus, the unique smallest tangent vector $\Delta \in T_P\mathrm{Gr}(n,p)$ such that $\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta) = F$ is called the *Riemannian logarithm* of F at P. We propose an algorithm that calculates the Riemannian logarithm. Moreover, in the case of cut points, the algorithm is able to return any of the (two or infinitely many) smallest $\Delta \in T_P\mathrm{Gr}(n,p)$ such that $\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta) = F$. This ability comes from the indeterminacy of the SVD operation invoked by the algorithm.

The horizontal lift of the exponential map (2.3.9) depends explicitly on the so called *principal angles* between two points and allows us to give explicit formulae for different geodesics between P and a cut point F. We observe that the inherent ambiguity of the SVD, see Appendix 2.B, corresponds to the different geodesics connecting the same points.

Our approach allows data processing schemes to explicitly map any given set of points on the Grassmannian to any tangent space $T_P\mathrm{Gr}(n,p)$, with the catch that possibly a subset of the points (namely those that are in the cut locus of P), is mapped to a set of tangent vectors each, instead of just a single one.

The cut locus, and the related injectivity radius, play an important role in curve fitting methods on manifolds [GMA19] and the analysis of Riemannian optimization problems [ATV13]. The ability to tackle cut points numerically is of special importance for computing so-called *almost gradients*, which enable the computation of Riemannian barycenters for not necessarily localized point sets, see [ATV13, Section 6.2].

2.5.1 Cut Locus

We can introduce the cut locus of the Grassmannian by applying the definitions of [Lee18, Chap. 10] about cut points to $\mathrm{Gr}(n,p)$. In the following, let $P \in \mathrm{Gr}(n,p)$ and $\Delta \in T_P\mathrm{Gr}(n,p)$ and $\gamma_\Delta \colon t \mapsto \mathrm{Exp}_P^{\mathrm{Gr}}(t\Delta)$. Then the *cut time of* (P,Δ) is defined as

$$t_{\mathrm{cut}}(P,\Delta) := \sup\{b > 0 \mid \text{the restriction of } \gamma_\Delta \text{ to } [0,b] \text{ is minimizing}\}.$$

The *cut point of* P *along* γ_Δ is given by $\gamma_\Delta(t_{\mathrm{cut}}(P,\Delta))$ and the *cut locus of* P is defined as

$$\mathrm{Cut}_P := \{F \in \mathrm{Gr}(n,p) \mid F = \gamma_\Delta(t_{\mathrm{cut}}(P,\Delta)) \text{ for some } \Delta \in T_P\mathrm{Gr}(n,p)\}.$$

In [Won67; Sak77], it is shown that the cut locus of $P = UU^T \in \mathrm{Gr}(n,p)$ is the set of all (projectors onto) subspaces with at least one direction orthogonal to all directions in the subspace onto which P projects, i.e.

$$\mathrm{Cut}_P = \left\{F = YY^T \in \mathrm{Gr}(n,p) \;\middle|\; \mathrm{rank}(U^TY) < p\right\}. \tag{2.5.1}$$

This means that the cut locus can be described in terms of *principal angles*: The principal angles $\theta_1,\dots,\theta_p \in [0,\frac{\pi}{2}]$ between two subspaces $\mathcal{U}$ and $\widetilde{\mathcal{U}}$ are defined recursively by

$$\cos(\theta_k) := u_k^T v_k := \max_{\substack{u \in \mathcal{U},\, \|u\|=1\\ u \perp u_1,\dots,u_{k-1}}} \; \max_{\substack{v \in \widetilde{\mathcal{U}},\, \|v\|=1\\ v \perp v_1,\dots v_{k-1}}} u^T v.$$

They can be computed via $\theta_k := \arccos(s_k) \in [0,\frac{\pi}{2}]$, where $s_k \leq 1$ is the k-largest singular value of $U^T\tilde{U} \in \mathbb{R}^{p\times p}$ for any two Stiefel representatives U and $\tilde{U}$. According to this definition, the principal angles are listed in ascending order: $0 \leq \theta_1 \leq \ldots \leq \theta_p \leq \frac{\pi}{2}$. In other words, the cut locus of P consists of all points $F \in \mathrm{Gr}(n,p)$ with at least one principal angle between P and F being equal to $\frac{\pi}{2}$.

Furthermore, as in [Lee18], we introduce the *tangent cut locus of P* by

$$\mathrm{TCL}_P := \{\Delta \in T_P\mathrm{Gr}(n,p) \mid \|\Delta\| = t_{\mathrm{cut}}(P, \Delta/\|\Delta\|)\}$$

and the *injectivity domain of P* by

$$\mathrm{ID}_P := \{\Delta \in T_P\mathrm{Gr}(n,p) \mid \|\Delta\| < t_{\mathrm{cut}}(P, \Delta/\|\Delta\|)\}.$$

The cut time can be explicitly calculated by the following proposition.

Proposition 2.5.1. Let $P = UU^T \in \mathrm{Gr}(n,p)$ and $\Delta \in T_P\mathrm{Gr}(n,p)$. Denote the largest singular value of $\Delta_U^{\mathsf{hor}} \in \mathsf{Hor}_U \mathrm{St}(n,p)$ by σ_1. Then

$$t_{\mathrm{cut}}(P, \Delta) = \frac{\pi}{2\sigma_1}. \tag{2.5.2}$$

Proof. Since $\gamma_\Delta(t_{\mathrm{cut}}(P,\Delta)) \in \mathrm{Cut}_P$, by (2.3.9) we have

$$\mathrm{rank}(U^T(UV\cos(t_{\mathrm{cut}}(P,\Delta)\Sigma)V^T + \hat{Q}\sin(t_{\mathrm{cut}}(P,\Delta)\Sigma)V^T)) < p,$$

which is equivalent to $\cos(t_{\mathrm{cut}}(P,\Delta)\sigma_1) = 0$. □

Now we see that the tangent cut locus TCL_P consists of those tangent vectors for which σ_1 (the largest singular value of the horizontal lift) fulfills $\sigma_1 = \frac{\pi}{2}$ and the injectivity domain ID_P contains the tangent vectors with $\sigma_1 < \frac{\pi}{2}$.

The *geodesic distance* is a natural notion of distance between two points on a Riemannian manifold. It is defined as the length of the shortest curve(s) between two points as measured with the Riemannian metric, if such a curve exists. On the Grassmannian, it can be calculated as the two-norm of the vector of principal angles between the two subspaces, cf. [Won67], i.e.

$$\mathrm{dist}(\mathcal{U}, \widetilde{\mathcal{U}}) = \left(\sum_{i=1}^{p} \sigma_i^2\right)^{\frac{1}{2}}. \tag{2.5.3}$$

This shows that for any two points on the Grassmann manifold $\mathrm{Gr}(n,p)$, the geodesic distance is bounded by

$$\mathrm{dist}(\mathcal{U}, \widetilde{\mathcal{U}}) \leq \sqrt{p}\frac{\pi}{2},$$

which was already stated in [Won67, Theorem 8].

Remark. There are other notions of distance on the Grassmannian that can also be computed from the principal angles, but which are not equal to the geodesic distance, see [EAS98, §4.5], [QZL05], [YL16, Table 2]. In the latter reference, it is also shown that all these distances can be generalized to subspaces of different dimensions by introducing Schubert varieties and adding $\frac{\pi}{2}$ for the "missing" angles.

The *injectivity radius* at $P \in \mathrm{Gr}(n,p)$ is defined as the distance from P to its cut locus, or equivalently, as the supremum of the radii r for which $\mathrm{Exp}^{\mathrm{Gr}}_P$ is a diffeomorphism from the open ball $B_r(0) \subset T_P\mathrm{Gr}(n,p)$ onto its image. The injectivity radius at every P is equal to $\mathrm{inj}(P) = \frac{\pi}{2}$, since there is always a subspace F for which the principal angles between P and F are all equal to zero, except one, which is equal to $\frac{\pi}{2}$. For such an F it holds that $\mathrm{dist}(P,F) = \frac{\pi}{2}$, c.f. (2.5.3), and $F \in \mathrm{Cut}_P$. For all other points $\widetilde{F}$ with $\mathrm{dist}(P,\widetilde{F}) < \frac{\pi}{2}$, all principal angles are strictly smaller than $\frac{\pi}{2}$, and therefore $\widetilde{F} \notin \mathrm{Cut}_P$.

Proposition 2.5.2. Let $P = UU^T \in \mathrm{Gr}(n,p)$ and $\Delta \in T_P\mathrm{Gr}(n,p)$. Consider the geodesic segment $\gamma_\Delta \colon [0,1] \ni t \mapsto \mathrm{Exp}^{\mathrm{Gr}}_P(t\Delta)$. Let the SVD of the horizontal lift of Δ be given by $\hat{Q}\Sigma V^T = \Delta^{\mathsf{hor}}_U \in \mathsf{Hor}_U\,\mathrm{St}(n,p)$, where $\Sigma = \mathrm{diag}(\sigma_1,\dots,\sigma_p)$.

a) If the largest singular value $\sigma_1 < \pi/2$, then the geodesic segment γ_Δ is unique minimizing.

b) If the largest singular value $\sigma_1 = \pi/2$, then the geodesic segment γ_Δ is non-unique minimizing.

c) If the largest singular value $\sigma_1 > \pi/2$, then the geodesic segment γ_Δ is not minimizing.

Proof. In case of a), γ_Δ is minimizing by definition of the cut locus. It is unique by [Lee18, Thm. 10.34 c)]. In case of b), γ_Δ is still minimizing by the definition of the cut locus. For non-uniqueness, replace σ_1 by $-\frac{\pi}{2}$ (instead of $\frac{\pi}{2}$) and observe that we get a different geodesic with the same length and same endpoints. Case c) holds by definition of the cut locus. □

2.5.2 Riemannian Logarithm

For any $P \in \mathrm{Gr}(n,p)$, the restriction of $\mathrm{Exp}^{\mathrm{Gr}}_P$ to the injectivity domain ID_P is a diffeomorphism onto $\mathrm{Gr}(n,p) \setminus \mathrm{Cut}_P$ by [Lee18, Theorem 10.34]. This means that for any $F \in \mathrm{Gr}(n,p) \setminus \mathrm{Cut}_P$ there is a unique tangent vector $\Delta \in \mathrm{ID}_P$ such that $\mathrm{Exp}^{\mathrm{Gr}}_P(\Delta) = F$. The mapping that finds this Δ is conventionally called the *Riemannian logarithm*. Furthermore, [Lee18, Thm. 10.34] states that the restriction of $\mathrm{Exp}^{\mathrm{Gr}}_P$ to the union of the injectivity domain and the tangent cut locus $\mathrm{ID}_P \cup \mathrm{TCL}_P$ is surjective. Therefore for any $F \in \mathrm{Cut}_P$ we find a (non-unique) tangent vector which is mapped to F via the exponential map. We propose Algorithm 2.5.3, which computes the unique $\Delta \in \mathrm{ID}_P \subset T_P\mathrm{Gr}(n,p)$ in case of $F \in \mathrm{Gr}(n,p) \setminus \mathrm{Cut}_P$ and one possible

$\Delta \in \mathrm{TCL}_P \subset \mathrm{Gr}(n,p)$ for $F \in \mathrm{Cut}_P$. In the latter case, all other possible $\tilde{\Delta} \in \mathrm{TCL}_P$ such that $\mathrm{Exp}^{\mathrm{Gr}}_P(\tilde{\Delta}) = F$ can explicitly be derived from that result.

Algorithm 2.5.3 Extended Grassmann Logarithm with Stiefel representatives

Input: $U, Y \in \mathrm{St}(n,p)$ representing $P = UU^T,\ F = YY^T \in \mathrm{Gr}(n,p)$, respectively

1: $\widetilde{Q}\widetilde{S}\widetilde{R}^T \overset{\text{SVD}}{:=} Y^T U$ ▷ SVD
2: $Y_* := Y(\widetilde{Q}\widetilde{R}^T)$ ▷ Procrustes processing
3: $\hat{Q}\hat{S}R^T \overset{\text{SVD}}{:=} (I_n - UU^T)Y_*$ ▷ compact SVD
4: $\Sigma := \arcsin(\hat{S})$ ▷ element-wise on the diagonal
5: $\Delta^{\mathsf{hor}}_U := \hat{Q}\Sigma R^T$

Output: smallest $\Delta^{\mathsf{hor}}_U \in \mathsf{Hor}_U \mathrm{St}(n,p)$ such that $\mathrm{Exp}^{\mathrm{Gr}}_P(\Delta) = F$

Remark: In Step 1, the expression $\overset{\text{SVD}}{:=}$ is to be understood as "is *an* SVD". In case of $F \in \mathrm{Cut}_P$, i.e. singular values equal to zero, different choices of decompositions lead to different valid output vectors Δ^{hor}_U. The non-uniqueness of the compact SVD in Step 3 does not matter, because $\Sigma = \arcsin(\hat{S})$, and arcsin maps zero to zero and repeated singular values to repeated singular values. Therefore any non-uniqueness cancels out in the definition of Δ^{hor}_U.

Before we prove the claimed properties of Algorithm 2.5.3, let us state the following: An algorithm for the Grassmann logarithm with Stiefel representatives only was derived in [AMS04]. The Stiefel representatives are however not retained in this algorithm, i.e., coupling the exponential map and the logarithm recovers the input subspace but produces a different Stiefel representative $\tilde{Y} = \mathrm{Exp}^{\mathrm{Gr}}_U(\mathrm{Log}^{\mathrm{Gr}}_U(Y)) \neq Y$ as an output. Furthermore, it requires the matrix inverse of U^TY, which also means that it only works for points not in the cut locus, see (2.5.1). By slightly modifying this algorithm we get Algorithm 2.5.3, which retains the Stiefel representative, does not require the calculation of the matrix inverse $(U^TY)^{-1}$ and works for all pairs of points. The computational procedure of Algorithm 2.5.3 was first published in the book chapter preprint [Zim19].

In the following Theorem 2.5.4, we show that Algorithm 2.5.3 indeed produces the Grassmann logarithm for points not in the cut locus.

Theorem 2.5.4. Let $P = UU^T \in \mathrm{Gr}(n,p)$ and $F = YY^T \in \mathrm{Gr}(n,p) \setminus \mathrm{Cut}_P$ be two points on the Grassmannian. Then Algorithm 2.5.3 computes the horizontal lift of the Grassmann logarithm $\mathrm{Log}^{\mathrm{Gr}}_P(F) = \Delta \in T_P\mathrm{Gr}(n,p)$ to $\mathsf{Hor}_U \mathrm{St}(n,p)$. It retains the Stiefel representative Y_* when coupled with the Grassmann exponential on the level of Stiefel representatives (2.3.10), i.e.

$$Y_* = \mathrm{Exp}^{\mathrm{Gr}}_U(\Delta^{\mathsf{hor}}_U).$$

Proof. First, Algorithm 2.5.3 aligns the given subspace representatives U and Y by producing a representative of the equivalence class $[Y]$ that is "closest" to U. To this end, the Procrustes method is used, cf. [Hig08]. Procrustes gives

$$QR^T = \arg\min_{\Phi \in O(p)} \|U - Y\Phi\|_F,$$

by means of the SVD

$$Y^T U = QSR^T, \tag{2.5.4}$$

chosen here to be with singular values in ascending order from the top left to the bottom right. Therefore $Y_* := YQR^T$ represents the same subspace $[Y_*] = [Y]$, but

$$U^T Y_* = RSR^T$$

is symmetric. Now, we can split Y_* with the projector $P = UU^T$ onto $\operatorname{span}(U)$ and the projector $I_n - UU^T$ onto the orthogonal complement of $\operatorname{span}(U)$ via

$$Y_* = UU^T Y_* + (I_n - UU^T)Y_* = URSR^T + (I_n - UU^T)Y_*. \tag{2.5.5}$$

If we denote the part of Y_* that lies in $\operatorname{span}(U)^\perp$ by $L := (I_n - UU^T)Y_*$, we see that

$$L^T L = Y_*^T (I_n - UU^T) Y_* = I_n - RS^2R^T = R(I_n - S^2)R^T.$$

That means that $\hat{S} := \sqrt{(I_n - S^2)}$ is the diagonal matrix of singular values of L, with the singular values in descending order. The square root is well-defined, since $(I_n - S^2)$ is diagonal with values between 0 and 1. Note also that the column vectors of R are a set of orthonormal eigenvectors of $L^T L$, i.e., a compact singular value decomposition of L is of the form

$$L = (I_n - UU^T)Y_* = \hat{Q}\hat{S}R^T, \tag{2.5.6}$$

where again $\hat{Q} \in \mathrm{St}(n,p)$. Define $\Sigma := \arccos(S)$, where the arcus cosine (and sine and cosine in the following) is applied entry-wise on the diagonal. Then $S = \cos(\Sigma)$ and $\hat{S} = \sin(\Sigma)$. Inserting in (2.5.5) gives

$$Y_* = UR\cos(\Sigma)R^T + \hat{Q}\sin(\Sigma)R^T.$$

This is exactly the exponential with Stiefel representatives (2.3.10), i.e., $\mathrm{Exp}_U^{\mathrm{Gr}}(\Delta_U^{\mathsf{hor}}) = Y_*$, where $\Delta_U^{\mathsf{hor}} = \hat{Q}\Sigma R^T$. We also see that the exact matrix representative Y_*, and not just any equivalent representative, is computed by plugging Δ_U^{hor} into the exponential $\mathrm{Exp}_U^{\mathrm{Gr}}$.

The singular value decomposition in (2.5.4) differs from the usual SVD – with singular values in descending order – only by a permutation of the columns of Q and R. But if $Y^T U = QSR^T$ is an SVD with singular values in ascending order and $Y^T U = \widetilde{Q}\widetilde{S}\widetilde{R}^T$ is an SVD with singular values in descending order, the product $QR^T = \widetilde{Q}\widetilde{R}^T$ does not change, i.e., the computation of Y_* is not affected. Therefore we can compute the usual SVD for an easier implementation and keep in mind that $\widetilde{S}^2 + \hat{S}^2 \neq I_n$.

It remains to show that $\Delta \in \mathrm{ID}_P$, so that it is actually the Riemannian logarithm. Since F is not in the cut locus Cut_P, we have $\operatorname{rank}(U^T Y) = p$, which means that the smallest singular value of $U^T Y$ is larger than zero (and smaller than or equal to one). Therefore the entries of $\Sigma = \arccos(S)$ are smaller than $\frac{\pi}{2}$, which shows the claim. □

The next theorem gives an explicit description of the shortest geodesics between a point and another point in its cut locus.

Theorem 2.5.5. For $P = UU^T \in \mathrm{Gr}(n,p)$ and some $F = YY^T \in \mathrm{Cut}_P$, let r denote the number of principal angles between P and F equal to $\frac{\pi}{2}$. Then $\Delta \in \mathrm{TCL}_P \subset T_P\mathrm{Gr}(n,p)$ is a minimizing solution of

$$\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta) = F \tag{2.5.7}$$

if and only if the horizontal lift Δ_U^{hor} is an output of Algorithm 2.5.3.

Consider the compact SVD $\Delta_U^{\mathsf{hor}} = \hat{Q}\Sigma R^T$. Then the horizontal lifts of all other minimizing solutions of (2.5.7) are given by

$$(\Delta_W)_U^{\mathsf{hor}} := \hat{Q}\Sigma\,\mathrm{diag}(W, I_{p-r})R^T,$$

where $W \in \mathrm{O}(r)$ and $\mathrm{diag}(W, I_{p-r}) = \left(\begin{smallmatrix} W & 0 \\ 0 & I_{p-r}\end{smallmatrix}\right)$ denotes a block diagonal matrix. The shortest geodesics between P and F are given by

$$\gamma_W(t) := \mathrm{Exp}_P^{\mathrm{Gr}}(t\Delta_W) = [UR\,\mathrm{diag}(W^T, I_{p-r})\cos(t\Sigma) + \hat{Q}\sin(t\Sigma)].$$

Proof. Algorithm 2.5.3 continues to work for points in the cut locus, but the result is not unique. With an SVD of Y^TU with singular values in ascending order, the first r singular values are zero. By Proposition 2.B.1,

$$Y^TU = QSR^T = Q\,\mathrm{diag}(W_1, D)S\,\mathrm{diag}(W_2, D^T)R^T,$$

where $D \in \mathrm{O}(p-r)$ with $(D)_{ij} = 0$ for $s_i \neq s_j$ and $W_1, W_2 \in \mathrm{O}(r)$ arbitrary. Then Y_* is not unique anymore, but is given as the set of matrices

$$\left\{Y_{*,W_1,W_2} := YQ\,\mathrm{diag}(W_1W_2, I_{p-r})R^T \mid W_1, W_2 \in O(r)\right\}.$$

Define $W := W_1W_2$ and $\hat{W} := \mathrm{diag}(W, I_{p-r})$. Then

$$\begin{aligned}(I_n - UU^T)Y_{*,W} &= (I_n - UU^T)YQ\hat{W}R^T \\ &= \underbrace{(I_n - UU^T)YQR^T}_{\hat{Q}\hat{S}R^T} R\hat{W}R^T = \hat{Q}\hat{S}\hat{W}R^T.\end{aligned}$$

With $\Sigma = \arcsin(\hat{S}) = \arccos(S)$, every matrix

$$(\Delta_W)_U^{\mathsf{hor}} := \hat{Q}\Sigma\hat{W}R^T$$

is the horizontal lift of a tangent vector at P of a geodesic towards F: For the exponential, it holds that

$$\begin{aligned}\mathrm{Exp}_P^{\mathrm{Gr}}(\Delta_W) &= [UR\hat{W}^T\cos(\Sigma)\hat{W}R^T + \hat{Q}\sin(\Sigma)\hat{W}R^T] \\ &= [UR\hat{W}^T\cos(\Sigma) + \hat{Q}\sin(\Sigma)] = [UR\cos(\Sigma) + \hat{Q}\sin(\Sigma)] = [Y],\end{aligned}$$

where the third equality holds, since $\Sigma = \operatorname{diag}(\frac{\pi}{2}, \dots, \frac{\pi}{2}, \sigma_{r+1}, \dots, \sigma_p)$. But the geodesics γ_W starting at $[U]$ in the directions Δ_W differ, i.e.

$$\begin{aligned}\gamma_W(t) &= [UR\hat{W}^T \cos(t\Sigma)\hat{W}R^T + \hat{Q}\sin(t\Sigma)\hat{W}R^T]\\ &= [UR\hat{W}^T \cos(t\Sigma) + \hat{Q}\sin(t\Sigma)].\end{aligned}$$

Hence, the ambiguity factor $\hat{W}^T = \operatorname{diag}(W^T, I_{p-r})$ does not vanish for $0 < t < 1$. The geodesics are all of the same (minimal) length, since the singular values do not change and $\gamma_W(1) = \operatorname{Exp}_P^{\mathrm{Gr}}(\Delta_W)$.

To show that there are no other solutions, let $\bar{\Delta} \in \mathrm{TCL}_P$ fulfill $\operatorname{Exp}_P^{\mathrm{Gr}}(\bar{\Delta}) = F$ and $\bar{\Delta}_U^{\mathsf{hor}} = \bar{Q}\bar{\Sigma}\bar{R}^T$. Then by (2.3.10) there is some $M \in \mathrm{O}(p)$ such that

$$YM = U\bar{R}\cos(\bar{\Sigma})\bar{R}^T + \bar{Q}\sin(\bar{\Sigma})\bar{R}^T,$$

which implies $U^TYM = \bar{R}\cos(\bar{\Sigma})\bar{R}^T$, which is an SVD of U^TYM. Therefore $\bar{Y}_* := YM\bar{R}\bar{R}^T = YM$ fulfills the properties of Y_* of Algorithm 2.5.3. Now

$$(I_n - UU^T)\bar{Y}_* = (I_n - UU^T)(U\bar{R}\cos(\bar{\Sigma})\bar{R}^T + \bar{Q}\sin(\bar{\Sigma})\bar{R}^T) = \bar{Q}\sin(\bar{\Sigma})\bar{R}^T,$$

which means that $\bar{Q}\sin(\bar{\Sigma})\bar{R}^T$ is a compact SVD of $(I_n - UU^T)\bar{Y}_*$. Therefore $\bar{\Delta}_U^{\mathsf{hor}}$ is an output of Algorithm 2.5.3 and the claim is shown. □

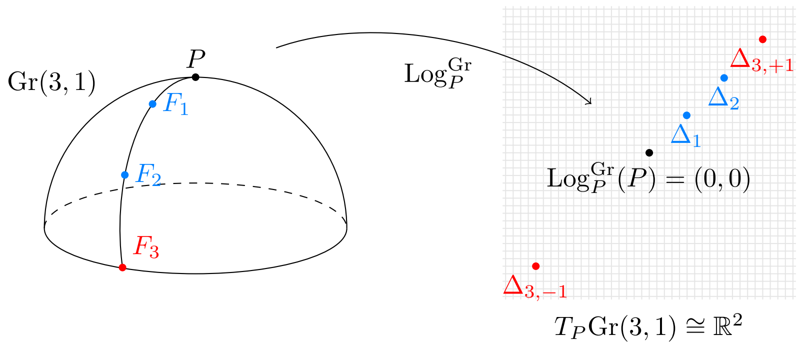

Figure 2.2: The manifold of one-dimensional subspaces of $\mathbb{R}^3$, i.e., $\mathrm{Gr}(3,1)$, can be seen as the upper half sphere with half of the equator removed. For points in the cut locus of a point $P \in \mathrm{Gr}(3,1)$ (like F_3 in the figure), there is no unique velocity vector in $T_P\mathrm{Gr}(3,1)$ that sends a geodesic from P to the point in question, but instead a set of two starting velocities ($\Delta_{3,+1}$ and $\Delta_{3,-1}$) that can be calculated according to Theorem 2.5.5. Since the points actually mark one dimensional subspaces through the origin, F_3 is identical to its antipode on the equator.

Together, Theorem 2.5.4 and Theorem 2.5.5 allow to map any set of points on $\mathrm{Gr}(n,p)$ to a single tangent space. The situation of multiple tangent vectors that correspond to one and the same point in the cut locus is visualized in Figure 2.2. Notice that if $r = 1$ in Theorem 2.5.5, there are only two possible geodesics $\gamma_{\pm 1}(t)$. For $r > 1$ there is a smooth variation of geodesics.

In [Bat+15, Theorem 3.3] a closed formula for the logarithm for Grassmann locations represented by orthogonal projectors was derived. We recast this result in form of the following proposition.

Proposition 2.5.6 (Grassmann Logarithm: Projector perspective)**.** Let a point $P \in \mathrm{Gr}(n,p)$ and $F \in \mathrm{Gr}(n,p) \setminus \mathrm{Cut}_P$. Then $\Delta = [\Omega, P] \in \mathrm{ID}_P \subset T_P\mathrm{Gr}(n,p)$ such that $\mathrm{Exp}_P^{\mathrm{Gr}}([\Omega, P]) = F$ is determined by

$$\Omega = \frac{1}{2} \log_m \left((I_n - 2F)(I_n - 2P) \right) \in \mathfrak{so}_P(n).$$

Consequently $\mathrm{Log}_P^{\mathrm{Gr}}(F) = [\Omega, P]$.

This proposition gives the logarithm explicitly, but it relies on $n \times n$ matrices. Lifting the problem to the Stiefel manifold reduces the computational complexity. A method to compute the logarithm that uses an orthogonal completion of the Stiefel representative U and the CS decomposition was proposed in [Gal+03].

2.5.3 Numerical Performance of the Logarithm

In this section, we assess the numerical accuracy of Algorithm 2.5.3 as opposed to the algorithm introduced in [AMS04, Section 3.8], for brevity hereafter referred to as the *new log algorithm* and the *standard log algorithm*, respectively.

For a random subspace representative $U \in \mathrm{St}(1000, 200)$ and a random horizontal tangent vector $\Delta_U^{\mathsf{hor}} \in \mathsf{Hor}_U \mathrm{St}(1000, 200)$ with largest singular value set to $\frac{\pi}{2}$, the subspace representative

$$U_1(\tau) = \mathrm{Exp}_U^{\mathrm{Gr}} \left((1-\tau)\Delta_U^{\mathsf{hor}} \right), \quad \tau \in [10^{-20}, 10^{0}],$$

is calculated. Observe that $U_1(0)U_1(0)^T$ is in the cut locus of UU^T. Then the logarithm $(\tilde{\Delta}(\tau))_U^{\mathsf{hor}} = \mathrm{Log}_U^{\mathsf{hor}}(U, U_1(\tau))$ is calculated according to the new log algorithm and the standard log algorithm, respectively. In the latter case, $(\tilde{\Delta}(\tau))_U^{\mathsf{hor}}$ is projected to the horizontal space $\mathsf{Hor}_U \mathrm{St}(1000, 200)$ by (2.3.4) to ensure $U^T(\tilde{\Delta}(\tau))_U^{\mathsf{hor}} = 0$. For $\tilde{U}_1(\tau) = \mathrm{Exp}_U^{\mathrm{Gr}}((\tilde{\Delta}(\tau))_U^{\mathsf{hor}})$, the error is then calculated according to (2.5.3) as

$$\mathrm{dist}\left(U_1(\tau), \tilde{U}_1(\tau)\right) = \|\arccos(S)\|_F,$$

where $QSR^T = U_1(\tau)^T\tilde{U}_1(\tau)$ is an SVD. Even though theoretically impossible, entries of values larger than one may arise in S due to the finite machine precision. In order to catch such numerical errors, the real part $\Re(\arccos(S))$ is used in the actual calculations of the subspace distance.

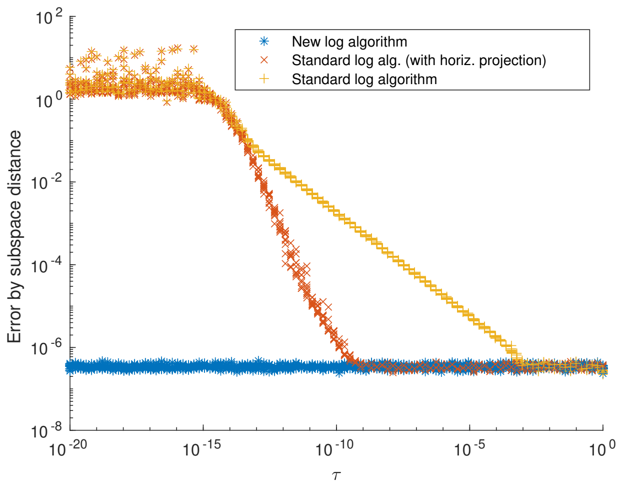

Figure 2.3: The error of the new log algorithm (blue stars) versus standard log algorithm with horizontal projection (red crosses) by subspace distance over τ. For comparison, the error of the standard log algorithm without projection onto the horizontal space is also displayed (yellow plus). The cut locus is approached by $\mathrm{Exp}_U^{\mathrm{Gr}}\left((1-\tau)\Delta_U^{\mathsf{hor}}\right)$. It can be observed that the new log algorithm still produces reliable results close to the cut locus.

In Figure 2.3, the subspace distance between $U_1(\tau)$ and $\tilde{U}_1(\tau)$ is displayed for 100 logarithmically spaced values τ between 10^{-20} and 10^0. The Stiefel representative $\tilde{U}_1(\tau)$ is here calculated with the new log algorithm, with the standard log algorithm, and with the standard log algorithm with projection onto the horizontal space. This is repeated for 10 random subspace representatives U with random horizontal tangent vectors Δ_U^{hor}, and the results are plotted individually. As expected, Algorithm 2.5.3 shows favorable behaviour when approaching the cut locus. When the result of the standard log algorithm is not projected onto the horizontal space, it can be seen that its subspace error starts to increase already at $\tau \approx 10^{-3}$. The baseline error (in Figure 2.3) is due to the numerical accuracy of the subspace distance calculation procedure.

Even though this experiment addresses the extreme-case behavior, it is of practical importance. In fact, the results of [AEK06] show that for large-scale n and two subspaces drawn from the uniform distribution on $\mathrm{Gr}(n,p)$, the largest principal angle between the subspaces is with high probability close to $\frac{\pi}{2}$.

2.6 Local Parameterizations of the Grassmann Manifold

In this section, we construct local parameterizations and coordinate charts of the Grassmannian. To this end, we work with the Grassmann representation as orthogonal projector $P = UU^T$. The dimension of $\mathrm{Gr}(n,p)$ is $(n-p)p$. Here, we recap how explicit local parameterizations from open subsets of $\mathbb{R}^{(n-p)\times p}$ onto open subsets of $\mathrm{Gr}(n,p)$ (and the corresponding coordinate charts) can be constructed.

The Grassmannian $\mathrm{Gr}(n,p)$ can be parameterized by the so called *normal coordinates* via the exponential map, which was also done in [HHT07]. Let $P = UU^T \in \mathrm{Gr}(n,p)$ and $U_\perp$ some orthogonal completion of $U \in \mathrm{St}(n,p)$. By making use of (2.2.11), a parameterization of $\mathrm{Gr}(n,p)$ around P is given via

$$\begin{aligned}\rho\colon \mathbb{R}^{(n-p)\times p} &\to \mathrm{Gr}(n,p),\\ \rho(B) &:= \mathrm{Exp}^{\mathrm{Gr}}_P(U_\perp B U^T + U B^T U_\perp^T)\\ &= \begin{pmatrix} U & U_\perp\end{pmatrix} \exp_{\mathrm{m}}\left(\begin{pmatrix} 0 & -B^T \\ B & 0\end{pmatrix}\right) P_0 \exp_{\mathrm{m}}\left(\begin{pmatrix} 0 & B^T \\ -B & 0\end{pmatrix}\right) \begin{pmatrix} U & U_\perp\end{pmatrix}^T.\end{aligned}$$

A different approach that avoids matrix exponentials, and which is also briefly introduced in [HM94, Appendix C.4], works as follows: Let $\mathcal{B} \subset \mathbb{R}^{(n-p)\times p}$ be an open ball around the zero-matrix $0 \in \mathbb{R}^{(n-p)\times p}$ for some induced matrix norm $\|\cdot\|$. Consider

$$\varphi\colon \mathcal{B} \to \mathbb{R}^{n\times n},\ B \mapsto \begin{pmatrix} I_p \\ B\end{pmatrix} (I_p + B^T B)^{-1} \begin{pmatrix} I_p & B^T\end{pmatrix}.$$

Note that B is mapped to the orthogonal projector onto $\mathrm{colspan}(\left(\begin{smallmatrix} I_p \\ B\end{smallmatrix}\right))$, so that actually $\varphi(\mathcal{B}) \subset \mathrm{Gr}(n,p)$. In particular, $\varphi(0) = P_0$. Let $P \in \mathrm{Gr}(n,p)$ be written block-wise as $P = \left(\begin{smallmatrix} A & B^T \\ B & C\end{smallmatrix}\right)$. Next, we show that the image of φ is the set of such projectors P with an invertible $p\times p$-block A and that φ is a bijection onto its image. To this end, assume that $A \in \mathbb{R}^{p\times p}$ has full rank p. Because P is idempotent, it holds

$$P = \begin{pmatrix} A & B^T \\ B & C\end{pmatrix} = \begin{pmatrix} A^2 + B^T B & AB^T + B^T C \\ BA + CB & BB^T + C^2\end{pmatrix} = P^2.$$

As a consequence, $(I_p + A^{-1}B^T B A^{-1})^{-1} = A(\overbrace{A^2 + B^T B}^{=A})^{-1} A = A$. Moreover, since $p = \operatorname{rank} P = \operatorname{rank} A = \operatorname{rank}(\left(\begin{smallmatrix} A \\ B\end{smallmatrix}\right))$, the blocks $\left(\begin{smallmatrix} B^T \\ C\end{smallmatrix}\right)$ can be expressed as a linear combination $\left(\begin{smallmatrix} A \\ B\end{smallmatrix}\right) X = \left(\begin{smallmatrix} B^T \\ C\end{smallmatrix}\right)$ with $X \in \mathbb{R}^{p\times(n-p)}$. This shows that $X = A^{-1}B^T$ and $C = BA^{-1}B^T$. In summary,

$$\begin{aligned} P &= \begin{pmatrix} A & B^T \\ B & BA^{-1}B^T\end{pmatrix} = \begin{pmatrix} I_p \\ BA^{-1}\end{pmatrix} A \begin{pmatrix} I_p & A^{-1}B^T\end{pmatrix}\\ &= \begin{pmatrix} I_p \\ BA^{-1}\end{pmatrix} (I_p + A^{-1}B^T B A^{-1})^{-1} \begin{pmatrix} I_p & A^{-1}B^T\end{pmatrix} = \varphi(BA^{-1}).\end{aligned}$$

Let $\psi : \left(\begin{smallmatrix} A & B^T \\ B & C\end{smallmatrix}\right) \mapsto BA^{-1}$. Then, for any $B \in \mathcal{B}$, $(\psi\circ\varphi)(B) = B$ so that $\psi\circ\varphi = \mathrm{id}|_{\mathcal{B}}$. Conversely, for any $P \in \mathrm{Gr}(n,p)$ with full rank upper $(p\times p)$-diagonal block A, $(\varphi\circ\psi)(P) = P$. Therefore, $\varphi : \mathcal{B} \to \varphi(\mathcal{B})$ is a local parameterization around $0 \in \mathbb{R}^{(n-p)\times p}$ and $x := \psi|_{\varphi(\mathcal{B})} : \varphi(\mathcal{B}) \to \mathcal{B}$ is the associated coordinate chart $x = \varphi^{-1}$. With the group action Φ, we can move this local parameterization to obtain local parameterizations around any other point of $\mathrm{Gr}(n,p)$ via $\varphi_Q(B) := Q\varphi(B)Q^T$, which (re)establishes the fact that $\mathrm{Gr}(n,p)$ is an embedded $(n-p)p$-dimensional submanifold of $\mathbb{R}^{n\times n}$.

The tangent space at P is the image $\mathrm{colspan}(\mathrm{d}\varphi_P)$ for a suitable parameterization φ around P. At P_0, we obtain

$$T_{P_0}\mathrm{Gr}(n,p) = \{\mathrm{d}\varphi_{P_0}(B) \mid B \in \mathbb{R}^{(n-p)\times p}\} = \{\begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix} \mid B \in \mathbb{R}^{(n-p)\times p}\},$$

in consistency with (2.2.8).

In principle, φ and ψ can be used as a replacement for the Riemannian exp- and log-mappings in data processing procedures. For example, for a set of data points contained in $\varphi(\mathcal{B}) \subset \mathrm{Gr}(n,p)$, Euclidean interpolation can be performed on the coordinate images in $\mathcal{B} \subset \mathbb{R}^{(n-p)\times p}$. Likewise, for an objective function $f : \mathrm{Gr}(n,p) \supset \mathcal{D} \to \mathbb{R}$ with domain $\mathcal{D} \subset \varphi(\mathcal{B})$, the associated function $f \circ \varphi : \mathbb{R}^{(n-p)\times p} \supset \varphi^{-1}(\mathcal{D}) \to \mathbb{R}$ can be optimized relying entirely on standard Euclidean tools; no evaluation of neither matrix exponentials nor matrix logarithms is required. Yet, these parameterizations do not enjoy the metric special properties of the Riemannian normal coordinates. Another reason to be wary of interpolation in coordinates is that the values on the Grassmannian will never leave $\varphi(\mathcal{B})$, and this can be very unnatural for some data sets. Furthermore, the presence of a domain $\mathcal{D}$ can be unnatural, as $\varphi(\mathcal{B})$ is an open subset of $\mathrm{Gr}(n,p)$, whereas the whole Grassmannian is compact, a desirable property for optimization. If charts are a switched, then information gathered by the solver may lose interest. Nevertheless, working in charts can be a successful approach [UM14].

2.7 Jacobi Fields and Conjugate Points

In this section, we describe Jacobi fields vanishing at one point and the conjugate locus of the Grassmannian. *Jacobi fields* are vector fields along a geodesic fulfilling the Jacobi equation (2.7.1). They can be viewed as vector fields pointing towards another "close-by" geodesic, see for example [Lee18]. The *conjugate points* of P are all those $F \in \mathrm{Gr}(n,p)$ such that there is a non-zero Jacobi field along a (not necessarily minimizing) geodesic from P to F, which vanishes at P and F. The set of all conjugate points of P is the *conjugate locus* of P. In general, there are not always multiple distinct (possibly non-minimizing) geodesics between two conjugate points, but on the Grassmannian there are. The conjugate locus on the Grassmannian was first treated in [Won68a], but the description there is not complete. This is for example pointed out in [Sak77] and [Ber97]. The latter gives a description of the conjugate locus in the complex case, which we show can be transferred to the real case.

Jacobi fields and conjugate points are of interest when variations of geodesics are considered. They arise for example in geodesic regression [Fle13] and curve fitting problems on manifolds [BG18].

2.7.1 Jacobi Fields

A *Jacobi field* is smooth vector field J along a geodesic γ satisfying the ordinary differential equation

$$D_t^2 J + R(J, \dot{\gamma})\dot{\gamma} = 0, \tag{2.7.1}$$

called *Jacobi equation*. Here $R(\cdot, \cdot)$ is the curvature tensor and D_t denotes the covariant derivative along the curve γ. This means that for every extension $\hat{J}$ of J, which is to be understood as a smooth vector field on a neighborhood of the image of γ that coincides with J on $\gamma(t)$ for every t, it holds that $(D_t J)(t) = \nabla_{\dot{\gamma}(t)} \hat{J}$. For a detailed introduction see for example [Lee18, Chapter 10]. A Jacobi field is the variation field of a variation through geodesics. That means intuitively that J points from the geodesic γ to a "close-by" geodesic, and, by linearity and scaling, to a whole family of such close-by geodesics. Jacobi fields that vanish at a point can be explicitly described via [Lee18, Proposition 10.10], which states that the Jacobi field J along the geodesic γ, with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$, and initial conditions $J(0) = 0 \in T_p M$ and $D_t J(0) = w \in T_v(T_p M) \cong T_p M$ is given by

$$J(t) = \mathrm{d}(\exp_p)_{tv}(tw). \tag{2.7.2}$$

The concept is visualized in Figure 2.4.

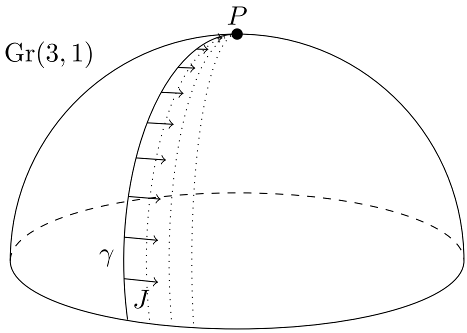

Figure 2.4: The Jacobi field J points from the geodesic γ towards close-by geodesics (dotted) and vanishes at P. Note that $J(t) \in T_{\gamma(t)}\mathrm{Gr}(3,1)$ is a tangent vector and not actually the offset vector between points on the respective geodesics in $\mathbb{R}^3$. Nevertheless, J is the variation field of a variation of γ through geodesics, c.f. [Lee18, Proposition 10.4].

By making use of the derivative of the exponential mapping derived in Proposition 2.3.4, we can state the following proposition for Jacobi fields vanishing at a point on the Grassmannian.

Proposition 2.7.1. Let $P = UU^T \in \mathrm{Gr}(n,p)$ and let $\Delta_1, \Delta_2 \in T_P\mathrm{Gr}(n,p)$ be two tangent vectors, where the singular values of $(\Delta_1)^{\mathsf{hor}}_U$ are mutually distinct and non-zero. Define the geodesic γ by $\gamma(t) := \mathrm{Exp}^{\mathrm{Gr}}_P(t\Delta_1)$. Furthermore, let $(t\Delta_1)^{\mathsf{hor}}_U = Q(t\Sigma)V^T$ and $(t(\Delta_1+s\Delta_2))^{\mathsf{hor}}_U = Q(s)(t\Sigma(s))V(s)^T$ be given via the compact SVDs of the horizontal lifts, i.e., $Q(s) \in \mathrm{St}(n,p)$, $\Sigma(s) = \mathrm{diag}(\sigma_1(s), \dots, \sigma_p(s))$ and $V(s) \in \mathrm{O}(p)$, as well as $Q(0) = Q, \Sigma(0) = \Sigma$ and $V(0) = V$. Finally, define

$$Y(t) := UV\cos(t\Sigma) + Q\sin(t\Sigma) \in \mathrm{St}(n,p)$$

and

$$\Gamma(t) := U\dot{V}\cos(t\Sigma) - tUV\sin(t\Sigma)\dot{\Sigma} + \dot{Q}\sin(t\Sigma) + tQ\cos(t\Sigma)\dot{\Sigma} \in T_{Y(t)}\mathrm{St}(n,p).$$ [3]

Then the Jacobi field J along γ fulfilling $J(0) = 0$ and $D_tJ(0) = \Delta_2$ is given by

$$J(t) = \Gamma(t)Y(t)^T + Y(t)\Gamma(t)^T \in T_{\gamma(t)}\mathrm{Gr}(n,p).$$

The horizontal lift of $J(t)$ to $Y(t)$ is accordingly given by

$$(J(t))^{\mathsf{hor}}_{Y(t)} = \Gamma(t) + Y(t)\Gamma(t)^TY(t) = (I_n - Y(t)Y(t)^T)\Gamma(t).$$

It is the variation field of the variation of γ through geodesics given by $\Xi(s,t) := \mathrm{Exp}^{\mathrm{Gr}}_P(t(\Delta_1 + s\Delta_2))$.

Proof. The proof works analogously to the one of Proposition 2.3.4, since according to (2.7.2)

$$J(t) = \mathrm{d}(\mathrm{Exp}^{\mathrm{Gr}}_P)_{t\Delta_1}(t\Delta_2) = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \mathrm{Exp}^{\mathrm{Gr}}_P(t(\Delta_1 + s\Delta_2)).$$

□

2.7.2 Conjugate Locus

In the following, let $p \leq \frac{n}{2}$. The reason for this restriction is that for $p > \frac{n}{2}$ there are automatically principal angles equal to zero, yet these do not contribute to the conjugate locus, as one can see by switching to the orthogonal complement. We will see that the conjugate locus Conj_P of $P \in \mathrm{Gr}(n,p)$ is given by all $F \in \mathrm{Gr}(n,p)$ such that at least two principal angles between P and F coincide, or there is at least one principal angle equal to zero if $p < \frac{n}{2}$. This obviously includes the case of two or more principal angles equal to $\frac{\pi}{2}$. In the complex case, the conjugate locus also includes points with one principal angle of $\frac{\pi}{2}$, as is shown in [Ber97]. Only in the cases of principal angles of $\frac{\pi}{2}$ is there a nontrivial

[3] As in the case of the derivative of the Grassmann exponential, the matrices $\dot{Q} = \frac{\mathrm{d}Q}{\mathrm{d}s}(0), \dot{\Sigma} = \frac{\mathrm{d}\Sigma}{\mathrm{d}s}(0)$ and $\dot{V} = \frac{\mathrm{d}V}{\mathrm{d}s}(0)$ can be calculated by Algorithm 2.B.2.

Jacobi field vanishing at P and F along a *shortest* geodesic. It can be calculated from the variation of geodesics as above. In the other cases, the shortest geodesic is unique, but we can smoothly vary longer geodesics from P to F. This variation is possible because of the periodicity of sine and cosine and the indeterminacies of the SVD.

Theorem 2.7.2. Let $P \in \mathrm{Gr}(n,p)$ where $p \leq \frac{n}{2}$. The conjugate locus Conj_P of P consists of all points $F \in \mathrm{Gr}(n,p)$ with at least two identical principal angles or, when $p < \frac{\pi}{2}$, at least one zero principal angle between F and P.

Proof. Let P and F have $r = j - i + 1$ repeated principal angles $\sigma_i = \cdots = \sigma_j$. Obtain $\Delta_U^{\mathrm{hor}} = \hat{Q}\Sigma R^T$ by Algorithm 2.5.3. Define Σ' by adding π to one of the repeated angles. Then for every $D \in \mathrm{O}(r)$ and $\tilde{D} := \mathrm{diag}(I_{i-1}, D, I_{p-j})$, the curve

$$\gamma_D(t) = [UR\tilde{D}\cos(t\Sigma')\tilde{D}^T R^T + \hat{Q}\tilde{D}\sin(t\Sigma')\tilde{D}^T R^T]$$

is a geodesic from P to $\gamma_D(1) = F$. Since for $0 < t < 1$ the matrix $\cos(t\Sigma')$ does not have the same number of repeated diagonal entries as $\cos(t\Sigma)$, not all curves γ_D coincide. Then we can choose an open interval $\mathcal{I}$ around 0 and a smooth curve $D\colon \mathcal{I} \to \mathrm{O}(r)$ with $D(0) = I_r$ such that $\Gamma(s,t) = \gamma_{D(s)}(t)$ is a variation through geodesics as defined in [Lee18, Chap. 10, p. 284]. The variation field J of Γ is a Jacobi field along $\gamma_{D(0)}$ according to [Lee18, Theorem 10.1]. Furthermore, J is vanishing at $t = 0$ and $t = 1$, as $\gamma_{D(s)}(1) = \gamma_{D(\tilde{s})}(1)$ for all $s, \tilde{s} \in \mathcal{I}$ by Proposition 2.B.1, and likewise for $t = 0$. Since J is not constantly vanishing, P and F are conjugate along $\gamma_{D(0)}$ by definition.

When $p < \frac{n}{2}$ and there is at least one principal angle equal to zero, there is some additional freedom of variation. Let the last r principal angles between P and F be $\sigma_{p-r+1} = \cdots = \sigma_p = 0$. Obtain $\Delta_U^{\mathrm{hor}} = \hat{Q}\Sigma R^T$ by Algorithm 2.5.3. Since $p < \frac{n}{2}$, $\hat{Q}$ can be chosen such that $U^T\hat{Q} = 0$, and there is at least one unit vector $\hat{q} \in \mathbb{R}^n$, such that $\hat{q}$ is orthogonal to all column vectors in U and in $\hat{Q}$. Let $\hat{Q}_\perp$ be an orthogonal completion of $\hat{Q}$ with $\hat{q}$ as its first column vector. Define Σ' as the matrix Σ with π added to the $(p-r+1)$th diagonal entry. Then for every $W \in \mathrm{O}(2)$,

$$\gamma_W(t) = [UR\cos(t\Sigma') + \begin{pmatrix}\hat{Q} & \hat{Q}_\perp\end{pmatrix} \mathrm{diag}(I_{p-r}, W, I_{n-p+r-2}) \begin{pmatrix}\sin(t\Sigma') \\ 0\end{pmatrix}]$$

is a geodesic from P with $\gamma_W(1) = F$. With an argument as above, P and F are conjugate along γ_{I_2}.

There are no other points in the conjugate locus than those with repeated principal angles (or one zero angle in case of $p < \frac{n}{2}$), as the SVD is unique (up to order of the singular values) for matrices with no repeating and no zero singular values. As every geodesic on the Grassmannian is of the form (2.3.9), the claim can be shown by contradiction. □

By construction, the length of $\gamma_{D(0)}$ between P and F is longer than the length of the shortest geodesic, since $\|\Sigma\|_F < \|\Sigma'\|_F$. The same is true for the case of a zero angle.

It holds that the cut locus Cut_P is no subset of the conjugate locus Conj_P, since points with just one principal angle equal to $\frac{\pi}{2}$ are not in the conjugate locus. Likewise the conjugate locus is no subset of the cut locus. The points in the conjugate locus that are conjugate along a *minimizing* geodesic however are also in the cut locus, as those are exactly those with multiple principal angles equal to $\frac{\pi}{2}$.

Remark. The (incomplete) treatment in [Won68a] covered only the cases of at least two principal angles equal to $\frac{\pi}{2}$ or principal angles equal to zero, but not the cases of repeated arbitrary principal angles. We can nevertheless take from there that for $p > \frac{n}{2}$ we need at least $2p - n + 1$ principal angles equal to zero, instead of just one as for $p < \frac{n}{2}$. Points with repeated (nonzero) principal angles are however always in the conjugate locus, as the proof of Theorem 2.7.2 still holds for them.

2.8 Conclusion

In this work, we have collected the facts and formulae that we deem most important for Riemannian computations on the Grassmann manifold. This includes in particular explicit formulae and algorithms for computing local coordinates, the Riemannian normal coordinates (the Grassmann exponential and logarithm mappings), the Riemannian connection, the parallel transport of tangent vectors and the sectional curvature. All these concepts may appear as building blocks or tools for the theoretical analysis of, e.g., optimization problems, interpolation problems and, more generally speaking, data processing problems such as data averaging or clustering.

We have treated the Grassmannian both as a quotient manifold of the orthogonal group and the Stiefel manifold, and as the space of orthogonal projectors of fixed rank and have exposed (and exploited) the connections between these view points. While concepts from differential geometry arise naturally in the theoretical considerations, care has been taken that the final formulae are purely matrix-based and thus are fit for immediate use in algorithms. At last, the paper features an original approach to computing the Grassmann logarithm, which simplifies the theoretical analysis, extends its operational domain and features improved numerical properties. Eventually, this tool allowed us to conduct a detailed investigation of shortest curves to cut points as well as studying the conjugate points on the Grassmannian by basic matrix-algebraic means. These findings are more explicit and more complete than the previous results in the research literature.

2.A Basics from Riemannian Geometry

For the reader's convenience, we recap some fundamentals from Riemannian geometry. Concise introductions can be found in [HM94, Appendices C.3, C.4, C.5], [Gal11] and [AMS08]. For an in-depth treatment, see for example [Car92; KN96; Lee18].

An *n-dimensional differentiable manifold* $\mathcal{M}$ is a topological space $\mathcal{M}$ such that for every point $p \in \mathcal{M}$, there exists a so-called *coordinate chart* $x : \mathcal{M} \supset \mathcal{D}_p \to \mathbb{R}^n$ that bijectively maps an open neighborhood $\mathcal{D}_p \subset \mathcal{M}$ of a location p to an open neighborhood $D_{x(p)} \subset \mathbb{R}^n$ around $x(p) \in \mathbb{R}^n$ with the additional property that the *coordinate change*

$$x \circ \tilde{x}^{-1} : \tilde{x}(\mathcal{D}_p \cap \tilde{\mathcal{D}}_p) \to x(\mathcal{D}_p \cap \tilde{\mathcal{D}}_p)$$

of two such charts $x, \tilde{x}$ is a diffeomorphism, where their domains of definition overlap, see [Gal11, Fig. 18.2, p. 496]. This enables to transfer the most essential tools from calculus to manifolds. An *n-dimensional submanifold of* $\mathbb{R}^{n+d}$ is a subset $\mathcal{M} \subset \mathbb{R}^{n+d}$ that can be locally smoothly straightened, i.e. satisfies the local n-slice condition [Lee12, Thm. 5.8].

Theorem 2.A.1 ([Gal11, Prop. 18.7, p. 500])**.** Let $h : \mathbb{R}^{n+d} \supset \Omega \to \mathbb{R}^d$ be differentiable and $c_0 \in \mathbb{R}^d$ be defined such that the differential $Dh_p \in \mathbb{R}^{d\times(n+d)}$ has maximum possible rank d at every point $p \in \Omega$ with $h(p) = c_0$. Then, the preimage

$$h^{-1}(c_0) = \{p \in \Omega \mid \ h(p) = c_0\}$$

is an n-dimensional submanifold of $\mathbb{R}^{n+d}$.

This theorem establishes the Stiefel manifold $\mathrm{St}(n,p) = \left\{U \in \mathbb{R}^{n\times p} \;\middle|\; U^T U = I\right\}$ as an embedded submanifold of $\mathbb{R}^{n\times p}$, since $\mathrm{St}(n,p) = F^{-1}(I)$ for $F : U \mapsto U^T U$.

Tangent Spaces The *tangent space* of a submanifold $\mathcal{M}$ at a point $p \in \mathcal{M}$, in symbols $T_p\mathcal{M}$, is the space of velocity vectors of differentiable curves $c : t \mapsto c(t)$ passing through p, i.e.,

$$T_p\mathcal{M} = \{\dot{c}(t_0) \mid \ c : I \to \mathcal{M},\ c(t_0) = p\}.$$

The tangent space is a vector space of the same dimension n as the manifold $\mathcal{M}$.

Geodesics and the Riemannian Distance Function Riemannian metrics measure the lengths and angles between tangent vectors. Eventually, this allows to measure the lengths of curves on a manifold and the Riemannian distance between two manifold locations.

A *Riemannian metric* on $\mathcal{M}$ is a family $(\langle\cdot,\cdot\rangle_p)_{p\in\mathcal{M}}$ of inner products $\langle\cdot,\cdot\rangle_p : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}$ that is smooth in variations of the base point p. The *length* of a tangent vector $v \in T_p\mathcal{M}$ is $\|v\|_p := \sqrt{\langle v, v\rangle_p}$. The length of a curve $c : [a,b] \to \mathcal{M}$ is defined as

$$L(c) = \int_a^b \|\dot{c}(t)\|_{c(t)} dt = \int_a^b \sqrt{\langle \dot{c}(t), \dot{c}(t)\rangle_{c(t)}} dt.$$

A curve is said to be *parameterized by the arc length*, if $L(c|_{[a,t]}) = t - a$ for all $t \in [a,b]$. Obviously, *unit-speed curves* with $\|\dot{c}(t)\|_{c(t)} \equiv 1$ are parameterized by the arc length. Constant-speed curves with $\|\dot{c}(t)\|_{c(t)} \equiv \nu_0$ are parameterized proportional to the arc length. The *Riemannian distance* between two points $p, q \in \mathcal{M}$ with respect to a given metric is

$$\operatorname{dist}_{\mathcal{M}}(p,q) = \inf\{L(c) \mid c : [a,b] \to \mathcal{M} \text{ piecewise smooth}, c(a) = p, c(b) = q\}, \quad (2.A.1)$$

where, by convention, $\inf\{\emptyset\} = \infty$. A shortest path between $p, q \in \mathcal{M}$ is a curve c that connects p and q such that $L(c) = \operatorname{dist}_{\mathcal{M}}(p,q)$. Candidates for shortest curves between points are called *geodesics* and are characterized by a differential equation: A differentiable curve $c : [a,b] \to \mathcal{M}$ is a geodesic (w.r.t. to a given Riemannian metric), if the *covariant derivative* of its velocity vector field vanishes, i.e.,

$$\frac{D\dot{c}}{dt}(t) = 0 \quad \forall t \in [a,b]. \quad (2.A.2)$$

Intuitively, the covariant derivative can be thought of as the standard derivative (if it exists) followed by a point-wise projection onto the tangent space. In general, a covariant derivative, also known as a *linear connection*, is a bilinear mapping $(X,Y) \mapsto \nabla_X Y$ that maps two vector fields X, Y to a third vector field $\nabla_X Y$ in such a way that it can be interpreted as the directional derivative of Y in the direction of X, [Lee18, §4, §5]. Of importance is the *Riemannian connection* or *Levi-Civita connection* that is compatible with a Riemannian metric [AMS08, Thm 5.3.1], [Lee18, Thm 5.10]. It is determined uniquely by the Koszul formula

$$\begin{aligned} 2\langle \nabla_X Y, Z\rangle &= X(\langle Y, Z\rangle) + Y(\langle Z, X\rangle) - Z(\langle X, Y\rangle) \\ &\quad -\langle X, [Y,Z]\rangle - \langle Y, [X,Z]\rangle + \langle Z, [X,Y]\rangle \end{aligned}$$

and is used to define the *Riemannian curvature tensor*

$$(X,Y,Z) \mapsto R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$ [4]

A Riemannian manifold is flat if and only if it is locally isometric to the Euclidean space, which holds if and only if the Riemannian curvature tensor vanishes identically [Lee18, Thm. 7.10].

Lie Groups and Orbits A *Lie group* is a smooth manifold that is also a group with smooth multiplication and inversion. A *matrix Lie group* G is a subgroup of the general linear group $GL(n,\mathbb{C})$ that is closed in $GL(n,\mathbb{C})$ (but not necessarily in the ambient space $\mathbb{C}^{n\times n}$). Basic examples include $GL(n,\mathbb{R})$ and the orthogonal group $\mathrm{O}(n)$. Any matrix Lie group G is automatically an embedded submanifold of $\mathbb{C}^{n\times n}$ [Hal15, Corollary 3.45]. The tangent space $T_I G$ of G at the identity $I \in G$ has a special role. When endowed with the bracket operator or *matrix commutator* $[V,W] = VW - WV$ for $V, W \in T_I G$, the tangent space becomes an algebra, called the *Lie algebra* associated with the Lie group

[4] In these formulae, $[X,Y] = X(Y) - Y(X)$ is the Lie bracket of two vector fields.

G, see [Hal15, §3]. As such, it is denoted by $\mathfrak{g} = T_I G$. For any $A \in G$, the function "left-multiplication with A" is a diffeomorphism $L_A\colon G \to G,\ L_A(B) = AB$; its differential at a point $B \in G$ is the isomorphism $\mathrm{d}(L_A)_B\colon T_B G \to T_{L_A(B)} G,\ \mathrm{d}(L_A)_B(V) = AV$. Using this observation at $B = I$ shows that the tangent space at an arbitrary location $A \in G$ is given by the translates (by left-multiplication) of the tangent space at the identity [GR17, §5.6, p. 160],

$$T_A G = T_{L_A(I)} G = A\mathfrak{g} = \left\{ \Delta = AV \in \mathbb{R}^{n\times n} \middle|\ V \in \mathfrak{g} \right\}. \tag{2.A.3}$$

A *smooth left action* of a Lie group G on a manifold M is a smooth map $\phi\colon G \times M \to M$ fulfilling $\phi(g_1, \phi(g_2, p)) = \phi(g_1 g_2, p)$ and $\phi(e, p) = p$ for all $g_1, g_2 \in G$ and all $p \in M$, where $e \in G$ denotes the identity element. One often writes $\phi(g, p) = g \cdot p$. For each $p \in M$, the *orbit of* p is defined as

$$G \cdot p := \{g \cdot p \mid g \in G\}, \tag{2.A.4}$$

and the *stabilizer of* p is defined as

$$G_p := \{g \in G \mid g \cdot p = p\}. \tag{2.A.5}$$

For a detailed introduction see for example [Lee12]. We need the following well known result, see for example [HM94, Section 2.1], where the quotient manifold G/G_p refers to the set $\{gG_p \mid g \in G\}$ endowed with the unique manifold structure that turns the quotient map $g \mapsto gG_p$ into a submersion.

Proposition 2.A.2. Let G be a compact Lie group acting smoothly on a manifold M. Then for any $p \in M$, the orbit $G \cdot p$ is an embedded submanifold of M that is diffeomorphic to the quotient manifold G/G_p.

Proof. The continuous action of a compact Lie group is always proper, [Lee12, Corollary 21.6]. Therefore [AB15, Proposition 3.41] shows the claim. □

2.B Matrix Analysis Necessities

Throughout, we consider the matrix space $\mathbb{R}^{m\times n}$ as a Euclidean vector space with the standard metric

$$\langle A, B\rangle_0 = \operatorname{tr}(A^T B). \tag{2.B.1}$$

Unless noted otherwise, the singular value decomposition (SVD) of a matrix $X \in \mathbb{R}^{m\times n}$ is understood to be the compact SVD

$$X = U\Sigma V^T, \quad U \in \mathbb{R}^{m\times n}, \Sigma, V \in \mathbb{R}^{n\times n}.$$

The SVD is not unique.

Proposition 2.B.1 (Ambiguity of the Singular Value Decomposition). [HJ91, Theorem 3.1.1'] Let $X \in \mathbb{R}^{m\times n}$ have a (full) SVD $X = U\Sigma V^T$ with singular values in descending order and $\operatorname{rank}(X) = r$. Let $\sigma_1 > \cdots > \sigma_k > 0$ be the distinct nonzero singular values with respective multiplicity $\mu_1, \dots, \mu_k$. Then $X = \tilde{U}\Sigma\tilde{V}^T$ is another SVD if and only if $\tilde{U} = U \operatorname{diag}(D_1, \dots, D_k, W_1)$ and $\tilde{V} = V \operatorname{diag}(D_1, \dots, D_k, W_2)$, with $D_i \in \mathrm{O}(\mu_i)$, $W_1 \in \mathrm{O}(m-r)$, and $W_2 \in \mathrm{O}(n-r)$ arbitrary.

Differentiating the Singular Value Decomposition Let $p \leq n \in \mathbb{N}$ and suppose that $t \mapsto Y(t) \in \mathbb{R}^{n\times p}$ is a differentiable matrix curve around $t_0 \in \mathbb{R}$. If the singular values of $Y(t_0)$ are mutually distinct and non-zero, then the singular values and both the left and the right singular vectors depend differentiable on $t \in [t_0 - \delta t, t_0 + \delta t]$ for δt small enough.

Let $t \mapsto Y(t) = U(t)\Sigma(t)V(t)^T \in \mathbb{R}^{n\times p}$, where $U(t) \in \mathrm{St}(n,p)$, $V(t) \in O(p)$ and $\Sigma(t) \in \mathbb{R}^{p\times p}$ diagonal and positive definite. Let u_j and v_j, $j = 1, \dots, p$ denote the columns of $U(t_0)$ and $V(t_0)$, respectively. For brevity, write $Y = Y(t_0), \dot{Y} = \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=t_0} Y(t)$, likewise for the other matrices that feature in the SVD. The derivatives of the matrix factors of the SVD can be calculated with Alg. 2.B.2. A proof can for example be found in [HBP09; DE99].

Algorithm 2.B.2 Differentiating the SVD

Input: Matrices $Y, \dot{Y} \in \mathbb{R}^{n\times p}$, (compact) SVD $Y = U\Sigma V^T$.
1: $\dot{\sigma}_j = (u_j)^T \dot{Y} v_j$ for $j = 1, \dots, p$
2: $\dot{V} = V\Gamma$, where $\Gamma_{ij} = \begin{cases} \frac{\sigma_i(u_i^T\dot{Y}v_j)+\sigma_j(u_j^T\dot{Y}v_i)}{(\sigma_j+\sigma_i)(\sigma_j-\sigma_i)}, & i \neq j \\ 0, & i = j \end{cases}$ for $i, j = 1, \dots, p$
3: $\dot{U} = \left(\dot{Y}V + U(\Sigma\Gamma - \dot{\Sigma})\right)\Sigma^{-1}$.
Output: $\dot{U}, \dot{\Sigma} = \operatorname{diag}(\dot{\sigma}_1, \dots, \dot{\sigma}_m), \dot{V}$

Differentiating the QR-Decomposition Let $t \mapsto Y(t) \in \mathbb{R}^{n\times r}$ be a differentiable matrix function with Taylor expansion $Y(t_0 + h) = Y(t_0) + h\dot{Y}(t_0) + \mathcal{O}(h^2)$. Following [WLL12, Proposition 2.2], the QR-decomposition is characterized via the following set of matrix equations.

$$Y(t) = Q(t)R(t), \quad Q^T(t)Q(t) = I_r, \quad 0 = P_L \odot R(t).$$

In the latter, $P_L = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 \end{pmatrix}$ and '$\odot$' is the element-wise matrix product so that $P_L \odot R$ selects the strictly lower triangle of the square matrix R. For brevity, we write

$Y = Y(t_0)$, $\dot{Y} = \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=t_0} Y(t)$, likewise for $Q(t)$, $R(t)$. By the product rule

$$\dot{Y} = \dot{Q}R + Q\dot{R}, \quad 0 = \dot{Q}^T Q + Q^T \dot{Q}, \quad 0 = P_L \odot \dot{R}.$$

According to [WLL12, Proposition 2.2], the derivatives $\dot{Q}, \dot{R}$ can be obtained from Alg. 2.B.3. The trick is to compute $X = Q^T \dot{Q}$ first and then use this to compute $\dot{Q} = QQ^T\dot{Q} + (I_n - QQ^T)\dot{Q}$ by exploiting that $Q^T\dot{Q}$ is skew-symmetric and that $\dot{R}R^{-1}$ is upper triangular.

Algorithm 2.B.3 Differentiating the QR-decomposition, [WLL12, Proposition 2.2]

Input: matrices $T, \dot{T} \in \mathbb{R}^{n\times r}$, (compact) QR-decomposition $T = QR$.
1: $L := P_L \odot (Q^T \dot{T} R^{-1})$
2: $X = L - L^T$ ▷ Now, $X = Q^T\dot{Q}$
3: $\dot{R} = Q^T \dot{T} - XR$
4: $\dot{Q} = (I_n - QQ^T)\dot{T}R^{-1} + QX$
Output: $\dot{Q}, \dot{R}$

Matrix Exponential and the Principal Matrix Logarithm The matrix exponential and the principal matrix logarithm are defined by

$$\exp_m(X) := \sum_{j=0}^{\infty} \frac{X^j}{j!}, \quad \log_m(I + X) := \sum_{j=1}^{\infty} (-1)^{j+1} \frac{X^j}{j}. \tag{2.B.2}$$

The latter is well-defined for matrices that have no eigenvalues on $\mathbb{R}^-$.

References

[AB15] M. M. Alexandrino and R. G. Bettiol. *Lie Groups and Geometric Aspects of Isometric Actions*. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-16613-1. DOI: 10.1007/978-3-319-16613-1_2.

[AEK06] P.-A. Absil, A. Edelman, and P. Koev. "On the largest principal angle between random subspaces". In: *Linear Algebra and its Applications* 414.1 (2006), pp. 288–294. ISSN: 0024-3795. DOI: 10.1016/j.laa.2005.10.004.

[AF08] D. Amsallem and C. Farhat. "Interpolation Method for Adapting Reduced-Order Models and Application to Aeroelasticity". In: *AIAA Journal* 46.7 (2008), pp. 1803–1813. DOI: 10.2514/1.35374.

[AMS04] P.-A. Absil, R. Mahony, and R. Sepulchre. "Riemannian Geometry of Grassmann Manifolds with a View on Algorithmic Computation". English. In: *Acta Applicandae Mathematica* 80.2 (2004), pp. 199–220. ISSN: 0167-8019. DOI: 10.1023/B:ACAP.0000013855.14971.91.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, New Jersey: Princeton University Press, 2008. URL: http://press.princeton.edu/titles/8586.html.

[ATV13] B. Afsari, R. Tron, and R. Vidal. "On the Convergence of Gradient Descent for Finding the Riemannian Center of Mass". In: *SIAM Journal on Control and Optimization* 51.3 (2013), pp. 2230–2260. DOI: 10.1137/12086282X.

[BA15] N. Boumal and P.-A. Absil. "Low-rank matrix completion via preconditioned optimization on the Grassmann manifold". In: *Linear Algebra and its Applications* 475 (2015), pp. 200–239. DOI: 10.1016/j.laa.2015.02.027.

[Bat+15] E. Batzies et al. "Geometric mean and geodesic regression on Grassmannians". In: *Linear Algebra Appl.* 466 (2015), pp. 83–101. ISSN: 0024-3795. DOI: 10.1016/j.laa.2014.10.003.

[Ber97] S. Berceanu. "On the geometry of complex Grassmann manifold, its noncompact dual and coherent states". In: *Bull. Belg. Math. Soc. Simon Stevin* 4.2 (1997), pp. 205–243. ISSN: 1370-1444. DOI: 10.36045/bbms/1105731655.

[BG18] R. Bergmann and P.-Y. Gousenbourger. "A Variational Model for Data Fitting on Manifolds by Minimizing the Acceleration of a Bézier Curve". In: *Frontiers in Applied Mathematics and Statistics* 4 (2018), p. 59. ISSN: 2297-4687. DOI: 10.3389/fams.2018.00059.

[BN91] A. A. Borisenko and Y. A. Nikolaevskiĭ. "Grassmann manifolds and Grassmann image of submanifolds". In: *Uspekhi Mat. Nauk* 46.2(278) (1991), pp. 41–83, 240. ISSN: 0042-1316. DOI: 10.1070/RM1991v046n02ABEH002742.

[Bou14] N. Boumal. "Optimization and estimation on manifolds". PhD thesis. Université catholique de Louvain, 2014.

[BW15] L. Balzano and S. J. Wright. "Local Convergence of an Algorithm for Subspace Identification from Partial Data". In: *Foundations of Computational Mathematics* 15.5 (2015), pp. 1279–1314. DOI: 10.1007/s10208-014-9227-7.

[Car92] M. P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, 1992. ISBN: 9780817634902.

[CV19] R. Chakraborty and B. C. Vemuri. "Statistics on the Stiefel manifold: Theory and applications". In: *The Annals of Statistics* 47.1 (2019), pp. 415–438. DOI: 10.1214/18-AOS1692.

[DE99] L. Dieci and T. Eirola. "On Smooth Decompositions of Matrices". In: *SIAM Journal on Matrix Analysis and Applications* 20.3 (1999), pp. 800–819. DOI: 10.1137/S0895479897330182.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith. “The Geometry of Algorithms with Orthogonality Constraints”. In: *SIAM Journal on Matrix Analysis and Applications* 20.2 (Apr. 1998), pp. 303–353. ISSN: 0895-4798. DOI: 10.1137/S0895479895290954.

[Fle13] P. T. Fletcher. “Geodesic regression and the theory of least squares on Riemannian manifolds”. In: *International journal of computer vision* 105.2 (2013), pp. 171–185. DOI: 10.1007/s11263-012-0591-y.

[Gal+03] K. A. Gallivan et al. “Efficient algorithms for inferences on Grassmann manifolds”. In: *IEEE Workshop on Statistical Signal Processing*. 2003, pp. 315–318. DOI: 10.1109/SSP.2003.1289408.

[Gal11] J. H. Gallier. *Geometric Methods and Applications: For Computer Science and Engineering*. Texts in Applied Mathematics. New York: Springer, 2011. ISBN: 0-387-95044-3. DOI: 10.1007/978-1-4419-9961-0.

[GMA19] P.-Y. Gousenbourger, E. Massart, and P.-A. Absil. “Data fitting on manifolds with composite Bézier-like curves and blended cubic splines”. In: *Journal of Mathematical Imaging and Vision* 61.5 (2019), pp. 645–671.

[GR17] R. Godement and U. Ray. *Introduction to the Theory of Lie Groups*. Universitext. Springer International Publishing, 2017. ISBN: 9783319543734. DOI: 10.1007/978-3-319-54375-8.

[Hal15] B. C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. 2nd. Springer Graduate texts in Mathematics. New York – Berlin – Heidelberg: Springer–Verlag, 2015. DOI: 10.1007/978-3-319-13467-3.

[HBP09] A. Hay, J. T. Borggaard, and D. Pelletier. “Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition”. In: *Journal of Fluid Mechanics* 629 (2009), pp. 41–72. DOI: 10.1017/S0022112009006363.

[Hel01] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. Crm Proceedings & Lecture Notes. American Mathematical Society, 2001. ISBN: 9780821828489.

[HHT07] U. Helmke, K. Hüper, and J. Trumpf. *Newton’s method on Graßmann manifolds*. 2007. eprint: 0709.2205 (math.OC).

[Hig08] N. J. Higham. *Functions of Matrices: Theory and Computation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2008, pp. xx+425. ISBN: 978-0-898716-46-7. DOI: 10.1137/1.9780898717778.

[HJ91] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991, pp. viii+607. ISBN: 0-521-30587-X. DOI: 10.1017/CBO9780511840371.

[HLW06] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*. Second. Vol. 31. Springer Series in Computational Mathematics. Berlin: Springer-Verlag, 2006, pp. xviii+644. ISBN: 3-540-30663-3; 978-3-540-30663-4. DOI: 10.1007/3-540-30666-8.

[HM94] U. Helmke and J. B. Moore. *Optimization and Dynamical Systems*. Communications & Control Engineering. London: Springer–Verlag, 1994. ISBN: 978-1-4471-3469-5. DOI: 10.1007/978-1-4471-3467-1.

[KL07] O. Koch and C. Lubich. “Dynamical Low-Rank Approximation”. In: *SIAM J. Matrix Analysis Applications* 29 (2007), pp. 434–454. DOI: 10.1137/050639703.

[KN96] S. Kobayashi and K. Nomizu. *Foundations of differential geometry. Vol. I & II*. Wiley Classics Library. Reprint of the 1963 original. John Wiley & Sons, Inc., New York, 1996.

[Koz01] E. S. Kozlov. “Geometry of Real Grassmann Manifolds. Part I - VI”. In: *Journal of Mathematical Sciences* (2000 – 2001).

[Lee12] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York, 2012. ISBN: 9781441999825. DOI: 10.1007/978-1-4419-9982-5.

[Lee18] J. M. Lee. *Introduction to Riemannian Manifolds*. Vol. 176. Graduate Texts in Mathematics. Springer, Cham, 2018, pp. xiii+437. ISBN: 978-3-319-91754-2; 978-3-319-91755-9. DOI: 10.1007/978-3-319-91755-9.

[Lei61] K. Leichtweiss. “Zur Riemannschen Geometrie in Grassmannschen Mannigfaltigkeiten”. In: *Math. Z.* 76 (1961), pp. 334–366. ISSN: 0025-5874. DOI: 10.1007/BF01210982.

[Li+19] C. Li et al. *Convergence Analysis of Gradient Algorithms on Riemannian Manifolds Without Curvature Constraints and Application to Riemannian Mass*. 2019. eprint: 1910.02280 (math.OC).

[LLY20] Z. Lai, L.-H. Lim, and K. Ye. *Simpler Grassmannian optimization*. 2020. eprint: 2009.13502 (math.OC).

[Man12] Y. Man Lui. “Advances in matrix manifolds for computer vision”. In: *Image and Vision Computing* 30.6–7 (2012), pp. 380–388. ISSN: 0262-8856. DOI: 10.1016/j.imavis.2011.08.002.

[MM16] 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. Cham: Springer International Publishing, 2016. ISBN: 3319450255, 9783319450254. DOI: 10.1007/978-3-319-45026-1.

[MS85] A. Machado and I. Salavessa. “Grassmannian manifolds as subsets of Euclidean spaces”. In: *Differential geometry (Santiago de Compostela, 1984)*. Vol. 131. Res. Notes in Math. Pitman, Boston, MA, 1985, pp. 85–102.

[Ngu13] T. S. Nguyen. “A real time procedure for affinely dependent parametric model order reduction using interpolation on Grassmann manifolds”. In: *International Journal for Numerical Methods in Engineering* 93.8 (2013), pp. 818–833. DOI: 10.1002/nme.4408.

[NS15] T. Nguyen and T. Stykel. “Model order reduction of parameterized circuit equations based on interpolation”. In: *Adv. Comput. Math* 41 (2015), pp. 1321–1342. DOI: 10.1007/s10444-015-9418-z.

[ONe83] B. O’Neill. *Semi-Riemannian geometry - With applications to relativity*. Vol. 103. Pure and Applied Mathematics. Academic Press, New York, 1983, pp. xiii+468. ISBN: 0-12-526740-1.

[QZL05] L. Qiu, Y. Zhang, and C.-K. Li. “Unitarily Invariant Metrics on the Grassmann Space”. In: *SIAM Journal on Matrix Analysis and Applications* 27.2 (2005), pp. 507–25. DOI: 10.1137/040607605.

[Rah+05] I. U. Rahman et al. “Multiscale Representations for Manifold-Valued Data”. In: *SIAM Journal on Multiscale Modeling and Simulation* 4.4 (2005), pp. 1201–1232. DOI: 10.1137/050622729.

[Ren13] Q. Rentmeesters. “Algorithms for data fitting on some common homogeneous spaces”. PhD thesis. Louvain, Belgium: Université Catholique de Louvain, 2013.

[Saa92] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Algortihms and Architectures for Advanced Scientific Computing. Manchester, UK: Manchester University Press, 1992.

[Sak77] T. Sakai. “On cut loci of compact symmetric spaces”. In: *Hokkaido Math. J.* 6.1 (1977), pp. 136–161. ISSN: 0385-4035. DOI: 10.14492/hokmj/1381758555.

[Sak96] T. Sakai. *Riemannian Geometry*. Fields Institute Communications. American Mathematical Soc., 1996. ISBN: 9780821889565. DOI: 10.1090/mmono/149.

[SL05] A. Srivastava and X. Liu. “Tools for application-driven linear dimension reduction”. In: *Neurocomputing* 67 (2005). Geometrical Methods in Neural Networks and Learning, pp. 136–160. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2004.11.036.

[ST15] A. Srivastava and P. K. Turaga. *Riemannian computing in computer vision*. English (US). Springer International Publishing, 2015. ISBN: 978-3-319-22957-7. DOI: 10.1007/978-3-319-22957-7.

[UM14] K. Usevich and I. Markovsky. "Optimization on a Grassmann manifold with application to system identification". In: *Automatica* 50.6 (2014), pp. 1656–1662. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2014.04.010.

[WC88] G. L. Wu and W. H. Chen. "A matrix inequality and its geometric applications". In: *Acta Math. Sinica* 31.3 (1988), pp. 348–355. ISSN: 0583-1431.

[WLL12] S. F. Walter, L. Lehmann, and R. Lamour. "On evaluating higher-order derivatives of the QR decomposition of tall matrices with full column rank in forward and reverse mode algorithmic differentiation". In: *Optimization Methods and Software* 27.2 (2012), pp. 391–403. DOI: 10.1080/10556788.2011.610454.

[Won67] Y.-C. Wong. "Differential Geometry of Grassmann Manifolds". In: *Proceedings of the National Academy of Sciences of the United States of America* 57 (1967), pp. 589–594. DOI: 10.1073/pnas.57.3.589.

[Won68a] Y.-C. Wong. "Conjugate loci in Grassmann manifolds". In: *Bull. Amer. Math. Soc.* 74 (1968), pp. 240–245. ISSN: 0002-9904. DOI: 10.1090/S0002-9904-1968-11903-2.

[Won68b] Y.-C. Wong. "Sectional Curvatures of Grassmann Manifolds". In: *Proceedings of the National Academy of Sciences of the United States of America* 60.1 (1968), pp. 75–79. ISSN: 00278424. DOI: 10.1073/pnas.60.1.75.

[YL16] K. Ye and L.-H. Lim. "Schubert varieties and distances between subspaces of different dimensions". In: *SIAM J. Matrix Anal. Appl.* 37.3 (2016), pp. 1176–1197. ISSN: 0895-4798. DOI: 10.1137/15M1054201.

[ZB16] D. Zhang and L. Balzano. "Global Convergence of a Grassmannian Gradient Descent Algorithm for Subspace Estimation". In: Cadiz, Spain: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS, May 2016, pp. 1460–1468.

[Zim14] R. Zimmermann. "A Locally Parametrized Reduced Order Model for the Linear Frequency Domain Approach to Time-Accurate Computational Fluid Dynamics". In: *SIAM Journal on Scientific Computing* 36.3 (2014), B508–B537. DOI: 10.1137/130942462.

[Zim19] R. Zimmermann. *Manifold interpolation and model reduction.* 2019. arXiv: 1902.06502 [math.NA].

[Zim20] R. Zimmermann. "Hermite Interpolation and Data Processing Errors on Riemannian Matrix Manifolds". In: *SIAM Journal on Scientific Computing* 42.5 (2020), A2593–A2619. DOI: 10.1137/19M1282878.

[ZPW18] R. Zimmermann, B. Peherstorfer, and K. Willcox. "Geometric subspace updates with applications to online adaptive nonlinear model reduction". In: *SIAM Journal on Matrix Analysis and Application* 39.1 (2018), pp. 234–261. DOI: 10.1137/17M1123286.

Chapter 3

Efficient Quasi-Geodesics on the Stiefel Manifold

Thomas Bendokat[a1], Ralf Zimmermann[a]

[a] Department of Mathematics and Computer Science, University of Southern Denmark (SDU), Odense, Denmark, {bendokat,zimmermann}@imada.sdu.dk

Abstract Solving the so-called geodesic endpoint problem, i.e., finding a geodesic that connects two given points on a manifold, is at the basis of virtually all data processing operations, including averaging, clustering, interpolation and optimization. On the Stiefel manifold of orthonormal frames, this problem is computationally involved. A remedy is to use quasi-geodesics as a replacement for the Riemannian geodesics. Quasi-geodesics feature constant speed and covariant acceleration with constant (but possibly non-zero) norm. For a well-known type of quasi-geodesics, we derive a new representation that is suited for large-scale computations. Moreover, we introduce a new kind of quasi-geodesics that turns out to be much closer to the Riemannian geodesics.

Keywords Stiefel manifold, Geodesic, Quasi-geodesic, Geodesic endpoint problem.

3.1 Introduction

Connecting two points on the Stiefel manifold with a geodesic requires the use of an iterative algorithm [Zim17], which raises issues such as convergence and computational costs. An alternative is to use *quasi-geodesics* [BKL17; JMS19; Kra+17; MLB20; NA05]. The term is used inconsistently. Here, we mean *curves with constant speed and covariant acceleration with constant (but possibly non-zero) norm.* The term quasi-geodesics is motivated by the fact that actual geodesics feature a constant-zero covariant acceleration. Such quasi-geodesics have been considered in [BKL17; JMS19; Kra+17; MLB20], where a representation of the Stiefel manifold with square matrices was used.

We introduce an economic way to compute these quasi-geodesics at considerably reduced computational costs. Furthermore, we propose a new kind of quasi-geodesics, which turn out to be closer to the true Riemannian geodesic but come at a slightly higher

[1]corresponding author

This paper corresponds to [BZ21a]. The final authenticated publication is available online at https://doi.org/10.1007/978-3-030-80209-7_82.

computational cost than the aforementioned economic quasi-geodesics. Both kinds of quasi-geodesics can be used for a wide range of problems, including optimization and interpolation of a set of points.

3.2 The Stiefel Manifold

This introductory exposition follows mainly [EAS98]. The manifold of orthonormal frames in $\mathbb{R}^{n\times p}$, i.e. the *Stiefel manifold*, is $\mathrm{St}(n,p) := \left\{U \in \mathbb{R}^{n\times p} \;\middle|\; U^TU = \mathrm{I}_p\right\}$. The *tangent space* at any point $U \in \mathrm{St}(n,p)$ can be parameterized as

$$T_U\mathrm{St}(n,p) := \left\{UA + U_\perp B \;\middle|\; A \in \mathfrak{so}(p),\ B \in \mathbb{R}^{(n-p)\times p}\right\},$$

where $\mathfrak{so}(p)$ denotes the real $p\times p$ skew-symmetric matrices and $U_\perp$ denotes an arbitrary but fixed orthonormal completion of U such that $\begin{pmatrix}U & U_\perp\end{pmatrix} \in \mathrm{O}(n)$. A Riemannian metric on $T_U\mathrm{St}(n,p)$ is induced by the *canonical* inner product

$$g_U\colon T_U\mathrm{St}(n,p)\times T_U\mathrm{St}(n,p)\to\mathbb{R},\ g_U(\Delta_1,\Delta_2) := \mathrm{tr}\left(\Delta_1^T(\mathrm{I}_n - \frac{1}{2}UU^T)\Delta_2\right).$$

The metric defines geodesics, i.e. locally shortest curves. The *Riemannian exponential* gives the geodesic from a point $U \in \mathrm{St}(n,p)$ in direction $\Delta = UA + U_\perp B \in T_U\mathrm{St}(n,p)$. Via the QR-decomposition $(\mathrm{I}_n - UU^T)\Delta = QR$, with $Q \in \mathrm{St}(n,p)$ and $R \in \mathbb{R}^{p\times p}$, it can be calculated as

$$\mathrm{Exp}_U(t\Delta) := \begin{pmatrix}U & U_\perp\end{pmatrix}\exp_{\mathrm{m}} t\begin{pmatrix}A & -B^T\\ B & 0\end{pmatrix}\begin{pmatrix}\mathrm{I}_p\\ 0\end{pmatrix} = \begin{pmatrix}U & Q\end{pmatrix}\exp_{\mathrm{m}} t\begin{pmatrix}A & -R^T\\ R & 0\end{pmatrix}\begin{pmatrix}\mathrm{I}_p\\ 0\end{pmatrix},$$

where $\exp_{\mathrm{m}}$ denotes the matrix exponential. The inverse problem, i.e. given $U,\tilde{U} \in \mathrm{St}(n,p)$, find $\Delta \in T_U\mathrm{St}(n,p)$ with $\tilde{U} = \mathrm{Exp}_U(\Delta)$, is called the *geodesic endpoint problem* and is associated with computing the Riemannian logarithm [Zim17]. There is no known closed formula. Yet, as suggested in [BKL17; Kra+17], one can exploit the quotient relation between $\mathrm{St}(n,p)$ and the *Grassmann manifold* [EAS98; BZA20] of p-dimensional subspaces of $\mathbb{R}^n$: Let $U,\tilde{U} \in \mathrm{St}(n,p)$. Then the columns of U and $\tilde{U}$ span subspaces, i.e. points on the Grassmannian. For the Grassmannian, the Riemannian logarithm is known [BZA20], which means that we know how to find $U_\perp B \in T_U\mathrm{St}(n,p)$ and $R \in \mathrm{O}(p)$ such that $\tilde{U}R = \begin{pmatrix}U & U_\perp\end{pmatrix}\exp_{\mathrm{m}}\left(\begin{smallmatrix}0 & -B^T\\ B & 0\end{smallmatrix}\right)\left(\begin{smallmatrix}\mathrm{I}_p\\ 0\end{smallmatrix}\right)$. Denote $A := \log_{\mathrm{m}}(R^T)$, where $\log_{\mathrm{m}}$ is the principle matrix logarithm. Then

$$\tilde{U} = \begin{pmatrix}U & U_\perp\end{pmatrix}\exp_{\mathrm{m}}\begin{pmatrix}0 & -B^T\\ B & 0\end{pmatrix}\begin{pmatrix}\mathrm{I}_p\\ 0\end{pmatrix}\exp_{\mathrm{m}}(A). \tag{3.2.1}$$

If there is a Stiefel geodesic from U to $\tilde{U}$, then there is also $U\tilde{A} + U_\perp\tilde{B} \in T_U\mathrm{St}(n,p)$ such that

$$\tilde{U} = \begin{pmatrix}U & U_\perp\end{pmatrix}\exp_{\mathrm{m}}\begin{pmatrix}\tilde{A} & -\tilde{B}^T\\ \tilde{B} & 0\end{pmatrix}\begin{pmatrix}\mathrm{I}_p\\ 0\end{pmatrix}. \tag{3.2.2}$$

Given $U, \tilde{U}$, we cannot find $U\tilde{A} + U_\perp \tilde{B} \in T_U\mathrm{St}(n,p)$ directly for (3.2.2), but we can find $UA + U_\perp B \in T_U\mathrm{St}(n,p)$ for (3.2.1). On the other hand, given $U \in \mathrm{St}(n,p)$ and $\Delta = UA + U_\perp B \in T_U\mathrm{St}(n,p)$, we can define a $\tilde{U} \in \mathrm{St}(n,p)$ via (3.2.1).

3.3 Quasi-Geodesics on the Stiefel Manifold

We first reduce the computational effort associated with the quasi-geodesics of [Kra+17; JMS19; BKL17]. Then, we introduce a new technique to construct quasi-geodesics.

3.3.1 Economy-size Quasi-Geodesics

Similarly to [Kra+17; JMS19; BKL17], we use the notion of a retraction [AMS08] as a starting point of the construction. Let M be a smooth manifold with tangent bundle TM. A *retraction* is a smooth mapping $R\colon TM \to M$ with the following properties:

1.) $R_x(0) = x$, i.e. R maps the zero tangent vector at $x \in M$ to x.

2.) The derivative at 0, $\mathrm{d}R_x(0)$ is the identity mapping on $T_0T_xM \simeq T_xM$.

Here, R_x denotes the restriction of R to T_xM. An example of a retraction is the Riemannian exponential mapping. On the Stiefel manifold, we can for any retraction $R\colon T\mathrm{St}(n,p) \to \mathrm{St}(n,p)$ and any tangent vector $\Delta \in T_U\mathrm{St}(n,p)$ define a smooth curve $\gamma_\Delta\colon t \mapsto R_U(t\Delta)$, which fulfills $\gamma_\Delta(0) = U$ and $\dot{\gamma}_\Delta(0) = \Delta$. The essential difference to [Kra+17; JMS19; BKL17] is that we work mainly with $n \times p$ matrices instead of $n \times n$ representatives, which entails a considerable cost reduction, when $p \leq \frac{n}{2}$.

The idea is to connect the subspaces spanned by the Stiefel manifold points with the associated Grassmann geodesic, while concurrently moving along the equivalence classes to start and end at the correct Stiefel representatives. This principle is visualized in [BKL17, Fig. 1]. We define the economy-size quasi-geodesics similarly to [Kra+17, Prop. 6 and Thm. 7].

Proposition 3.3.1. Let $U \in \mathrm{St}(n,p)$ and $\Delta = UA + U_\perp B \in T_U\mathrm{St}(n,p)$ with compact SVD $(\mathrm{I}_n - UU^T)\Delta = U_\perp B \overset{\mathrm{SVD}}{=} Q\Sigma V^T$. The mapping $\mathcal{RS}\colon T\mathrm{St}(n,p) \to \mathrm{St}(n,p)$, defined by Δ maps to $\mathcal{RS}_U(\Delta) := (UV\cos(\Sigma) + Q\sin(\Sigma))V^T \exp_{\mathrm{m}}(A)$, is a retraction with corresponding quasi-geodesic

$$\gamma(t) = \mathcal{RS}_U(t\Delta) = (UV\cos(t\Sigma) + Q\sin(t\Sigma))V^T \exp_{\mathrm{m}}(tA). \tag{3.3.1}$$

An orthogonal completion of $\gamma(t)$ is $\gamma_\perp(t) = \begin{pmatrix} U & U_\perp \end{pmatrix} \exp_{\mathrm{m}}\left(t\left(\begin{smallmatrix} 0 & -B^T \\ B & 0 \end{smallmatrix}\right)\right)\left(\begin{smallmatrix} 0 \\ \mathrm{I}_{n-p} \end{smallmatrix}\right)$. The quasi-geodesic γ has the following properties:

1. $\gamma(0) = U$
2. $\dot{\gamma}(t) = \gamma(t)A + \gamma_\perp(t)B\exp_{\mathrm{m}}(tA)$
3. $\|\dot{\gamma}(t)\|^2 = \frac{1}{2}\operatorname{tr}(A^TA) + \operatorname{tr}(B^TB)$ (constant speed)

4. $\ddot{\gamma}(t) = \gamma(t)(A^2 - \exp_{\mathrm{m}}(tA^T)B^T B \exp_{\mathrm{m}}(tA)) + 2\gamma_\perp(t)BA\exp_{\mathrm{m}}(tA)$
5. $D_t\dot{\gamma}(t) = \gamma_\perp(t)BA\exp_{\mathrm{m}}(tA)$
6. $\|D_t\dot{\gamma}(t)\|^2 = \|BA\|_F^2$ (constant-norm covariant acceleration)

Furthermore, γ is a geodesic if and only if $BA = 0$.

Proof. The fact that $\mathcal{RS}_U(0) = U$ for all $U \in \mathrm{St}(n,p)$ is obvious. Furthermore

$$\begin{aligned}\mathrm{d}\mathcal{RS}_U(0)(\Delta) &= (-UV\sin(\varepsilon\Sigma)\Sigma + Q\cos(\varepsilon\Sigma)\Sigma)V^T\exp_m(\varepsilon A)\big|_{\varepsilon=0}\\ &\quad + (UV\cos(\varepsilon\Sigma) + Q\sin(\varepsilon\Sigma))V^T\exp_{\mathrm{m}}(\varepsilon A)A\big|_{\varepsilon=0}\\ &= Q\Sigma V^T + UVV^TA = \Delta,\end{aligned}$$

so $\mathcal{RS}$ is a retraction. Note that γ can also be written as

$$\gamma(t) = \begin{pmatrix} U & U_\perp\end{pmatrix}\exp_{\mathrm{m}}\left(t\begin{pmatrix}0 & -B^T\\ B & 0\end{pmatrix}\right)\begin{pmatrix}\exp_{\mathrm{m}}(tA)\\ 0\end{pmatrix},$$

by comparison with the Grassmann geodesics in [EAS98, Thm. 2.3]. Therefore one possible orthogonal completion is given by the stated formula. The formulas for $\dot{\gamma}(t)$ and $\ddot{\gamma}(t)$ can be calculated by taking the derivative of

$$t \mapsto \begin{pmatrix}\gamma(t) & \gamma_\perp(t)\end{pmatrix} = \begin{pmatrix} U & U_\perp\end{pmatrix}\exp_{\mathrm{m}}\left(t\begin{pmatrix}0 & -B^T\\ B & 0\end{pmatrix}\right)\begin{pmatrix}\exp_{\mathrm{m}}(tA) & 0\\ 0 & \mathrm{I}_{n-p}\end{pmatrix}.$$

It follows that $\|\dot{\gamma}(t)\|^2 = \mathrm{tr}(\dot{\gamma}(t)^T(\mathrm{I}_n - \frac{1}{2}\gamma(t)\gamma(t)^T)\dot{\gamma}(t)) = \frac{1}{2}\mathrm{tr}(A^TA) + \mathrm{tr}(B^TB)$. To calculate the covariant derivative $D_t\dot{\gamma}(t)$, we use $\gamma(t)^T\gamma_\perp(t) = 0$ and the formula for the covariant derivative of $\dot{\gamma}$ along γ from [EAS98, eq. (2.41), (2.48)],

$$D_t\dot{\gamma}(t) = \ddot{\gamma}(t) + \dot{\gamma}(t)\dot{\gamma}(t)^T\gamma(t) + \gamma(t)\left((\gamma(t)^T\dot{\gamma}(t))^2 + \dot{\gamma}(t)^T\dot{\gamma}(t)\right), \tag{3.3.2}$$

cf. [Kra+17]. Since it features constant speed and constant-norm covariant acceleration, γ is a quasi-geodesic. It becomes a true geodesic if and only if $BA = 0$. □ □

Connecting $U, \tilde{U} \in \mathrm{St}(n,p)$ with a quasi-geodesic from Proposition 3.3.1 requires the inverse of $\mathcal{RS}_U$. Since $\mathcal{RS}_U$ is the Grassmann exponential – lifted to the Stiefel manifold – followed by a change of basis, we can make use of the modified algorithm from [BZA20] for the Grassmann logarithm. The procedure is stated in Algorithm 3.3.2. Proposition 3.3.3 confirms that it yields a quasi-geodesic.

Proposition 3.3.3. Let $U, \tilde{U} \in \mathrm{St}(n,p)$. Then Algorithm 3.3.2 returns a quasi-geodesic γ connecting U and $\tilde{U}$, i.e. $\gamma(0) = U$ and $\gamma(1) = \tilde{U}$, in direction $\dot{\gamma}(0) = UA + Q\Sigma V^T$ and of length $L(\gamma) = (\frac{1}{2}\mathrm{tr}(A^TA) + \mathrm{tr}(\Sigma^2))^{\frac{1}{2}}$.

Proof. Follows from [BZA20, Alg. 1], Proposition 3.3.1 and a straightforward calculation. □

Algorithm 3.3.2 Economy-size quasi-geodesic between two given points

Input: $U, \tilde{U} \in \mathrm{St}(n,p)$
1: $\widetilde{Q}\widetilde{S}\widetilde{R}^T \overset{\text{SVD}}{:=} \tilde{U}^T U$ ▷ SVD
2: $R := \widetilde{Q}\widetilde{R}^T$
3: $\tilde{U}_* := \tilde{U}R$ ▷ Change of basis in the subspace spanned by $\tilde{U}$
4: $A := \log_{\mathrm{m}}(R^T)$
5: $QSV^T \overset{\text{SVD}}{:=} (\mathrm{I}_n - UU^T)\tilde{U}_*$ ▷ compact SVD
6: $\Sigma := \arcsin(S)$ ▷ element-wise on the diagonal
Output: $\gamma(t) = (UV\cos(t\Sigma) + Q\sin(t\Sigma))V^T \exp_{\mathrm{m}}(tA)$

3.3.2 Short Economy-size Quasi-Geodesics

To construct an alternative type of quasi-geodesics, we make the following observation: Denote B in (3.2.1) by $\hat{B}$ and calculate the SVD $U_\perp \hat{B} = Q\Sigma V^T$. Furthermore, compute $R \in \mathrm{O}(p)$ as in Algorithm 3.3.2 and denote $a := \log_{\mathrm{m}}(R^T) \in \mathfrak{so}(p)$ and $b := \Sigma V^T \in \mathbb{R}^{p\times p}$. Then we can rewrite (3.2.1) as

$$\begin{aligned}\tilde{U} &= \begin{pmatrix} U & Q \end{pmatrix} \exp_{\mathrm{m}}\begin{pmatrix} 0 & -b^T \\ b & 0 \end{pmatrix} \begin{pmatrix} \mathrm{I}_p \\ 0_{p\times p} \end{pmatrix} \exp_{\mathrm{m}}(a) \\ &= \begin{pmatrix} U & Q \end{pmatrix} \exp_{\mathrm{m}}\begin{pmatrix} 0 & -b^T \\ b & 0 \end{pmatrix} \exp_{\mathrm{m}}\begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} \mathrm{I}_p \\ 0 \end{pmatrix} \qquad \text{for any } c \in \mathfrak{so}(p).\end{aligned}$$

Without the factor c, this is exactly what lead to the quasi-geodesics (3.3.1). There are however also matrices $A \in \mathfrak{so}(p), B \in \mathbb{R}^{p\times p}$ and $C \in \mathfrak{so}(p)$ satisfying

$$\begin{pmatrix} A & -B^T \\ B & C \end{pmatrix} = \log_{\mathrm{m}}\left(\exp_{\mathrm{m}}\begin{pmatrix} 0 & -b^T \\ b & 0 \end{pmatrix} \exp_{\mathrm{m}}\begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix} \right). \tag{3.3.3}$$

This implies that

$$\rho(t) = \begin{pmatrix} U & Q \end{pmatrix} \exp_{\mathrm{m}}\left(t \begin{pmatrix} A & -B^T \\ B & C \end{pmatrix} \right) \begin{pmatrix} \mathrm{I}_p \\ 0 \end{pmatrix} \tag{3.3.4}$$

is a curve from $\rho(0) = U$ to $\rho(1) = \tilde{U}$. It is indeed the projection of the geodesic in $\mathrm{O}(n)$ from $\begin{pmatrix} U & U_\perp \end{pmatrix}$ to $\begin{pmatrix} \tilde{U} & \tilde{U}_\perp \end{pmatrix}$ for some orthogonal completion $\tilde{U}_\perp$ of $\tilde{U}$. If $C = 0$, then ρ is exactly the Stiefel geodesic. For $x := \left(\begin{smallmatrix} 0 & -b^T \\ b & 0 \end{smallmatrix}\right)$ and $y := \left(\begin{smallmatrix} a & 0 \\ 0 & c \end{smallmatrix}\right)$ with $\|x\| + \|y\| \leq \ln(\sqrt{2})$, we can express C with help of the (Dynkin-)Baker-Campbell-Hausdorff (BCH) series $Z(x,y) = \log_{\mathrm{m}}(\exp_{\mathrm{m}}(x)\exp_{\mathrm{m}}(y))$, [Ros06, §1.3, p. 22]. To get close to the Riemannian geodesics, we want to find a c such that C becomes small. Three facts are now helpful for the solution:

1. The series $Z(x,y)$ depends only on iterated commutators $[\cdot,\cdot]$ of x and y.
2. Since the Grassmannian is symmetric, the Lie algebra $\mathfrak{so}(n)$ has a Cartan decomposition $\mathfrak{so}(n) = \mathfrak{v} \oplus \mathfrak{h}$ with $[\mathfrak{v},\mathfrak{v}] \subseteq \mathfrak{v}$, $[\mathfrak{v},\mathfrak{h}] \subseteq \mathfrak{h}$, $[\mathfrak{h},\mathfrak{h}] \subseteq \mathfrak{v}$, [EAS98, (2.38)].

3. For x and y defined as above, we have $x \in \mathfrak{h}$ and $y \in \mathfrak{v} \subset \mathfrak{so}(2p)$.

Considering terms up to combined order 4 in x and y in the Dynkin formula and denoting the anti-commutator by $\{x, y\} = xy + yx$, the matrix C is given by

$$C = C(c) = c + \frac{1}{12}\Big(2bab^T - \{bb^T, c\}\Big) - \frac{1}{24}\Big(2[c, bab^T] - [c, \{bb^T, c\}]\Big) + \text{h.o.t.}$$

Ignoring the higher-order terms, we can consider this as a fixed point problem $0 = C(c) \Leftrightarrow c = -\frac{1}{12}(\ldots) + \frac{1}{24}(\ldots)$. Performing a single iteration starting from $c_0 = 0$ yields

$$c_1 = c_1(a, b) = -\frac{1}{6}bab^T. \tag{3.3.5}$$

With $\delta := \max\{\|a\|, \|b\|\} \leq \ln(\sqrt{2})$, where $\|\cdot\|$ denotes the 2-norm, one can show that this choice of $c(a, b)$ produces a C-block with $\|C(c)\| \leq (\frac{7}{216} + \frac{1}{1-\delta})\delta^5$.

A closer look at curves of the form (3.3.4) shows the following Proposition.

Proposition 3.3.4. Let $U, Q \in \mathrm{St}(n, p)$ with $U^T Q = 0$ and $A \in \mathfrak{so}(p)$, $B \in \mathbb{R}^{p\times p}$, $C \in \mathfrak{so}(p)$. Then the curve

$$\rho(t) = \begin{pmatrix} U & Q \end{pmatrix} \exp_{\mathrm{m}}\left(t\begin{pmatrix} A & -B^T \\ B & C \end{pmatrix}\right)\begin{pmatrix} \mathrm{I}_p \\ 0 \end{pmatrix} \tag{3.3.6}$$

has the following properties, where $\rho_\perp(t) := \begin{pmatrix} U & Q \end{pmatrix} \exp_{\mathrm{m}}\left(t\begin{pmatrix} A & -B^T \\ B & C \end{pmatrix}\right)\begin{pmatrix} 0 \\ \mathrm{I}_p \end{pmatrix}$:

1. $\rho(0) = U$
2. $\dot\rho(0) = UA + QB \in T_U\mathrm{St}(n, p)$
3. $\dot\rho(t) = \begin{pmatrix} U & Q \end{pmatrix} \exp_{\mathrm{m}}\left(t\begin{pmatrix} A & -B^T \\ B & C \end{pmatrix}\right)\begin{pmatrix} A \\ B \end{pmatrix} = \rho(t)A + \rho_\perp(t)B \in T_{\rho(t)}\mathrm{St}(n, p)$
4. $\|\dot\rho(t)\|^2 = \frac{1}{2}\operatorname{tr}(A^T A) + \operatorname{tr}(B^T B)$ (constant speed)
5. $\ddot\rho(t) = \begin{pmatrix} U & Q \end{pmatrix} \exp_{\mathrm{m}}\left(t\begin{pmatrix} A & -B^T \\ B & C \end{pmatrix}\right)\begin{pmatrix} A^2 - B^T B \\ BA + CB \end{pmatrix}$
$= \rho(t)(A^2 - B^T B) + \rho_\perp(t)(BA + CB)$
6. $D_t\dot\rho(t) = \rho_\perp(t)CB$
7. $\|D_t\dot\rho(t)\|^2 = \|CB\|_F^2$ (constant-norm covariant acceleration)

Proof. This can directly be checked by calculation and making use of the formula (3.3.2) for the covariant derivative $D_t\dot\rho(t)$. ◻

Note that the property $U^TQ = 0$ can only be fulfilled if $p \leq \frac{n}{2}$. Since they feature constant speed and constant-norm covariant acceleration, curves of the form (3.3.6) are *quasi-geodesics*. Now we can connect two points on the Stiefel manifold with Algorithm 3.3.5, making use of $\exp_m\left(\begin{smallmatrix} 0 & -\Sigma \\ \Sigma & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} \cos\Sigma & -\sin\Sigma \\ \sin\Sigma & \cos\Sigma \end{smallmatrix}\right)$ for diagonal Σ. Curves produced by

Algorithm 3.3.5 Short economy-size quasi-geodesic between two given points

Input: $U, \tilde{U} \in \mathrm{St}(n,p)$

1: $\widetilde{Q}\widetilde{S}\widetilde{R}^T \overset{\text{SVD}}{:=} \tilde{U}^T U$ ▷ SVD
2: $R := \widetilde{Q}\widetilde{R}^T$
3: $a := \log_m(R^T)$
4: $QSV^T \overset{\text{SVD}}{:=} (\mathrm{I}_n - UU^T)\tilde{U}R$ ▷ compact SVD
5: $\Sigma := \arcsin(S)$ ▷ element-wise on the diagonal
6: $b := \Sigma V^T$
7: $c := c(a,b) = -\frac{1}{6}bab^T$
8: $\begin{pmatrix} A & -B^T \\ B & C \end{pmatrix} = \log_m \begin{pmatrix} V\cos(\Sigma)V^TR^T & -V\sin(\Sigma)\exp_m(c) \\ \sin(\Sigma)V^TR^T & \cos(\Sigma)\exp_m(c) \end{pmatrix}$

Output: $\rho(t) = \begin{pmatrix} U & Q \end{pmatrix} \exp_m\left(t\begin{pmatrix} A & -B^T \\ B & C \end{pmatrix}\right)\begin{pmatrix} \mathrm{I}_p \\ 0 \end{pmatrix}$

Algorithm 3.3.5 are of the form (3.3.6). We numerically verified that they are closer to the Riemannian geodesic than the economy-size quasi-geodesics of Algorithm 3.3.2 in the cases we considered. As geodesics are locally shortest curves, this motivates the term *short economy-size quasi-geodesics* for the curves produced by Algorithm 3.3.5.

Note that Algorithm 3.3.5 allows to compute the quasi-geodesic $\rho(t)$ with initial velocity $\dot{\rho}(0) = UA + QB$ between two given points. The opposite problem, namely finding the quasi-geodesic $\rho(t)$ given a point $U \in \mathrm{St}(n,p)$ with tangent vector $UA+QB \in T_U\mathrm{St}(n,p)$, is however not solved, since the correct $C \in \mathfrak{so}(p)$ is missing. Nevertheless, since $\rho(t)$ is an approximation of a geodesic, the Riemannian exponential can be used to generate an endpoint.

3.4 Numerical Comparison

To compare the behaviour of the economy-size and the short economy-size quasi-geodesics, two random points on the Stiefel manifold $\mathrm{St}(200,30)$ with a distance of $d \in \{0.1\pi, 0.5\pi, 1.3\pi\}$ from each other are created. Then the quasi-geodesics according to Algorithms 3.3.2 and 3.3.5 are calculated. The Riemannian distance, i.e., the norm of the Riemannian logarithm, between the quasi-geodesics and the actual Riemannian geodesic is plotted in Figure 3.1. In all cases considered, the short quasi-geodesics turn out to be two to five orders of magnitude closer to the Riemannian geodesic than the economy-size quasi-geodesics.

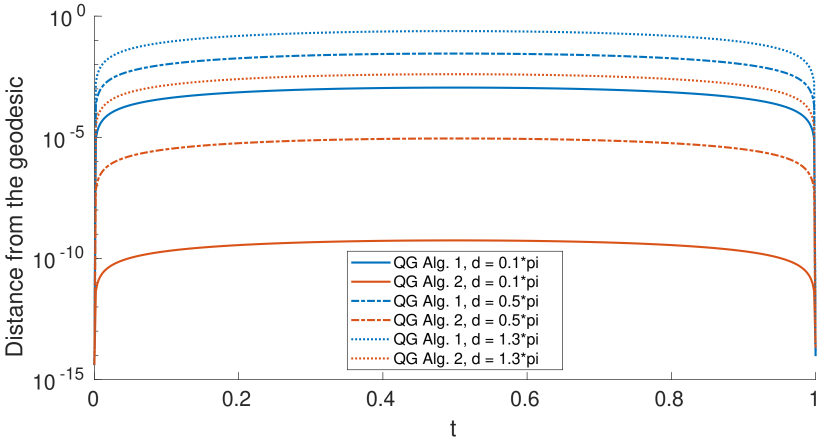

Figure 3.1: Comparison of the distance of the quasi-geodesics (QG) to the true geodesic between two random points. The distance between the two points is denoted by d.[2]

Table 3.1: Comparison of the relative deviation in length of the quasi-geodesics compared to the true geodesic between two random points on $\mathrm{St}(200, p)$ at a distance of $\frac{\pi}{2}$ for different values of p. The observable p-dependence suggests further investigations.

p	Algorithm 3.3.5	Algorithm 3.3.2	p	Algorithm 3.3.5	Algorithm 3.3.2
10	1.1997e-08	9.9308e-04	60	4.0840e-12	7.2356e-04
20	3.6416e-10	8.6301e-04	70	1.9134e-12	6.8109e-04
30	8.4590e-11	8.9039e-04	80	9.9672e-13	6.1304e-04
40	1.9998e-11	7.7265e-04	90	5.7957e-13	5.7272e-04
50	8.1613e-12	7.7672e-04	100	2.7409e-13	4.9691e-04

In Table 3.1, we display the relative deviation of the length of the quasi-geodesics from the Riemannian distance between two randomly generated points at a distance of $\frac{\pi}{2}$ on $\mathrm{St}(200, p)$, where p varies between 10 and 100. The outcome justifies the name *short* quasi-geodesics.

The essential difference in terms of the computational costs between the quasi-geodesics of Alg. 3.3.2, those of Alg. 3.3.5, and the approach in [Kra+17, Thm. 7] is that the they require matrix exp- and log-function evaluations of $(p \times p)$-, $(2p \times 2p)$- and $(n \times n)$-matrices, respectively.

[2] In the figure, Alg. 1 refers to Algorithm 3.3.2 and Alg. 2 refers to Algorithm 3.3.5.

3.5 Conclusion and Outlook

We have proposed a new efficient representation for a well-known type of quasi-geodesics on the Stiefel manifold, which is suitable for large-scale computations and has an exact inverse to the endpoint problem for a given tangent vector. Furthermore, we have introduced a new kind of quasi-geodesics, which are much closer to the Riemannian geodesics. These can be used for endpoint problems, but the exact curve for a given tangent vector is unknown. Both kinds of quasi-geodesics can be used, e.g., for interpolation methods like De Casteljau etc. [Kra+17; BKL17]. In future work, further and more rigorous studies of the quasi-geodesics' length properties and the p-dependence displayed in Table 3.1 are of interest.

References

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, New Jersey: Princeton University Press, 2008. URL: http://press.princeton.edu/titles/8586.html.

[BKL17] J. Batista, K. Krakowski, and F. S. Leite. "Exploring quasi-geodesics on Stiefel manifolds in order to smooth interpolate between domains". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. 2017, pp. 6395–6402.

[BZA20] T. Bendokat, R. Zimmermann, and P.-A. Absil. *A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects*. 2020. arXiv: 2011.13699.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith. "The Geometry of Algorithms with Orthogonality Constraints". In: *SIAM Journal on Matrix Analysis and Applications* 20.2 (Apr. 1998), pp. 303–353. ISSN: 0895-4798. DOI: 10.1137/S0895479895290954.

[JMS19] V. Jurdjevic, I. Markina, and F. Silva Leite. "Extremal Curves on Stiefel and Grassmann Manifolds". English. In: *The Journal of Geometric Analysis* (2019).

[Kra+17] K. A. Krakowski et al. "A modified Casteljau algorithm to solve interpolation problems on Stiefel manifolds". In: *Journal of Computational and Applied Mathematics* 311 (2017), pp. 84–99. ISSN: 0377-0427.

[MLB20] L. Machado, F. S. Leite, and E. Batzies. "Geometric Algorithm to Generate Interpolating Splines on Grassmann and Stiefel Manifolds". English. In: *CONTROLO 2020*. Cham: Springer International Publishing, 2020, pp. 180–189. ISBN: 1876-1100.

[NA05] Y. Nishimori and S. Akaho. "Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold". In: *Neurocomputing* 67 (2005), pp. 106–135. ISSN: 0925-2312.

[Ros06] W. Rossmann. *Lie Groups: An Introduction Through Linear Groups*. Oxford Graduate Texts in Mathematics. Oxford University Press, 2006. ISBN: 9780199202515.

[Zim17] R. Zimmermann. "A Matrix-Algebraic Algorithm for the Riemannian Logarithm on the Stiefel Manifold under the Canonical Metric". In: *SIAM J. Matrix. Anal. Appl.* 38.2 (2017), pp. 322–342.

Part II

The Symplectic Stiefel and Grassmann Manifolds

Chapter 4

The real symplectic Stiefel and Grassmann manifolds: metrics, geodesics and applications

Thomas Bendokat[a1], Ralf Zimmermann[a]

[a] Department of Mathematics and Computer Science, University of Southern Denmark (SDU), Odense, Denmark, {bendokat,zimmermann}@imada.sdu.dk

Abstract The real symplectic Stiefel manifold is the manifold of symplectic bases of symplectic subspaces of a fixed dimension. It features in a large variety of applications in physics and engineering. In this work, we study this manifold with the goal of providing theory and matrix-based numerical tools fit for basic data processing. Geodesics are fundamental for data processing. However, these are so far unknown. Pursuing a Lie group approach, we close this gap and derive efficiently computable formulas for the geodesics both with respect to a natural pseudo-Riemannian metric and a novel Riemannian metric. In addition, we provide efficiently computable and invertible retractions. Moreover, we introduce the real symplectic Grassmann manifold, i.e., the manifold of symplectic subspaces. Again, we derive efficient formulas for pseudo-Riemannian and Riemannian geodesics and invertible retractions. The findings are illustrated by numerical experiments, where we consider optimization via gradient descent on both manifolds and compare the proposed methods with the state of the art. In particular, we treat the 'nearest symplectic matrix' problem and the problem of optimal data representation via a low-rank symplectic subspace. The latter task is associated with the problem of finding a 'proper symplectic decomposition', which is important in structure-preserving model order reduction of Hamiltonian systems.

Keywords symplectic Stiefel manifold, symplectic Grassmann manifold, pseudo-Riemannian metric, Riemannian metric, geodesic, Riemannian optimization, Hamiltonian model order reduction, proper symplectic decomposition, symplectic group

AMS subject classifications 22E70, 53-08, 53B20, 53B30, 53B50, 53Z05, 70G45

[1] corresponding author
The arXiv version of this paper can be found at [BZ21b]. The paper has been submitted to SIMAX.

4.1 Introduction

The central object under study in this work is the *real symplectic Stiefel manifold*. The elements of this matrix manifold are the symplectic bases of fixed order $2k$ of symplectic subspaces in $\mathbb{R}^{2n}$,

$$\mathrm{SpSt}(2n, 2k) := \left\{U \in \mathbb{R}^{2n\times 2k} \mid U^T J_{2n} U = J_{2k}\right\}, \quad J_{2m} = \begin{bmatrix} 0 & I_m \\ -I_m & 0 \end{bmatrix}, m \in \{n, k\}.$$

Symplectic structures feature in a large variety of applications in physics and engineering, most prominently in Hamiltonian mechanics [AG01]. Hamiltonian systems are used in applications ranging from molecular dynamics to celestial mechanics, see [HLW06] and references therein. The symplectic Stiefel manifold is of special importance to optimization problems of the form

$$\begin{aligned} \min_{U\in\mathbb{R}^{2n\times 2k}} \quad & f(U) \\ \text{s. t.} \quad & U^T J_{2n} U = J_{2k}, \end{aligned} \tag{4.1.1}$$

see [Gao+21b] and references therein, since it allows to tackle such *constrained* optimization problems on $\mathbb{R}^{2n\times 2k}$ as *unconstrained* optimization problems on $\mathrm{SpSt}(2n, 2k)$. Fields of applications include the symplectic eigenvalue problem [BJ15; Son+21] and projection-based structure-preserving model order reduction for Hamiltonian systems. Here, an optimization problem of the form (4.1.1) appears as the central task of computing a so-called *proper symplectic decomposition* [AH17; PM16; BBH19].

Riemannian optimization requires that we have explicit formulas for essential geometric quantities at hand, as well as efficient algorithms for practical computations. It is understood that the inner geometry of the symplectic Stiefel manifold depends on the chosen metric.

Related work and state of the art Optimization on the real symplectic group is considered in [Fio11] with respect to a pseudo-Riemannian metric and in [WSF18; BCC20] with respect to a left-invariant Riemannian metric. Quotients of the real symplectic group relating to the real symplectic Stiefel and Grassmann manifolds are briefly introduced in [Sed19, Subsection 2.1]. To the best of the authors' knowledge, the first treatment of the real symplectic Stiefel manifold with a view on numerical applications is [Gao+21b]. The optimization algorithm developed there forms the state of the art. The same team of authors compared this method with optimization with respect to a Riemannian metric that stems from a Euclidean metric in the later work [Gao+21a].

Main original contributions Starting from the classical real symplectic group, equipped with a bi-invariant pseudo-Riemannian metric, we use a Lie group approach to investigate the real symplectic Stiefel manifold. This original approach allows us to exploit quotient manifold results from semi-Riemannian geometry [ONe83] and enables us to derive the first closed-form expressions for the corresponding pseudo-Riemannian geodesics on $\mathrm{SpSt}(2n, 2k)$. Complementary to the pseudo-Riemannian approach, we also introduce a Riemannian metric and derive closed-form expressions for the corresponding Riemannian geodesics. In view of optimization tasks, we provide a formula for the gradient associated with the Riemannian metric and efficiently computable and invertible retractions that approximate the pseudo-Riemannian geodesics.

Moreover, we initiate a study of the manifold of symplectic subspaces that are spanned by symplectic bases, which we term the *real symplectic Grassmann manifold* in analogy to the classical Stiefel and Grassmann manifolds. Continuing the quotient manifold approach, we derive corresponding formulas for pseudo-Riemannian and Riemannian geodesics and retractions. We promote symplectic subspaces as the main objects of interest in structure-preserving model order reduction of parameterized Hamiltonian systems.

We illustrate the theoretical findings by means of numerical examples. More precisely, we investigate the numerical feasibility of the proposed methods, we tackle the nearest symplectic matrix problem on the real symplectic Stiefel manifold and compute the optimal symplectic subspace containing a given data matrix on the real symplectic Grassmann manifold. The latter problem is directly associated with finding a proper symplectic decomposition in the context of structure preserving model reduction. We juxtapose the methods' performance with the state of the art.

Organisation Section 4.2 reviews basic facts on the real symplectic group and states its geodesics associated with a natural bi-invariant pseudo-Riemannian metric and a right-invariant Riemannian metric. In Section 4.3 we investigate the real symplectic Stiefel manifold, where we cover basic geometry, Riemannian and pseudo-Riemannian metrics and their geodesics as well as the Riemannian gradient. Section 4.4 introduces the real symplectic Grassmann manifold as a quotient space and provides formulas for the inherited metrics and geodesics and the Riemannian gradient. Suitable retractions fit for replacing the actual geodesics in efficient implementations are given in Section 4.5. Numerical experiments are contained in Section 4.6 and Section 4.7 concludes on the paper.

4.2 The real symplectic group

Symplectic vectors spaces are the objects of interest for the (local) study of Hamiltonian systems. An introduction can be found in [AG01]. By definition, a *real symplectic vector space* is a real vector space $\mathcal{V}$ together with a nondegenerate, skew-symmetric bilinear form $\omega\colon \mathcal{V} \times \mathcal{V} \to \mathbb{R}$. This means ω is bilinear and fulfills

1. $\omega(x, y) = 0$ for all $y \in \mathcal{V}$ implies $x = 0$ (nondegenerate),

2. $\omega(x, y) = -\omega(y, x)$ (skew-symmetric).

As a standard result, such a $\mathcal{V}$ is even-dimensional. For any subspace $\mathcal{U} \subset \mathcal{V}$, the symplectic form ω allows to the define the *symplectic complement*

$$\mathcal{U}^{\perp} := \{v \in \mathcal{V} \mid \omega(v,u) = 0 \ \forall u \in \mathcal{U}\}.$$

Since in general $\mathcal{U}^{\perp} \cap \mathcal{U} \neq \{0\}$, four special cases of subspaces are classified. A subspace $\mathcal{U}$ of $(\mathcal{V}, \omega)$ is called

1. *isotropic*, if $\mathcal{U} \subset \mathcal{U}^{\perp}$,
2. *coisotropic*, if $\mathcal{U}^{\perp} \subset \mathcal{U}$,
3. *Langrangian*, if $\mathcal{U}^{\perp} = \mathcal{U}$, and
4. *symplectic*, if $\mathcal{U}^{\perp} \cap \mathcal{U} = \{0\}$.

The last case, a symplectic subspace, means that ω restricts to a symplectic form on $\mathcal{U}$, i.e., $(\mathcal{U}, \omega|_{\mathcal{U}})$ is a symplectic space in itself.

The *linear Darboux theorem* [AG01] states that for any two symplectic vector spaces of the same dimension, there is a linear isomorphism between them preserving the symplectic form. We can therefore restrict our considerations to the *standard symplectic vector space* $(\mathbb{R}^{2n}, \omega_0)$, where the *standard symplectic form* is defined as

$$\omega_0(x,y) := x^T J_{2n} y \quad \text{with} \quad J_{2n} := \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix},$$

where $n \in \mathbb{N}$ and I_n is the $n \times n$ identity matrix. Note that $J_{2n}^T = -J_{2n} = J_{2n}^{-1}$.

The real symplectic group is the matrix Lie group of transformations which leave the standard symplectic form invariant. It has been studied for example in [Fio11; WSF18; BCC20] with a view on applications, and in [AG01] from a more abstract point of view.

Define for any matrix $A \in \mathbb{R}^{2n \times 2k}$ the *symplectic inverse* [PM16]

$$A^+ := J_{2k}^T A^T J_{2n}.$$

The *real symplectic group* is then defined as

$$\mathrm{Sp}(2n, \mathbb{R}) := \left\{ M \in \mathbb{R}^{2n \times 2n} \;\middle|\; M^+ M = I_{2n} \right\} = \left\{ M \in \mathbb{R}^{2n \times 2n} \;\middle|\; M^T J_{2n} M = J_{2n} \right\}.$$

For any $M \in \mathrm{Sp}(2n, \mathbb{R})$, and any $x, y \in \mathbb{R}^{2n}$, it holds that

$$\omega_0(x,y) = x^T J_{2n} y = x^T M^T J_{2n} M y = \omega_0(Mx, My).$$

The corresponding Lie algebra is given by the *Hamiltonian matrices*

$$\mathfrak{sp}(2n, \mathbb{R}) := \left\{ \Omega \in \mathbb{R}^{2n \times 2n} \;\middle|\; \Omega^+ = -\Omega \right\}.$$

Accordingly, the tangent space of the real symplectic group at M is given by translation by M, i.e.

$$\begin{aligned} T_M \mathrm{Sp}(2n, \mathbb{R}) &= \left\{ M\Omega \in \mathbb{R}^{2n \times 2n} \;\middle|\; \Omega \in \mathfrak{sp}(2n, \mathbb{R}) \right\} \\ &= \left\{ \Omega M \in \mathbb{R}^{2n \times 2n} \;\middle|\; \Omega \in \mathfrak{sp}(2n, \mathbb{R}) \right\}, \end{aligned}$$

The dimension of the symplectic group is $\dim \mathrm{Sp}(2n, \mathbb{R}) = (2n+1)n$, see also [Fio11].

4.2.1 Pseudo-Riemannian metric

Similarly to [Fio11], we define a bi-invariant pseudo-Riemannian metric on $\mathrm{Sp}(2n,\mathbb{R})$ by $h_M \colon T_M\mathrm{Sp}(2n,\mathbb{R}) \times T_M\mathrm{Sp}(2n,\mathbb{R}) \to \mathbb{R}$,

$$h_M(X_1, X_2) := \langle X_1, X_2\rangle_M := \frac{1}{2}\operatorname{tr}(X_1^+ X_2), \quad X_1, X_2 \in T_M\mathrm{Sp}(2n,\mathbb{R}). \tag{4.2.1}$$

The factor $\frac{1}{2}$ is introduced for convenience. If $X_1 = M\Omega_1$ and $X_2 = M\Omega_2$, then

$$\langle X_1, X_2\rangle_M = \frac{1}{2}\operatorname{tr}(\Omega_1^+ M^+ M\Omega_2) = -\frac{1}{2}\operatorname{tr}(\Omega_1\Omega_2),$$

i.e. $\langle\cdot,\cdot\rangle_M$ is exactly $-\frac{1}{2}$ times the pseudo-Riemannian metric defined in [Fio11] and therefore $-\frac{1}{2}$ times the Khvedelidze–Mladenov metric [KM02] on the general linear group. By properties of the trace and the symplectic inverse, it can be immediately verified that the pseudo-Riemannian metric defined in this way is bi-invariant. Therefore, making use of [ONe83, Proposition 11.9], the (pseudo-Riemannian) geodesics are given by the one-parameter subgroups

$$\mathrm{Exp}_M^{\mathrm{Sp},h}(tX) := M\exp_{\mathrm{m}}(tM^+X) = M\exp_{\mathrm{m}}(t\Omega),$$

where $X = M\Omega \in T_M\mathrm{Sp}(2n,\mathbb{R})$ and $\exp_{\mathrm{m}}$ denotes the matrix exponential. This corresponds to [Fio11, Theorem 2.4].

4.2.2 Riemannian metric

The pseudo-Riemannian metric (4.2.1) is bi-invariant, but is not positive definite. Especially for optimization problems, a (by definition positive definite) Riemannian metric can be advantageous, and there exists a vast amount of literature concerning Riemannian optimization. While a left-invariant Riemannian metric on $\mathrm{Sp}(2n,\mathbb{R})$ was introduced in [WSF18], we introduce a right-invariant Riemannian metric in anticipation of the quotient structure that is considered in the upcoming Section 4.3. We also derive the corresponding gradient and geodesics.

The mapping $g_M \colon T_M\mathrm{Sp}(2n,\mathbb{R}) \times T_M\mathrm{Sp}(2n,\mathbb{R}) \to \mathbb{R}$,

$$g_M(X_1, X_2) := \frac{1}{2}\operatorname{tr}((X_1M^+)^T X_2M^+), \quad X_1, X_2 \in T_M\mathrm{Sp}(2n,\mathbb{R}), \tag{4.2.2}$$

defines point-wise a right-invariant Riemannian metric on the real symplectic group $\mathrm{Sp}(2n,\mathbb{R})$. The right-invariance follows from the fact that for every $N \in \mathrm{Sp}(2n,\mathbb{R})$, $g_{MN}(X_1N, X_2N) = \frac{1}{2}\operatorname{tr}((X_1NN^+M^+)^T X_2NN^+M^+) = g_M(X_1, X_2)$.

The Riemannian gradient for this metric is given as follows: Let $f \colon \mathrm{Sp}(2n,\mathbb{R}) \to \mathbb{R}$ be differentiable and let ∇f_M be the Euclidean gradient of a continuous extension of f to an open subset of $\mathbb{R}^{2n\times 2n}$ around $M \in \mathrm{Sp}(2n,\mathbb{R})$, evaluated at M. Then the Riemannian gradient of f at M (with respect to the metric g_M) is

$$\mathrm{grad}_f^g(M) = \nabla f_M M^T M + J_{2n}M(\nabla f_M)^T J_{2n}M \in T_M\mathrm{Sp}(2n,\mathbb{R}).$$

This follows from the fact that by definition $\operatorname{grad}_f^g(M)$ is the unique tangent vector at M such that $g_M(\operatorname{grad}_f^g(M), X) = \mathrm{d}f_M(X) = \operatorname{tr}((\nabla f_M)^T X)$ holds for all $X \in T_M\mathrm{Sp}(2n,\mathbb{R})$. By making use of the fact that $M^+X = -X^+M$, $\operatorname{grad}_f^g(M)$ solves this equation, and $\operatorname{grad}_f^g(M) \in T_M(\mathrm{Sp}(2n,\mathbb{R}))$ follows from $\operatorname{grad}_f^g(M)M^+ = -M(\operatorname{grad}_f^g(M))^+$.

We can derive the Riemannian geodesics corresponding to the Riemannian metric (4.2.2) analogously to [VAV13, Proposition 4.2], where the Riemannian geodesics corresponding to a right-invariant metric on the general linear group $\mathrm{GL}(n)$ were derived.

Proposition 4.2.1. Let $M \in \mathrm{Sp}(2n,\mathbb{R})$ and $X \in T_M\mathrm{Sp}(2n,\mathbb{R})$. The Riemannian geodesic γ with $\gamma(0) = M$ and $\dot\gamma(0) = X$ for the Riemannian metric (4.2.2) is given by

$$\gamma(t) := \mathrm{Exp}_M^{\mathrm{Sp},g}(tX) := \exp_{\mathrm{m}}(t(XM^+ - (XM^+)^T))\exp_{\mathrm{m}}(t(XM^+)^T)M.$$

Proof. The proof of [VAV13, Proposition 4.2] can be transferred straightforwardly to this setting. □

4.3 The real symplectic Stiefel manifold

The *real symplectic Stiefel manifold* is defined as

$$\mathrm{SpSt}(2n,2k) := \left\{U \in \mathbb{R}^{2n\times 2k} \;\middle|\; U^+U = I_{2k}\right\} = \left\{U \in \mathbb{R}^{2n\times 2k} \;\middle|\; U^TJ_{2n}U = J_{2k}\right\}.$$

It contains the matrices $U \in \mathbb{R}^{2n\times 2k}$, whose column vectors form symplectic bases for the $2k$-dimensional symplectic subspaces of $(\mathbb{R}^{2n},\omega_0)$ and was treated in [Gao+21b; Gao+21a; Son+21]. Note the formal similarity with the (compact) Stiefel manifold $\mathrm{St}(n,k) = \left\{U \in \mathbb{R}^{n\times k} \;\middle|\; U^TU = I_k\right\}$. As a novelty, and in contrast to the aforementioned references, we will pursue a Lie group-based approach to study the real symplectic Stiefel manifold. We will furthermore introduce a new pseudo-Riemannian and a new Riemannian metric and derive the geodesics for both.

Denote the projection onto the first k columns of a matrix, when multiplied from the right, by

$$I_{n,k} := \begin{bmatrix} I_k \\ 0 \end{bmatrix} \in \mathbb{R}^{n\times k}, \tag{4.3.1}$$

and the projection onto the first k and the $(n+1)$th to the $(n+k)$th column by

$$E := \begin{bmatrix} I_{n,k} & 0 \\ 0 & I_{n,k} \end{bmatrix} \in \mathbb{R}^{2n\times 2k}.$$

Our first goal is to recognize the real symplectic Stiefel manifold as a quotient of the real symplectic group. To this end, we introduce the following canonical projection:

$$\pi\colon \mathrm{Sp}(2n,\mathbb{R}) \to \mathrm{SpSt}(2n,2k),\ M \mapsto ME. \tag{4.3.2}$$

Proposition 4.3.1. The real symplectic Stiefel manifold is diffeomorphic to the quotient

$$\mathrm{SpSt}(2n, 2k) \cong \mathrm{Sp}(2n, \mathbb{R})/\mathrm{Sp}(2(n-k), \mathbb{R}).$$

It has dimension $\dim(\mathrm{SpSt}(2n, 2k)) = (4n - 2k + 1)k$.

Proof. The set $\mathrm{SpSt}(2n, 2k)$ is the orbit of E under the group action of $\mathrm{Sp}(2n, \mathbb{R})$ that is induced by left-multiplication. The stabilizer

$$\mathrm{stab}_E := \{M \in \mathrm{Sp}(2n, \mathbb{R}) \mid ME = E\}$$

of this group action is isomorphic to $\mathrm{Sp}(2(n-k), \mathbb{R})$. From [Lee12, Theorem 21.20], it follows that $\mathrm{SpSt}(2n, 2k)$ has a unique smooth manifold structure for which the group action is smooth. It furthermore follows that the dimension of the real symplectic Stiefel manifold is

$$\dim(\mathrm{SpSt}(2n, 2k)) = \dim(\mathrm{Sp}(2n, \mathbb{R})) - \dim(\mathrm{Sp}(2(n-k), \mathbb{R})) = (4n - 2k + 1)k,$$

in accordance with [Gao+21b]. From [Lee12, Theorem 21.18], the existence of a diffeomorphism between $\mathrm{SpSt}(2n, 2k)$ and $\mathrm{Sp}(2n, \mathbb{R})/\mathrm{Sp}(2(n-k), \mathbb{R})$ follows. □

The quotient manifold structure of the real symplectic Stiefel manifold (and of the real symplectic Grassmann manifold, which is to be discussed later on) with the symplectic group as the associated total space is visualized in Figure 4.1.

The Lie group approach allows to represent tangent vectors in a similar way as is common for the standard Stiefel manifold $\mathrm{St}(n, k)$. As the projection π in (4.3.2) is surjective, for every $U \in \mathrm{SpSt}(2n, 2k)$, there is an $M \in \mathrm{Sp}(2n, \mathbb{R})$ such that $U = ME$. Define a *symplectic complement* of U by $U^s := ME^s$, where

$$E^s := \begin{bmatrix} 0_{k\times(n-k)} & 0_{k\times(n-k)} \\ I_{n-k} & 0_{n-k} \\ 0_{k\times(n-k)} & 0_{k\times(n-k)} \\ 0_{n-k} & I_{n-k} \end{bmatrix} \in \mathbb{R}^{2n\times 2(n-k)},$$

i.e. the projection onto the columns complementary to those selected by E. Note that $E^+ = E^T$ and $(U^s)^+U^s = (E^s)^+M^+ME^s = I_{2(n-k)}$, i.e. $U^s \in \mathrm{SpSt}(2n, 2(n-k))$. Furthermore $U^+U^s = E^+M^+ME^s = 0$ and $I_{2n} - UU^+ = M(I_{2n} - EE^+)M^+ = U^s(U^s)^+$.

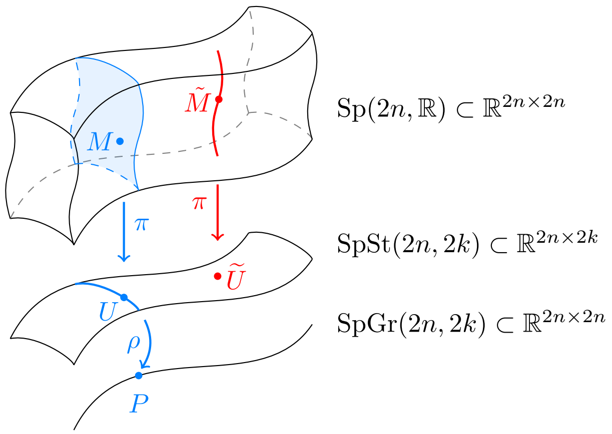

Figure 4.1: Visualization of the quotient structure of the real symplectic Grassmann and Stiefel manifold with respect to the symplectic group. Any point $P \in \mathrm{SpGr}(2n, 2k)$ has an equivalence class in $\mathrm{SpSt}(2n, 2k)$ as its pre-image under ρ, visualized by the blue line through $U \in \mathrm{SpSt}(2n, 2k)$. This equivalence class in turn has again an equivalence class of symplectic matrices in $\mathrm{Sp}(2n, \mathbb{R})$ as its pre-image under π, visualized by the blue area around M. The equivalence class of a single point $\tilde{U} \in \mathrm{SpSt}(2n, 2k)$ is of lower dimension, visualized by the red line through $\tilde{M} \in \mathrm{Sp}(2n, \mathbb{R})$.

Proposition 4.3.2. The tangent space at $U \in \mathrm{SpSt}(2n, 2k)$ is given by

$$\begin{aligned} T_U\mathrm{SpSt}(2n, 2k) &= \left\{ UA + U^s B \in \mathbb{R}^{2n\times 2k} \;\middle|\; A \in \mathfrak{sp}(2k, \mathbb{R}), B \in \mathbb{R}^{2(n-k)\times 2k} \right\} \\ &= \left\{ \Delta \in \mathbb{R}^{2n\times 2k} \;\middle|\; U^+\Delta \in \mathfrak{sp}(2k, \mathbb{R}) \right\}. \end{aligned} \tag{4.3.3}$$

Proof. For every tangent vector $\Delta \in T_U\mathrm{SpSt}(2n, 2k)$, there is a curve $\gamma\colon (-\varepsilon, \varepsilon) \to \mathrm{SpSt}(2n, 2k)$, with $\gamma(0) = U$ and $\dot\gamma(0) = \Delta$. Since $\gamma(t)^+\gamma(t) = I_{2k}$, differentiating and evaluating at $t = 0$ leads to $U^+\Delta = -\Delta^+U$.
Therefore $(U^+\Delta)^+ = \Delta^+U = -U^+\Delta$, and $A := U^+\Delta \in \mathfrak{sp}(2k, \mathbb{R})$. Furthermore,

$$\Delta = UU^+\Delta + (I_{2n} - UU^+)\Delta = UA + U^s(U^s)^+\Delta.$$

Counting dimensions and defining $B := (U^s)^+\Delta$ yields the result. □

Note that the tangent space parametrization [Gao+21b, (3.8b)] is similar to (4.3.3), but the chosen complement there is not necessarily a symplectic complement.

4.3.1 Pseudo-Riemannian metric on $\mathrm{SpSt}(2n, 2k)$

According to our quotient Lie group approach, $\mathrm{SpSt}(2n, 2k)$ inherits a pseudo-Riemannian metric from the pseudo-Riemannian metric (4.2.1) on the total space $\mathrm{Sp}(2n, \mathbb{R})$ in a natural way, by making use of horizontal lifts. A big advantage of this construction is that the corresponding geodesics can then be obtained via the projection of *horizontal geodesics*, i.e., geodesics with horizontal tangent vectors on $\mathrm{Sp}(2n, \mathbb{R})$ [ONe83, Corollary 7.46].

Splitting the Lie algebra $\mathfrak{sp}(2n, \mathbb{R})$ into a vertical and horizontal part with respect to the projection (4.3.2) and the pseudo-Riemannian metric h from (4.2.1) gives

$$\mathfrak{sp}(2n, \mathbb{R}) = \mathsf{Ver}^{\pi}\, \mathfrak{sp}(2n, \mathbb{R}) \oplus \mathsf{Hor}^{\pi,h}\, \mathfrak{sp}(2n, \mathbb{R}), \tag{4.3.4}$$

with *vertical space*

$$\mathsf{Ver}^{\pi}\, \mathfrak{sp}(2n, \mathbb{R}) := \ker \mathrm{d}\pi_E = \left\{ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & A & 0 & B \\ 0 & 0 & 0 & 0 \\ 0 & C & 0 & -A^T \end{bmatrix} \;\middle|\; \begin{array}{c} A \in \mathbb{R}^{(n-k)\times(n-k)} \\ B, C \in \mathrm{Sym}_{n-k} \end{array} \right\},$$

and *horizontal space*

$$\begin{aligned} \mathsf{Hor}^{\pi,h}\, \mathfrak{sp}(2n, \mathbb{R}) :=& (\mathsf{Ver}^{\pi}\, \mathfrak{sp}(2n, \mathbb{R}))^{\perp,h} \subset \mathfrak{sp}(2n, \mathbb{R}) \\ =& \left\{ \begin{bmatrix} A_1 & A_2^T & B_1 & B_2^T \\ A_3 & 0 & B_2 & 0 \\ C_1 & C_2^T & -A_1^T & -A_3^T \\ C_2 & 0 & -A_2 & 0 \end{bmatrix} \;\middle|\; \begin{array}{c} A_1 \in \mathbb{R}^{k\times k} \\ B_1, C_1 \in \mathrm{Sym}_k, \\ A_2, A_3, B_2, C_2 \in \mathbb{R}^{(n-k)\times k} \end{array} \right\}, \end{aligned}$$

where the orthogonal complement is taken with respect to the pseudo-Riemannian metric h of (4.2.1).

Proposition 4.3.3. For any $\Omega \in \mathfrak{sp}(2n, \mathbb{R})$, it holds that $\Omega \in \mathsf{Hor}^{\pi,h}\, \mathfrak{sp}(2n, \mathbb{R})$ if and only if

$$\begin{aligned} \Omega &= \left(I_{2n} - \frac{1}{2}EE^+\right)\Omega EE^+ - EE^+\Omega^+\left(I_{2n} - \frac{1}{2}EE^+\right) \\ &= \Omega EE^+ + EE^+\Omega - EE^+\Omega EE^+. \end{aligned} \tag{4.3.5}$$

Proof. Follows by a straightforward calculation. □

Eventually, we will exploit abstract results of semi-Riemannian geometry [ONe83, §11] for determining the geodesics. To enable this, we show next that

$$\mathrm{SpSt}(2n, 2k) \cong \mathrm{Sp}(2n, \mathbb{R}) / \mathrm{stab}_E$$

is *reductive*, and *naturally reductive* with respect to the pseudo-Riemannian metric h of (4.2.1), see [ONe83, Definition 11.21 & 11.23] for an explanation of these terms.

Lemma 4.3.4. The real symplectic Stiefel manifold $\mathrm{SpSt}(2n, 2k)$ is *reductive*. With respect to the pseudo-Riemannian metric h, it is *naturally reductive*.

Proof. In view of (4.3.4) and the fact that $\mathsf{Ver}^{\pi}\,\mathfrak{sp}(2n,\mathbb{R})$ is isomorphic to the Lie algebra of $\mathrm{Sp}(2(n-k),\mathbb{R}) \cong \mathrm{stab}_E$, we need to show that the complementary subspace $\mathsf{Hor}^{\pi,h}\,\mathfrak{sp}(2n,\mathbb{R})$ is $\mathrm{Ad}(\mathrm{stab}_E)$-invariant in order to establish reductiveness, where $\mathrm{Ad}_M(\Omega) = M\Omega M^+$, see [ONe83, §11, p. 303]. In fact, for every $M \in \mathrm{stab}_E$ (and therefore $M^+ \in \mathrm{stab}_E$) and every $\Omega \in \mathsf{Hor}^{\pi,h}\,\mathfrak{sp}(2n,\mathbb{R})$, it holds that

$$\begin{aligned}\mathrm{Ad}_M(\Omega) = M\Omega M^+ &= M\Omega EE^+M^+ + MEE^+\Omega M^+ - MEE^+\Omega EE^+M^+ \\ &= M\Omega M^+EE^+ + EE^+M\Omega M^+ - EE^+M\Omega M^+EE^+.\end{aligned}$$

By Proposition (4.3.3), it follows that $\mathrm{Ad}_M(\Omega) \in \mathsf{Hor}^{\pi,h}\,\mathfrak{sp}(2n,\mathbb{R})$, which means that $\mathsf{Hor}^{\pi,h}\,\mathfrak{sp}(2n,\mathbb{R})$ is $\mathrm{Ad}(\mathrm{stab}_E)$-invariant. Therefore, $\mathrm{SpSt}(2n, 2k)$ is reductive.

The fact that $\mathrm{SpSt}(2n, 2k)$ is naturally reductive with respect to h follows from a direct calculation, by making use of the fact that the projection from $\mathfrak{sp}(2n,\mathbb{R})$ onto $\mathsf{Hor}^{\pi,h}\,\mathfrak{sp}(2n,\mathbb{R})$ is given by $\Omega \mapsto \Omega - (I_{2n} - EE^+)\Omega(I_{2n} - EE^+)$. □

Recall that the Lie algebra $\mathfrak{sp}(2n,\mathbb{R}) = T_I\mathrm{Sp}(2n,\mathbb{R})$ is the tangent space at the identity. By left translation, every tangent space $T_M\mathrm{Sp}(2n,\mathbb{R})$ can be split into a vertical and horizontal part,

$$\begin{aligned}\mathsf{Ver}^{\pi}_M\,\mathrm{Sp}(2n,\mathbb{R}) &:= \{M\Omega \mid \Omega \in \mathsf{Ver}^{\pi}\,\mathfrak{sp}(2n,\mathbb{R})\} \\ \mathsf{Hor}^{\pi,h}_M\,\mathrm{Sp}(2n,\mathbb{R}) &:= \left\{M\Omega \;\middle|\; \Omega \in \mathsf{Hor}^{\pi,h}\,\mathfrak{sp}(2n,\mathbb{R})\right\}.\end{aligned} \tag{4.3.6}$$

The horizontal space at $M \in \mathrm{Sp}(2n,\mathbb{R})$, i.e., $\mathsf{Hor}^{\pi,h}_M\,\mathrm{Sp}(2n,\mathbb{R})$, is isomorphic to the tangent space $T_{\pi(M)}\mathrm{SpSt}(2n, 2k)$. For $M\Omega \in \mathsf{Hor}^{\pi,h}_M\,\mathrm{Sp}(2n,\mathbb{R})$, it holds that

$$\mathrm{d}\pi_M(M\Omega) = M\Omega E = M\begin{bmatrix} A_1 & A_2^T & B_1 & B_2^T \\ A_3 & 0 & B_2 & 0 \\ C_1 & C_2^T & -A_1^T & -A_3^T \\ C_2 & 0 & -A_2 & 0\end{bmatrix}E = M\begin{bmatrix} A_1 & B_1 \\ A_3 & B_2 \\ C_1 & -A_1^T \\ C_2 & -A_2\end{bmatrix} = UA + U^sB,$$

where $A = \begin{bmatrix} A_1 & B_1 \\ C_1 & -A_1^T\end{bmatrix}$, $B = \begin{bmatrix} A_3 & B_2 \\ C_2 & -A_2\end{bmatrix}$ and $U = ME$. It follows therefore that the application of $\mathrm{d}\pi_M$ to $\mathsf{Hor}^{\pi,h}_M\,\mathrm{Sp}(2n,\mathbb{R})$ gives (4.3.3). This implies that $T_U\mathrm{SpSt}(2n, 2k)$ can also be parameterized as

$$T_U\mathrm{SpSt}(2n, 2k) = \left\{M\Omega E \;\middle|\; M \in \pi^{-1}(U),\ \Omega \in \mathsf{Hor}^{\pi,h}\,\mathfrak{sp}(2n,\mathbb{R})\right\}.$$

Proposition 4.3.5 (Alternative tangent vector parameterization)**.** Every tangent vector $\Delta \in T_U\mathrm{SpSt}(2n, 2k)$ is of the form $\Delta = \tilde{\Omega}U$, where

$$\tilde{\Omega} = M\Omega M^+ \in \mathfrak{sp}(2n,\mathbb{R}),$$

with $\Omega \in \mathsf{Hor}^{\pi,h}\, \mathfrak{sp}(2n,\mathbb{R})$, is unique. It can be calculated via

$$\tilde{\Omega}(U,\Delta) = \left(I_{2n} - \frac{1}{2}UU^+\right)\Delta U^+ - U\Delta^+\left(I_{2n} - \frac{1}{2}UU^+\right). \tag{4.3.7}$$

Proof. This can be seen by making use of $U = ME$ and Proposition 4.3.3. □

Equation (4.3.7) corresponds to $S_{X,Y}J$ from [Gao+21b, Proposition 4.3].

Via horizontal lifts, a pseudo-Riemannian metric on the real symplectic Stiefel manifold can be defined as follows: For two tangent vectors $\Delta_1, \Delta_2 \in T_U\mathrm{SpSt}(2n,2k)$ and $U = ME$, calculate $\tilde{\Omega}(U,\Delta_i)$ according to (4.3.7), $i = 1,2$. The horizontal lift to $\mathsf{Hor}^{\pi,h}_M\, \mathrm{Sp}(2n,\mathbb{R})$ is then given by

$$(\Delta_i)^{\mathsf{hor}}_M = M\Omega_i = \tilde{\Omega}(U,\Delta_i)M,$$

where $\Omega_i = M^+\tilde{\Omega}(U,\Delta_i)M$. This follows from

$$\mathrm{d}\pi_M((\Delta_i)^{\mathsf{hor}}_M) = (\Delta_i)^{\mathsf{hor}}_M E = \tilde{\Omega}(U,\Delta_i)ME = \tilde{\Omega}(U,\Delta_i)U = \Delta_i,$$

and the fact that Ω_i fulfills Proposition 4.3.3. In the following, we exploit that pseudo-Riemannian submersions [ONe83, Definition 7.44] are particularly useful to find the geodesics in a quotient space, given that the geodesics in the associated total space are known, see [ONe83, Corollary 7.46].

Proposition 4.3.6. Let $\Delta_i = UA_i + U^sB_i \in T_U\mathrm{SpSt}(2n,2k)$, $i = 1,2$. A pseudo-Riemannian metric is defined by $h^{\mathrm{SpSt}}_U \colon T_U\mathrm{SpSt}(2n,2k) \times T_U\mathrm{SpSt}(2n,2k) \to \mathbb{R}$,

$$\begin{aligned} h^{\mathrm{SpSt}}_U(\Delta_1,\Delta_2) &:= \langle \Delta_1,\Delta_2\rangle_U := \left\langle (\Delta_2)^{\mathsf{hor}}_M, (\Delta_2)^{\mathsf{hor}}_M \right\rangle_M \\ &= \mathrm{tr}\left(\Delta_1^+\left(I_{2n} - \frac{1}{2}UU^+\right)\Delta_2\right) = \frac{1}{2}\,\mathrm{tr}(A_1^+A_2) + \mathrm{tr}(B_1^+B_2). \end{aligned} \tag{4.3.8}$$

With respect to the metric h^{SpSt}_U, π is a pseudo-Riemannian submersion.

Proof. A direct calculation shows the identities of (4.3.8). Since h is a pseudo-Riemannian metric, h^{SpSt}_U is a pseudo-Riemannian metric as well. The projection π is then a pseudo-Riemannian submersion by Lemma 4.3.4 and [ONe83, Lemma 11.24]. □

The geodesics, calculated via the exponential mapping with respect to h^{SpSt}_U, can be found via the projection of the exponential mapping in $\mathrm{Sp}(2n,\mathbb{R})$.

Proposition 4.3.7. Let $U \in \mathrm{SpSt}(2n,2k)$ and $M \in \pi^{-1}(U) \subset \mathrm{Sp}(2n,\mathbb{R})$. Furthermore, let $\Delta = M\Omega \in T_U\mathrm{SpSt}(2n,2k)$. The geodesic γ with respect to the pseudo-Riemannian metric (4.3.8) that starts from $\gamma(0) = U$ in direction $\dot{\gamma}(0) = \Delta$

is

$$\gamma(t) := \mathrm{Exp}_U^{\mathrm{SpSt},h}(t\Delta) := \pi(\mathrm{Exp}_M^{\mathrm{Sp},h}(t\Delta_M^{\mathsf{hor}})) \\ = M\exp_{\mathrm{m}}(t\Omega)E = \exp_{\mathrm{m}}(t\tilde{\Omega}(U,\Delta))U, \tag{4.3.9}$$

with $\tilde{\Omega}(U,\Delta) = M\Omega M^+$ from (4.3.7).

Proof. By [ONe83, Corollary 7.46], horizontal geodesics in $\mathrm{Sp}(2n,\mathbb{R})$ are mapped to geodesics in the quotient $\mathrm{SpSt}(2n,2k)$ under the pseudo-Riemannian submersion π. The facts that $\gamma(0) = U$ and $\dot{\gamma}(0) = M\Omega = \Delta$ are immediate. □

In the form of (4.3.9), the exponential mapping depends on the matrix exponential of a $2n \times 2n$ matrix. For tangent vectors $\Delta \in T_U\mathrm{SpSt}(2n,2k)$ with $\Delta^+(I_{2n} - UU^+)\Delta$ invertible, we can reduce the computational complexity to $4k \times 4k$. This is rendered possible by the fact that for $X, Y \in \mathbb{R}^{n\times k}$ with $Y^TX \in \mathbb{R}^{k\times k}$ non-singular, we have from [CI00, Prop. 3] that

$$\exp_{\mathrm{m}}(XY^T) = I_n + X(\exp_{\mathrm{m}}(Y^TX) - I_k)(Y^TX)^{-1}Y^T. \tag{4.3.10}$$

Proposition 4.3.8. Let $U \in \mathrm{SpSt}(2n,2k)$ and $\Delta \in T_U\mathrm{SpSt}(2n,2k)$. Define $A = U^+\Delta$ and $H = \Delta - UA$. If H^+H is invertible, then the geodesic from U in direction Δ is given by

$$\mathrm{Exp}_U^{\mathrm{SpSt},h}(t\Delta) = \begin{bmatrix} U & \tfrac{1}{2}UA + H \end{bmatrix} \exp_{\mathrm{m}}\left(t\begin{bmatrix} \tfrac{1}{2}A & \tfrac{1}{4}A^2 - H^+H \\ I_{2k} & \tfrac{1}{2}A \end{bmatrix}\right)\begin{bmatrix} I_{2k} \\ 0 \end{bmatrix}. \tag{4.3.11}$$

Proof. We start from (4.3.9) with $\tilde{\Omega}(U,\Delta)$ according to (4.3.7). Introducing

$$X = \begin{bmatrix} (I_{2n} - \tfrac{1}{2}UU^+)\Delta & -U \end{bmatrix} = \begin{bmatrix} \tfrac{1}{2}UA + H & -U \end{bmatrix} \in \mathbb{R}^{2n\times 4k}$$

and $Y^T = \begin{bmatrix} U^+ \\ \Delta^+(I_{2n} - \tfrac{1}{2}UU^+) \end{bmatrix} \in \mathbb{R}^{4k\times 2n}$, we have $\tilde{\Omega}(U,\Delta) = XY^T$. Furthermore

$$Y^TX = \begin{bmatrix} \tfrac{1}{2}U^+\Delta & -I_{2k} \\ \Delta^+(I_{2n} - \tfrac{3}{4}UU^+)\Delta & -\tfrac{1}{2}\Delta^+U \end{bmatrix} = \begin{bmatrix} \tfrac{1}{2}A & -I_{2k} \\ H^+H - \tfrac{1}{4}A^2 & \tfrac{1}{2}A \end{bmatrix}.$$

If $(Y^TX)^{-1}$ exists, it is given by

$$(Y^TX)^{-1} = \begin{bmatrix} \tfrac{1}{2}(H^+H)^{-1}A & (H^+H)^{-1} \\ \tfrac{1}{4}A(H^+H)^{-1}A - I_{2k} & \tfrac{1}{2}A(H^+H)^{-1} \end{bmatrix},$$

and therefore $Y^T X$ is invertible if and only if H^+H is invertible. It furthermore holds that $(Y^TX)^{-1}Y^TU = (Y^TX)^{-1}Y^TX\left[\begin{smallmatrix}0\\-I_{2k}\end{smallmatrix}\right] = \left[\begin{smallmatrix}0\\-I_{2k}\end{smallmatrix}\right]$. By (4.3.9) it holds that $\mathrm{Exp}_U^{\mathrm{SpSt},h}(t\Delta) = \exp_{\mathrm{m}}(t\tilde{\Omega})U = \exp_{\mathrm{m}}(tXY^T)U$. Now by (4.3.10), it holds for $t \neq 0$ that

$$\begin{aligned}\exp_{\mathrm{m}}(tXY^T)U &= (I_{2n} + tX(\exp_{\mathrm{m}}(tY^TX) - I_{2k})(tY^TX)^{-1}Y^T)U\\ &= U + X(\exp_{\mathrm{m}}(tY^TX) - I_{2k})(Y^TX)^{-1}Y^TU\\ &= U + X\exp_{\mathrm{m}}(tY^TX)\begin{bmatrix}0\\-I_{2k}\end{bmatrix} - X\begin{bmatrix}0\\-I_{2k}\end{bmatrix}\\ &= X\exp_{\mathrm{m}}(tY^TX)\begin{bmatrix}0\\-I_{2k}\end{bmatrix}.\end{aligned}$$

Moreover, $\lim_{t\to 0} X\exp_{\mathrm{m}}(tY^TX)\left[\begin{smallmatrix}0\\-I_{2k}\end{smallmatrix}\right] = U = \mathrm{Exp}_U^{\mathrm{SpSt},h}(0\cdot\Delta)$.
The form (4.3.11) is obtained by

$$X\exp_{\mathrm{m}}(tY^TX)\begin{bmatrix}0\\-I_{2k}\end{bmatrix} = XJ_{4k}^T\exp_{\mathrm{m}}(tJ_{4k}Y^TXJ_{4k}^T)J_{4k}\begin{bmatrix}0\\-I_{2k}\end{bmatrix}.$$

□

Note that for the calculation of (4.3.11) we don't need the invertibility of H^+H, and one can check that the right hand side is always an element in the symplectic Stiefel manifold $\mathrm{SpSt}(2n,2k)$. We can therefore always apply (4.3.11) to calculate a curve.

Remark. The simplified formula for the symplectic Stiefel exponential (4.3.11) is similar to the so-called quasigeodesic retraction defined in [Gao+21b, Lemma 5.1], which in our notation is given as

$$\mathcal{R}_U^{\mathrm{qgeo}}(\Delta) = \begin{bmatrix}U & \Delta\end{bmatrix}\exp_{\mathrm{m}}\left(\begin{bmatrix}U^+\Delta & -\Delta^+\Delta\\ I_{2k} & U^+\Delta\end{bmatrix}\right)\begin{bmatrix}I_{2k}\\0\end{bmatrix}\exp_{\mathrm{m}}(-U^+\Delta). \tag{4.3.12}$$

The two curves are however not identical. Note also the structural similarity with the formula for the Euclidean Stiefel geodesics of [EAS98, Section 2.2.2].

4.3.2 Right-invariant Riemannian metric on $\mathrm{SpSt}(2n,2k)$

The real symplectic Stiefel manifold may be equipped with different Riemannian metrics. The so-called canonical-like metric has been studied in [Gao+21b], while [Gao+21a] considers a restriction of the Euclidean metric. To the best of the authors' knowledge, the geodesics for these metrics are unknown. Complementary to the aforementioned Riemannian metrics, we use the Riemannian metric g_M of (4.2.2) on $\mathrm{Sp}(2n,\mathbb{R})$ to introduce a third Riemannian metric on $\mathrm{SpSt}(2n,2k)$ via a horizontal lift, which allows us to find the corresponding geodesics. This metric is invariant under the group action of $\mathrm{Sp}(2k,\mathbb{R})$ from the right and induces therefore a Riemannian metric on the symplectic subspaces that will be considered in Subsection 4.4.2.

We begin by splitting the tangent space $T_M\mathrm{Sp}(2n,\mathbb{R})$ at $M \in \mathrm{Sp}(2n,\mathbb{R})$ into a vertical and a horizontal part with respect to g_M and the projection π from (4.3.2),

$$T_M\mathrm{Sp}(2n,\mathbb{R}) = \mathsf{Ver}_M^\pi \mathrm{Sp}(2n,\mathbb{R}) \oplus \mathsf{Hor}_M^{\pi;g} \mathrm{Sp}(2n,\mathbb{R}).$$

As the vertical part is defined as the kernel of $\mathrm{d}\pi_M$, it is the same as in (4.3.6). The horizontal part, however, is different, since it is now given as the orthogonal complement of $\mathsf{Ver}_M^\pi \mathrm{Sp}(2n,\mathbb{R})$ with respect to the metric g_M of (4.2.2). This yields

$$\mathsf{Hor}_M^{\pi;g} \mathrm{Sp}(2n,\mathbb{R}) = \left\{ \bar{\Omega} M \;\middle|\; \bar{\Omega} = \bar{\Omega} P + P\bar{\Omega} - P\bar{\Omega}P \in \mathfrak{sp}(2n,\mathbb{R}) \right\}. \tag{4.3.13}$$

Here, $P = J_{2n}^T U U^+ J_{2n}$ and $U = \pi(M) = ME$. Equation (4.3.13) can be established as follows: Any horizontal tangent vector $X = \bar{\Omega} M$, with $\bar{\Omega} \in \mathfrak{sp}(2n,\mathbb{R})$, fulfills for all $Y = M\Omega \in \mathsf{Ver}_M^\pi \mathrm{Sp}(2n,\mathbb{R})$

$$0 = g_M(X,Y) = \frac{1}{2}\operatorname{tr}(\bar{\Omega}^T M\Omega M^+) = \frac{1}{2}\operatorname{tr}(M^+ J_{2n}\bar{\Omega} J_{2n} M\Omega).$$

Define $\hat{\Omega} := M^+ J_{2n} \bar{\Omega} J_{2n} M$. Then $0 = \frac{1}{2}\operatorname{tr}(\hat{\Omega}\Omega)$ for all $\Omega \in \mathsf{Ver}^\pi \mathfrak{sp}(2n,\mathbb{R})$ implies that $\hat{\Omega}$ fulfills (4.3.5). By making use of $E = M^+ME = M^+U$, it follows that

$$\begin{aligned}
\bar{\Omega} &= J_{2n} M \hat{\Omega} M^+ J_{2n} \\
&= J_{2n} M(\hat{\Omega} EE^+ + EE^+\hat{\Omega} - EE^+\hat{\Omega}EE^+) M^+ J_{2n} \\
&= \bar{\Omega} J_{2n} UU^+ J_{2n}^T + J_{2n} UU^+ J_{2n}^T \bar{\Omega} - J_{2n} UU^+ J_{2n}^T \bar{\Omega} J_{2n} UU^+ J_{2n}^T.
\end{aligned}$$

Conversely, if $\bar{\Omega}$ fulfills the above equation, then it follows that $\hat{\Omega} := M^+ J_{2n} \bar{\Omega} J_{2n} M$ fulfills (4.3.5) and therefore $\bar{\Omega} M \in \mathsf{Hor}_M^{\pi;g} \mathrm{Sp}(2n,\mathbb{R})$.

As usual, for $U = \pi(M)$ we can identify the tangent space $T_U\mathrm{SpSt}(2n,2k)$ with the horizontal space $\mathsf{Hor}_M^{\pi;g} \mathrm{Sp}(2n,\mathbb{R})$. Any $\Delta \in T_U\mathrm{SpSt}(2n,2k)$ is of the form $\Delta = \pi(\bar{\Omega}M) = \bar{\Omega}U$ for some $\bar{\Omega}M \in \mathsf{Hor}_M^{\pi;g} \mathrm{Sp}(2n,\mathbb{R})$, and we can find the horizontal lift

$$\Delta_M^{\mathsf{hor},g} = \bar{\Omega}(\Delta) M \tag{4.3.14}$$

via

$$\bar{\Omega}(\Delta) = \Delta(U^TU)^{-1}U^T + J_{2n}U(U^TU)^{-1}\Delta^T(I_{2n} - J_{2n}^T U(U^TU)^{-1}U^T J_{2n})J_{2n}. \tag{4.3.15}$$

This follows from the facts that

1. $\bar{\Omega}(\Delta)^+ = -\bar{\Omega}(\Delta)$, so $\bar{\Omega}(\Delta) \in \mathfrak{sp}(2n,\mathbb{R})$,
2. $\bar{\Omega}(\Delta)U = \Delta$, and
3. $\bar{\Omega}(\Delta) = \bar{\Omega}(\Delta)P + P\bar{\Omega}(\Delta) - P\bar{\Omega}(\Delta)P$, with $P = J_{2n}UU^+J_{2n}^T$,

where the last equation follows from a straightforward calculation.

The right action of $\mathrm{Sp}(2(n-k),\mathbb{R}) \cong \mathrm{stab}_E$ on $\mathrm{Sp}(2n,\mathbb{R})$ is vertical, i.e., $\pi(MN) = \pi(M)$ for all $N \in \mathrm{stab}_E$. The action is also transitive on fibers, i.e., for $M, M' \in \mathrm{Sp}(2n,\mathbb{R})$ with $\pi(M) = U = \pi(M')$ it holds that $MM^+M' = M'$ and $M^+M' \in \mathrm{stab}_E$. It is furthermore isometric, by right-invariance of (4.2.2). From [Lee18, Theorem 2.28], it follows that there is a unique Riemannian metric on $\mathrm{SpSt}(2n,2k)$ such that π is a Riemannian submersion. This Riemannian metric is given via the horizontal lift.

Proposition 4.3.9. The Riemannian metric on $\mathrm{SpSt}(2n,2k)$, for which π is a Riemannian submersion, is right-invariant and given point-wise by

$$\begin{aligned} g_U^{\mathrm{SpSt}}&\colon T_U\mathrm{SpSt}(2n,2k) \times T_U\mathrm{SpSt}(2n,2k) \to \mathbb{R},\\ g_U^{\mathrm{SpSt}}(\Delta_1,\Delta_2) &:= g_M((\Delta_1)_M^{\mathsf{hor},g}, (\Delta_2)_M^{\mathsf{hor},g})\\ &= \mathrm{tr}\left(\Delta_1^T\left(I_{2n} - \frac{1}{2}J_{2n}^T U(U^TU)^{-1}U^TJ_{2n}\right)\Delta_2(U^TU)^{-1}\right). \end{aligned} \tag{4.3.16}$$

Proof. The Riemannian submersion property and right-invariance hold by the definition of g_U^{SpSt} via the horizontal lift. The second equality follows from the combination of (4.2.2), (4.3.14) and (4.3.15). □

The Riemannian gradient of a function $f\colon \mathrm{SpSt}(2n,2k) \to \mathbb{R}$ with respect to g^{SpSt} is given by

$$\mathrm{grad}_f^g(U) = \nabla f(U)U^TU + J_{2n}U(\nabla f(U))^TJ_{2n}U, \tag{4.3.17}$$

where $\nabla f(U)$ denotes the Euclidean gradient of a smooth extension of f around $U \in \mathrm{SpSt}(2n,2k)$ in $\mathbb{R}^{2n\times 2k}$ at U. This holds because $U^+\,\mathrm{grad}_f^g(U) = -(\mathrm{grad}_f^g(U))^+U$, which implies $\mathrm{grad}_f^g(U) \in T_U\mathrm{SpSt}(2n,2k)$, and because $\mathrm{grad}_f^g(U)$ solves

$$g_U^{\mathrm{SpSt}}(\mathrm{grad}_f^g(U),\Delta) = \mathrm{d}f_U(\Delta) = \mathrm{tr}((\nabla f(U))^T\Delta)$$

for all $\Delta \in T_U\mathrm{SpSt}(2n,2k)$.

Riemannian geodesics on $\mathrm{Sp}(2n,\mathbb{R})$ with a horizontal tangent vector at every point project to Riemannian geodesics on $\mathrm{SpSt}(2n,2k)$ by [ONe83, Corollary 7.46]. (Mind that the referenced result is stated in the pseudo-Riemannian setting, but also holds true in the Riemannian case.) We show that Riemannian geodesics on $\mathrm{Sp}(2n,\mathbb{R})$ with initial horizontal tangent vector have a horizontal tangent vector throughout.

Lemma 4.3.10. Let $M \in \mathrm{Sp}(2n,\mathbb{R})$ and $X \in \mathsf{Hor}_M^{\pi,g}\,\mathrm{Sp}(2n,\mathbb{R})$. Define $\gamma(t) := \mathrm{Exp}_M^{\mathrm{Sp},g}(tX)$. Then $\dot\gamma(t) \in \mathsf{Hor}_{\gamma(t)}^{\pi,g}\,\mathrm{Sp}(2n,\mathbb{R})$.

Proof. Let $U = \pi(M)$ and $U(t) = \pi(\gamma(t)) = \gamma(t)E$. Furthermore, let

$$P(t) := J_{2n}U(t)U(t)^+J_{2n}^T.$$

By the structure of the horizontal space, it holds $X = \bar{\Omega} M$. Define $x(t) := \dot{\gamma}(t)\gamma(t) \in \mathfrak{sp}(2n, \mathbb{R})$. Then, by (4.3.13), $\dot{\gamma}(t) \in \mathsf{Hor}^{\pi,g}_{\gamma(t)} \operatorname{Sp}(2n,\mathbb{R})$ is equivalent to

$$x(t) = P(t)x(t) + x(t)P(t) - P(t)x(t)P(t).$$

With

- $x(t) = \exp_{\mathrm{m}}(t(\bar{\Omega} - \bar{\Omega}^T))\bar{\Omega}\exp_{\mathrm{m}}(-t(\bar{\Omega} - \bar{\Omega}^T))$,
- $U(t) = \exp_{\mathrm{m}}(t(\bar{\Omega} - \bar{\Omega}^T))\exp_{\mathrm{m}}(t\bar{\Omega}^T)U$,
- $J_{2n}^T \exp_{\mathrm{m}}(t(\bar{\Omega} - \bar{\Omega}^T))J_{2n} = \exp_{\mathrm{m}}(t(\bar{\Omega} - \bar{\Omega}^T))$,
- $J_{2n}\exp_{\mathrm{m}}(t\bar{\Omega})J_{2n} = \exp_{\mathrm{m}}(-t\bar{\Omega})$ and
- $P(t) = \exp_{\mathrm{m}}(t(\bar{\Omega} - \bar{\Omega}^T))\exp_{\mathrm{m}}(-t\bar{\Omega})P(0)\exp_{\mathrm{m}}(t\bar{\Omega})\exp_{\mathrm{m}}(-t(\bar{\Omega} - \bar{\Omega}^T))$,

the claim follows by a straightforward calculation. □

We are now ready to state the Riemannian geodesics on $\operatorname{SpSt}(2n, 2k)$ with respect to the Riemannian metric g^{SpSt} from (4.3.16).

Proposition 4.3.11. Let $U \in \operatorname{SpSt}(2n, 2k)$ and $\Delta \in T_U\operatorname{SpSt}(2n, 2k)$. Let $M \in \pi^{-1}(U) \subset \operatorname{Sp}(2n, \mathbb{R})$. Then the geodesic from U in direction Δ is given by

$$\begin{aligned} \operatorname{Exp}^{\operatorname{SpSt},g}_U(t\Delta) &:= \pi(\operatorname{Exp}^{\operatorname{Sp},g}_M(t\Delta^{\mathsf{hor},g}_M)) \\ &= \exp_{\mathrm{m}}(t(\bar{\Omega}(\Delta) - \bar{\Omega}(\Delta)^T))\exp_{\mathrm{m}}(t\bar{\Omega}(\Delta)^T)U \end{aligned} \tag{4.3.18}$$

with $\bar{\Omega}(\Delta)$ as in (4.3.15).

Proof. This follows directly from the preceding discussion and the definition of the horizontal lift. □

Equation (4.3.18) is formulated with $2n \times 2n$-matrices, but may in practical calculations be reduced to work with tall, skinny $2n \times 8k$ matrices and matrix exponentials of an $8k \times 8k$ and a $4k \times 4k$ matrix, respectively. To this end, define $\bar{A} \in \mathfrak{so}(2k, \mathbb{R})$ by

$$\bar{A} := J_{2k}U^T\Delta(U^TU)^{-1}J_{2k} + (U^TU)^{-1}\Delta^TU - (U^TU)^{-1}\Delta^TJ_{2n}^TU(U^TU)^{-1}J_{2k}$$

and define

$$\bar{H} := (I_{2n} - UU^+)J_{2n}\Delta(U^TU)^{-1}J_{2k}.$$

With $\bar{\Delta} := U\bar{A} + \bar{H} \in T_U\operatorname{SpSt}(2n, 2k)$ it holds that $\bar{\Omega}(\Delta) = YX^T$, where

$$X := \begin{bmatrix}(I - \tfrac{1}{2}UU^+)\bar{\Delta} & -U\end{bmatrix} \in \mathbb{R}^{2n \times 4k}$$

and

$$Y := \begin{bmatrix} J_{2n}^T U J_{2k} & (\bar{\Delta}^+(I_{2n} - \frac{1}{2}UU^+))^T \end{bmatrix} \in \mathbb{R}^{2n\times 4k}.$$

This follows from (4.3.15) and solving $\bar{\Omega}(\Delta)^T = (I_{2n} - \frac{1}{2}UU^+)\bar{\Delta}U^+ - U\bar{\Delta}^+(I_{2n} - \frac{1}{2}UU^+)$ for $\bar{\Delta}$, i.e., $\bar{A} = U^+\bar{\Omega}(\Delta)^T U$ and $\bar{H} = (I_{2n} - UU^+)\bar{\Omega}(\Delta)^T U$. Furthermore, define $\hat{X} := \begin{bmatrix} Y & -X \end{bmatrix} \in \mathbb{R}^{2n\times 8k}$ and $\hat{Y} := \begin{bmatrix} X & Y \end{bmatrix} \in \mathbb{R}^{2n\times 8k}$.

Proposition 4.3.12. With notation as above, it holds that

$$\mathrm{Exp}_U^{\mathrm{SpSt},g}(\Delta) = \hat{X}\exp_{\mathrm{m}}(\hat{Y}^T\hat{X})\begin{bmatrix} 0_{4k} \\ I_{4k} \end{bmatrix}\exp_{\mathrm{m}}\left(Y^T X\right)\begin{bmatrix} 0_{2k} \\ I_{2k} \end{bmatrix}. \tag{4.3.19}$$

Proof. First, note that since $\bar{\Omega}(\Delta) = YX^T$, it holds that $\exp_{\mathrm{m}}(\bar{\Omega}(\Delta)^T)U = \exp_{\mathrm{m}}(XY^T)U$. We make use of (4.3.10), which implies

$$\exp_{\mathrm{m}}(XY^T)U = U + X(\exp_{\mathrm{m}}(Y^TX) - I_k)(Y^TX)^{-1}Y^TU.$$

Since $(Y^TX)^{-1}Y^TU = (Y^TX)^{-1}Y^TX\left[\begin{smallmatrix} 0_{2k} \\ -I_{2k} \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0_{2k} \\ -I_{2k} \end{smallmatrix}\right]$, the simplified expression $\exp_{\mathrm{m}}(XY^T)U = X\exp_{\mathrm{m}}(Y^TX)\left[\begin{smallmatrix} 0_{2k} \\ -I_{2k} \end{smallmatrix}\right]$ follows.

Secondly, it holds that $\exp_{\mathrm{m}}(\bar{\Omega}(\Delta) - \bar{\Omega}(\Delta)^T) = \exp_{\mathrm{m}}(\hat{X}\hat{Y}^T)$. Repeating the steps above and noting $(\hat{Y}^T\hat{X})^{-1}\hat{Y}^TX = \left[\begin{smallmatrix} 0_{4k} \\ -I_{4k} \end{smallmatrix}\right]$ leads to the claimed result. □

4.4 The real symplectic Grassmann manifold

Similar to the usual Grassmann manifold [Bat+15; EAS98] of linear subspaces of a fixed dimension, we define the *real symplectic Grassmann manifold* as the manifold of symplectic subspaces of dimension $2k$ of $(\mathbb{R}^{2n}, \omega_0)$. This must not be confused with the Lagrangian Grassmannian, the manifold of Lagrangian subspaces, which is also referred to as the symplectic Grassmann manifold by some authors. The quotient manifold approach which we use is similar to the course of action in [BZA20]. As in the case of linear subspaces, we identify a symplectic subspace with the associated symplectic projection onto it.

Proposition 4.4.1. The set

$$\mathrm{SpGr}(2n, 2k) := \{P \in \mathbb{R}^{2n\times 2n} \mid P^2 = P,\ \mathrm{rank}(P) = 2k,\ P^+ = P\} \tag{4.4.1}$$

consists of the symplectic projections onto the $2k$-dimensional symplectic subspaces of the standard symplectic space $(\mathbb{R}^{2n}, \omega_0)$. It has a smooth manifold structure and is called the *real symplectic Grassmann manifold*. It features the quotient representation

$$\mathrm{SpGr}(2n,2k) \cong \mathrm{Sp}(2n,\mathbb{R})/(\mathrm{Sp}(2k,\mathbb{R}) \times \mathrm{Sp}(2(n-k),\mathbb{R})) \tag{4.4.2}$$

and has dimension

$$\dim \mathrm{SpGr}(2n,2k) = 4(n-k)k.$$

Proof. We show first that the thus defined space $\mathrm{SpGr}(2n,2k)$ is the orbit of

$$E_0 := EE^+$$

under the group action of $\mathrm{Sp}(2n,\mathbb{R})$ defined by

$$\phi\colon \mathrm{Sp}(2n,\mathbb{R}) \times \mathbb{R}^{2n\times 2n},\ (M,X) \mapsto MXM^+. \tag{4.4.3}$$

Because every $U \in \mathrm{SpSt}(2n,2k)$ has a representation $U = ME$, $M \in \mathrm{Sp}(2n,\mathbb{R})$, it is sufficient to show that every $P \in \mathrm{SpGr}(2n,2k)$ is equal to $P = UU^+$ for some $U \in \mathrm{SpSt}(2n,2k)$. This fact is established as follows: Since $PJ_{2n}P^T$ is skew-symmetric, it features a 'Schur-like decomposition' [Xu03, eq. (5)] of the form

$$PJ_{2n}P^T = Q\begin{bmatrix} 0 & \Sigma^2 & 0 \\ -\Sigma^2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}Q^T,$$

where $Q \in \mathrm{O}(2n)$ is a real orthogonal matrix. Moreover $\Sigma = \mathrm{diag}(\sigma_1,\dots,\sigma_k)$, where $\sigma_i > 0$ for all $i = 1,\dots,k$, because $\mathrm{rank}(P) = 2k$ [Xu03, Proposition 3]. From $P^+ = P = P^2$, it follows that $P = Q\begin{bmatrix} 0 & \Sigma^2 & 0 \\ -\Sigma^2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}Q^T J_{2n}^T$. For $U := QI_{2n,2k}\left[\begin{smallmatrix} \Sigma & 0 \\ 0 & \Sigma \end{smallmatrix}\right] \in \mathbb{R}^{2n\times 2k}$, with $I_{2n,2k}$ as in (4.3.1), it furthermore holds that $P = UU^+$. The fact that $U^+U = I_{2k}$, i.e., $U \in \mathrm{SpSt}(2n,2k)$, follows from $P^2 = P$. The other inclusion, i.e., $\phi(M,E_0) \in \mathrm{SpGr}(2n,2k)$ for all $M \in \mathrm{Sp}(2n,\mathbb{R})$ is immediate. The stabilizer of the group action $\phi(\cdot,E_0)$ is given by

$$\begin{aligned}\mathrm{stab}_{E_0} &= \{M \in \mathrm{Sp}(2n,\mathbb{R}) \mid ME_0M^+ = E_0\} \\ &= \left\{ \begin{bmatrix} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{bmatrix} \in \mathrm{Sp}(2n,\mathbb{R}) \right\},\end{aligned}$$

where $\left[\begin{smallmatrix} A_1 & B_1 \\ C_1 & D_1 \end{smallmatrix}\right] \in \mathrm{Sp}(2k,\mathbb{R})$ and $\left[\begin{smallmatrix} A_2 & B_2 \\ C_2 & D_2 \end{smallmatrix}\right] \in \mathrm{Sp}(2(n-k),\mathbb{R})$. Hence,

$$\mathrm{stab}_{E_0} \cong \mathrm{Sp}(2k,\mathbb{R}) \times \mathrm{Sp}(2(n-k),\mathbb{R}).$$

The manifold structure now follows from [Lee12, Theorem 21.20]. The real symplectic Grassmann manifold is by [Lee12, Theorem 21.18] diffeomorphic to the homogeneous space

$$\mathrm{SpGr}(2n,2k) \cong \mathrm{Sp}(2n,\mathbb{R})/(\mathrm{Sp}(2k,\mathbb{R}) \times \mathrm{Sp}(2(n-k),\mathbb{R})).$$

The dimension of $\mathrm{SpGr}(2n, 2k)$ is obtained via the standard formula

$$\begin{aligned}\dim \mathrm{SpGr}(2n, 2k) &= \dim \mathrm{Sp}(2n, \mathbb{R}) - \dim \mathrm{Sp}(2k, \mathbb{R}) \cdot \dim \mathrm{Sp}(2(n-k), \mathbb{R}) \\ &= 4(n-k)k.\end{aligned}$$

□

Note that the real symplectic Grassmann manifold $\mathrm{SpGr}(2n, 2k)$ has the same dimension as the Grassmann manifold $\mathrm{Gr}(2n, 2k)$. The manifolds are not the same however, since not every $2k$ dimensional subspace of $\mathbb{R}^{2n}$ is also a symplectic subspace of the standard symplectic space $(\mathbb{R}^{2n}, \omega_0)$.

Similarly to the Grassmann case [Bat+15], for every $P \in \mathrm{SpGr}(2n, 2k)$, we can define a set

$$\begin{aligned}\mathfrak{sp}_P(2n) &:= \left\{\tilde{\Omega} \in \mathfrak{sp}(2n, \mathbb{R}) \;\middle|\; \tilde{\Omega} = \tilde{\Omega}P + P\tilde{\Omega}\right\} \\ &= \left\{M\Omega M^+ \in \mathfrak{sp}(2n, \mathbb{R}) \;\middle|\; P = ME_0M^+,\ \Omega \in \mathfrak{sp}_{E_0}(2n)\right\}.\end{aligned}$$

The tangent space of $\mathrm{SpGr}(2n, 2k)$ at P is characterized by the following proposition.

Proposition 4.4.2. Let $P \in \mathrm{SpGr}(2n, 2k)$. The tangent space at P is given by

$$\begin{aligned}T_P\mathrm{SpGr}(2n, 2k) &= \{[\Omega, P] \mid \Omega \in \mathfrak{sp}(2n, \mathbb{R})\} \\ &= \{[\tilde{\Omega}, P] \mid \tilde{\Omega} \in \mathfrak{sp}_P(2n)\}.\end{aligned} \tag{4.4.4}$$

Proof. The first equality follows by a straightforward calculation, as every tangent vector is the derivative of a curve defined via (4.4.3). The second equality follows from $\{[\tilde{\Omega}, P] \mid \tilde{\Omega} \in \mathfrak{sp}_P(2n)\} \subset \{[\Omega, P] \mid \Omega \in \mathfrak{sp}(2n, \mathbb{R})\}$ and the fact that for every $\Omega \in \mathfrak{sp}(2n, \mathbb{R})$, it holds that $\tilde{\Omega} := \Omega P + P\Omega - 2P\Omega P \in \mathfrak{sp}_P(2n)$ and $[\Omega, P] = [\tilde{\Omega}, P]$. □

4.4.1 Pseudo-Riemannian metric on $\mathrm{SpGr}(2n, 2k)$

We can connect the real symplectic Stiefel manifold, i.e. the manifold of symplectic bases, with the real symplectic Grassmann manifold in the following way.

Proposition 4.4.3. The map

$$\rho\colon \mathrm{SpSt}(2n, 2k) \to \mathrm{SpGr}(2n, 2k), \quad U \mapsto \rho(U) := UU^+ \tag{4.4.5}$$

is a surjective submersion. Every tangent space $T_U\mathrm{SpSt}(2n, 2k)$ splits into a vertical and horizontal part with respect to ρ and the pseudo-Riemannian metric h_U^{SpSt}, namely $T_U\mathrm{SpSt}(2n, 2k) = \mathsf{Ver}_U^{\rho}\,\mathrm{SpSt}(2n, 2k) \oplus \mathsf{Hor}_U^{\rho;h}\,\mathrm{SpSt}(2n, 2k)$, where

$$\mathsf{Ver}_U^{\rho}\,\mathrm{SpSt}(2n, 2k) := \ker \mathrm{d}\rho_U = \{UA \mid A \in \mathfrak{sp}(2k, \mathbb{R})\}$$

and

$$\mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n,2k) := (\ker \mathrm{d}\rho_U)^{\perp,h_U^{\mathrm{SpSt}}} = \{U^s B \mid B \in \mathbb{R}^{2(n-k)\times 2k}\}.$$

Proof. This is a standard construction. We only show that ρ is a surjective submersion. As $\mathrm{SpGr}(2n,2k)$ is the orbit of E_0 under the group action ϕ of (4.4.3), for every $P \in \mathrm{SpGr}(2n,2k)$ there is $M \in \mathrm{Sp}(2n,\mathbb{R})$ such that $P = MEE^+M^+$. Defining $U = ME \in \mathrm{SpSt}(2n,2k)$ shows that $P = \rho(U)$ and therefore that the map ρ is surjective. To show that ρ is a submersion, we show that the differential $\mathrm{d}\rho_U$ is surjective for every $U \in \mathrm{SpSt}(2n,2k)$: Let $P = UU^+ \in \mathrm{SpGr}(2n,2k)$ and $[\Omega,P] \in T_P\mathrm{SpGr}(2n,2k)$. Then $\Delta := \Omega U \in T_U\mathrm{SpSt}(2n,2k)$ and $\mathrm{d}\rho_U(\Delta) = \Delta U^+ + U\Delta^+ = \Omega UU^+ - UU^+\Omega = [\Omega,P]$. ☐

Let $M \in \mathrm{Sp}(2n,\mathbb{R})$ such that $P = ME_0M^+$ and define $U = ME \in \mathrm{SpSt}(2n,2k)$ and $U^s = ME^s$. Then $P = UU^+$, and it follows from the preceding proposition that $T_P\mathrm{SpGr}(2n,2k)$ can be identified with $\mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n,2k)$ via the horizontal lift. For $\Gamma \in T_P\mathrm{SpGr}(2n,2k)$, this horizontal lift is explicitly given by

$$\Gamma_U^{\mathsf{hor}} = \Gamma U \in \mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n,2k),$$

as can be seen by the fact that there is $\Omega \in \mathfrak{sp}(2n,\mathbb{R})$ such that $\Gamma = [\Omega, UU^+]$ and

$$\begin{aligned}\mathrm{d}\rho_U(\Gamma U) &= \Gamma UU^+ + UU^+\Gamma^+ = \Omega UU^+ - UU^+\Omega UU^+ + UU^+\Omega^+ - UU^+\Omega^+UU^+\\ &= \Omega UU^+ - UU^+\Omega = [\Omega, UU^+] = \Gamma.\end{aligned}$$

By making use of the horizontal lift to $\mathrm{SpSt}(2n,2k)$, we can define a pseudo-Riemannian metric on the real symplectic Grassmann manifold $\mathrm{SpGr}(2n,2k)$.

Proposition 4.4.4. Let $\Gamma_1,\Gamma_2 \in T_P\mathrm{SpGr}(2n,2k)$ and $U \in \rho^{-1}(P)$. There is $B_i \in \mathbb{R}^{2(n-k)\times 2k}$ such that $(\Gamma_i)_U^{\mathsf{hor}} = U^sB_i$, i=1,2. The mapping $g_P^{\mathrm{SpGr}} \colon T_P\mathrm{SpGr}(2n,2k) \times T_P\mathrm{SpGr}(2n,2k) \to \mathbb{R}$,

$$g_P^{\mathrm{SpGr}}(\Gamma_1,\Gamma_2) := h_U^{\mathrm{SpSt}}\left((\Gamma_1)_U^{\mathsf{hor}}, (\Gamma_2)_U^{\mathsf{hor}}\right) = \mathrm{tr}\left(U^+\Gamma_1^+\Gamma_2 U\right) = \mathrm{tr}(B_1^+B_2) \qquad (4.4.6)$$

defines point-wise a pseudo-Riemannian metric on $\mathrm{SpGr}(2n,2k)$.

Proof. Similar to Proposition 4.3.6. ☐

In contrast to $\mathrm{SpSt}(2n,2k)$, which is a naturally reductive space, $\mathrm{SpGr}(2n,2k)$ is even *symmetric* with respect to the pseudo-Riemannian metric (4.4.6). To see this, let $X = \mathrm{diag}(-I_k, I_{n-k}, -I_k, I_{n-k})$ block-diagonal and observe that the involutive automorphism $\sigma\colon \mathrm{Sp}(2n,\mathbb{R}) \to \mathrm{Sp}(2n,R)$, $\sigma(M) = XMX$ fulfills [ONe83, Theorem 11.29]. By [ONe83, Lemma 11.24], it therefore holds that $\rho\circ\pi$ is a pseudo-Riemannian submersion with respect to (4.4.6), since any symmetric space is naturally reductive [ONe83, p. 317].

The connection between $\mathrm{SpSt}(2n,2k)$ and $\mathrm{SpGr}(2n,2k)$ allows us to state the following decomposition of real symplectic Stiefel matrices.

Corollary 4.4.5. Every $U \in \mathrm{SpSt}(2n, 2k)$ is of the form

$$U = Y \operatorname{diag}(\Sigma, \Sigma) N,$$

where $N \in \mathrm{Sp}(2k, \mathbb{R})$, $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_k)$ with $\sigma_i > 0$ for all $i = 1, \dots, k$ and $Y \in \mathrm{St}(2n, 2k)$ fulfills $Y^+ Y = \operatorname{diag}(\Sigma, \Sigma)^{-2}$.

Proof. As in the proof of Proposition 4.4.1, it follows from the Schur-like decomposition [Xu03, Equation (5)] that the matrix $P := UU^+ \in \mathrm{SpGr}(2n, 2k)$ is of the form $P = \tilde{U}\tilde{U}^+$, where $\tilde{U} = QI_{2n,2k} \operatorname{diag}(\Sigma, \Sigma) \in \mathrm{SpSt}(2n, 2k)$ and $Q \in \mathrm{O}(2n)$ is orthogonal. Define $Y := QI_{2n,2k}$. It holds that $Y^T Y = I_{2k}$, so $Y \in \mathrm{St}(2n, 2k)$, and from $\tilde{U}^+ \tilde{U} = I_{2k}$ it follows that $Y^+ Y = \operatorname{diag}(\Sigma, \Sigma)^{-2}$. The claim now follows from the fact that $U = \tilde{U} N$ for some $N \in \mathrm{Sp}(2k, \mathbb{R})$. ☐

As the metric g^{SpGr} of (4.4.6) is defined via a horizontal lift, we obtain the associated geodesics by projection.

Proposition 4.4.6. Let $P \in \mathrm{SpGr}(2n, 2k)$ and $\Gamma \in T_P \mathrm{SpGr}(2n, 2k)$. Furthermore, let $U \in \rho^{-1}(P) \subset \mathrm{SpSt}(2n, 2k)$. The geodesic starting at P in direction Γ with respect to the metric (4.4.6) is

$$\mathrm{Exp}_P^{\mathrm{SpGr}}(t\Gamma) := \rho(\mathrm{Exp}_U^{\mathrm{SpSt},h}(t\Gamma_U^{\mathsf{hor}})) = \exp_{\mathrm{m}}(t[\Gamma, P]) P \exp_{\mathrm{m}}(-t[\Gamma, P]). \tag{4.4.7}$$

Proof. By [ONe83, Proposition 11.31], the pseudo-Riemannian geodesics on the real symplectic Grassmannian $\mathrm{SpGr}(2n, 2k)$ are the projections of the one-parameter subgroups in $\mathrm{Sp}(2n, \mathbb{R})$ under the pseudo-Riemannian submersion $\rho \circ \pi$. Since

$$\rho(\mathrm{Exp}_U^{\mathrm{SpSt},h}(\Gamma_U^{\mathsf{hor}})) = (\rho \circ \pi) \left(\mathrm{Exp}_M^{\mathrm{Sp},h} \left((\Gamma_U^{\mathsf{hor}})_M^{\mathsf{hor}} \right) \right),$$

where $M \in \pi(U)^{-1}$, the claim follows. ☐

By making use of Proposition 4.3.8, we can reduce the computational complexity of (4.4.7). To this end note that $H := \Gamma_U^{\mathsf{hor}} \in \mathsf{Hor}_U^{\rho,h} \mathrm{SpSt}(2n, 2k)$. Therefore, if $H^+ H$ is invertible,

$$\mathrm{Exp}_U^{\mathrm{SpSt},h}(t\Gamma_U^{\mathsf{hor}}) = \begin{bmatrix} -H & U \end{bmatrix} \exp_{\mathrm{m}} \left(t \begin{bmatrix} 0 & -I_{2k} \\ H^+ H & 0 \end{bmatrix} \right) \begin{bmatrix} 0 \\ I_{2k} \end{bmatrix},$$

which implies

$$\begin{aligned} \mathrm{Exp}_P^{\mathrm{SpGr}}(t\Gamma) &= \mathrm{Exp}_U^{\mathrm{SpSt},h}(t\Gamma_U^{\mathsf{hor}})(\mathrm{Exp}_U^{\mathrm{SpSt},h}(t\Gamma_U^{\mathsf{hor}}))^+ \\ &= \begin{bmatrix} -H & U \end{bmatrix} \exp_{\mathrm{m}} \left(t \begin{bmatrix} 0 & -I_{2k} \\ H^+ H & 0 \end{bmatrix} \right) \begin{bmatrix} 0 & 0 \\ 0 & I_{2k} \end{bmatrix} \exp_{\mathrm{m}} \left(t \begin{bmatrix} 0 & H^+ H \\ -I_{2k} & 0 \end{bmatrix} \right) \begin{bmatrix} -H^+ \\ U^+ \end{bmatrix}. \end{aligned} \tag{4.4.8}$$

Lifting to another representative $\tilde{U} = UN$ of P, where $N \in \mathrm{Sp}(2k,\mathbb{R})$, implies $\tilde{H} = \Gamma^{\mathrm{hor}}_{UN} = HN$, with which one can check that (4.4.8) does not depend on the chosen representative.

Finding the (local) inverse of (4.4.7), i.e. given two points $P, F \in \mathrm{SpGr}(2n, 2k)$, find the tangent vector $\Gamma \in T_P\mathrm{SpGr}(2n, 2k)$ such that $\mathrm{Exp}^{\mathrm{SpGr}}_P(\Gamma) = F$, is called the geodesic endpoint problem, or also pseudo-Riemannian logarithm. The structure of the real symplectic Grassmann manifold allows us to find it similarly to the case of the standard Grassmann manifold [Bat+15, Theorem 3.3].

Proposition 4.4.7. Let $P, F \in \mathrm{SpGr}(2n, 2k)$. If

$$\tilde{\Omega} = \frac{1}{2} \log_{\mathrm{m}} \left((I_{2n} - 2F)(I_{2n} - 2P)\right) \tag{4.4.9}$$

is well defined and $\tilde{\Omega} \in \mathfrak{sp}_P(2n)$, it holds for $\Gamma := [\tilde{\Omega}, P] \in T_P\mathrm{SpGr}(2n, 2k)$ that $\mathrm{Exp}^{\mathrm{SpGr}}_P(\Gamma) = F$.

Proof. We have to show that $F = \exp_{\mathrm{m}}([\Gamma, P])P\exp_{\mathrm{m}}(-[\Gamma, P])$. Since by assumption $\tilde{\Omega} \in \mathfrak{sp}_P(2n)$ and therefore $[\Gamma, P] = \tilde{\Omega}$, this is equivalent to showing $F = \exp_{\mathrm{m}}(\tilde{\Omega})P\exp_{\mathrm{m}}(-\tilde{\Omega})$. The fact that $\Gamma \in T_P\mathrm{SpGr}(2n, 2k)$ holds by Proposition 4.4.2. For $\tilde{\Omega}$ defined in (4.4.9), it holds that $(I_{2n} - 2P)\tilde{\Omega}(I_{2n} - 2P) = \frac{1}{2}\log_{\mathrm{m}}((I_{2n} - 2P)(I_{2n} - 2F)) = -\tilde{\Omega}$, since $(I_{2n} - 2P)^{-1} = (I_{2n} - 2P)$. Therefore $(I_{2n} - 2P)\exp_{\mathrm{m}}(-\tilde{\Omega}) = (I_{2n} - 2P)\exp_{\mathrm{m}}(-\tilde{\Omega})(I_{2n} - 2P)^2 = \exp_{\mathrm{m}}(\tilde{\Omega})(I_{2n} - 2P)$. This leads to $\exp_{\mathrm{m}}(\tilde{\Omega})P\exp_{\mathrm{m}}(-\tilde{\Omega}) = \frac{1}{2}I_{2n} + \exp_{\mathrm{m}}(2\tilde{\Omega})(-\frac{1}{2}I_{2n} + P) = F$, which shows the claim. □

4.4.2 Riemannian metric on $\mathrm{SpGr}(2n, 2k)$

As the real symplectic Grassmann manifold $\mathrm{SpGr}(2n, 2k)$ is a quotient of $\mathrm{SpSt}(2n, 2k)$ (and of $\mathrm{Sp}(2n,\mathbb{R})$), we can obtain a Riemannian metric from a right-invariant Riemannian metric on $\mathrm{SpSt}(2n, 2k)$.

Again, we split the tangent space $T_U\mathrm{SpSt}(2n, 2k)$ at $U \in \mathrm{SpSt}(2n, 2k)$ into a vertical part with respect to ρ and a horizontal part with respect to ρ and g^{SpSt}_U from (4.3.16). The former yields (4.4.3), as the vertical space is independent of the metric. The latter gives

$$\begin{aligned}\mathsf{Hor}^{\rho;g}_U \mathrm{SpSt}(2n, 2k) &= (\mathsf{Ver}^{\rho}_U \mathrm{SpSt}(2n, 2k))^{\perp, g} \\ &= \left\{(UH^+ - HU^+)^T U \;\middle|\; U^+H = 0\right\}.\end{aligned}$$

This follows from $g^{\mathrm{SpSt}}_U\left(UA, (UH^+ - HU^+)^T U\right) = 0$ for all $UA \in \mathsf{Ver}^{\rho}_U \mathrm{SpSt}(2n, 2k)$ and all $H \in \mathbb{R}^{2n\times 2k}$ with $U^+H = 0$, and by counting degrees of freedom. For any $\Delta = (UH^+ - HU^+)^T U \in \mathsf{Hor}^{\rho;g}_U \mathrm{SpSt}(2n, 2k)$, the corresponding H (which is not to be confused with $(I_{2n} - UU^+)\Delta$ here) can by calculated via

$$H = (I_{2n} - UU^+)J^T_{2n}\Delta(U^T U)^{-1}J_{2k}. \tag{4.4.10}$$

We can identify the horizontal space $\mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$ with the tangent space $T_{\rho(U)}\operatorname{SpGr}(2n,2k)$ and define a Riemannian metric on $\operatorname{SpGr}(2n,2k)$ via the restriction of the Riemannian metric g^{SpSt} of (4.3.16) to the horizontal spaces.

For $\Gamma_i \in T_{UU^+}\operatorname{SpGr}(2n,2k)$, let $(\Gamma_i)_U^{\mathsf{hor},g} \in \mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$ be the horizontal lift to $\mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$, i.e. $\mathrm{d}\rho_U((\Gamma_i)_U^{\mathsf{hor},g}) = \Gamma_i$. The mapping

$$g_{UU^+}^{\mathrm{SpGr}} \colon T_{UU^+}\operatorname{SpGr}(2n,2k) \times T_{UU^+}\operatorname{SpGr}(2n,2k) \to \mathbb{R},$$
$$g_{UU^+}^{\mathrm{SpGr}}(\Gamma_1,\Gamma_2) := g_U^{\mathrm{SpSt}}((\Gamma_1)_U^{\mathsf{hor},g}, (\Gamma_2)_U^{\mathsf{hor},g})$$

defines pointwise a Riemannian metric. We are not aware of an explicit mapping to calculate the horizontal lift with respect to g^{SpSt} for a given $\Gamma \in T_{UU^+}\operatorname{SpGr}(2n,2k)$. Nevertheless, one can directly work with symplectic Stiefel representatives and horizontal tangent vectors, i.e., with $2n \times 2k$-matrices.

Lemma 4.4.8. For two horizontal tangent vectors

$$\Delta_i := (\Gamma_i)_U^{\mathsf{hor},g} = (UH_i^+ - H_iU^+)^T U \in \mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k),$$

it holds that

$$\begin{aligned} g_U^{\mathrm{SpSt}}((\Gamma_1)_U^{\mathsf{hor},g}, (\Gamma_2)_U^{\mathsf{hor},g}) &= \operatorname{tr}\left((U^TU)^{-1}\Delta_1^T(I_{2n} - UU^+)\Delta_2\right) \\ &= \operatorname{tr}\left(U^TU(H_2^TH_1)^+ - (U^TH_1)^+H_2^TU\right). \end{aligned}$$

Proof. This follows by a direct calculation from the properties of the trace. □

Let f be a function on the real symplectic Grassmannian, given on symplectic Stiefel representatives by $f\colon \operatorname{SpSt}(2n,2k) \to \mathbb{R}$, with $f(U) = f(UN)$ for all $N \in \operatorname{Sp}(2k,\mathbb{R})$. We assume that f can (locally) be extended to a smooth function on $2n \times 2k$-matrices, for convenience again denoted by f. The Riemannian gradient of f with respect to g^{SpGr} is given by

$$\operatorname{grad}_f^g(U) = (UH^+ - HU^+)^TU = J_{2n}^THJ_{2k}U^TU - J_{2n}^TUJ_{2k}H^TU,$$

with

$$H = (I_{2n} - UU^+)J_{2n}^T\nabla f_U J_{2k},$$

where ∇f_U denotes the Euclidean gradient of a smooth extension of f around U in $\mathbb{R}^{2n\times 2k}$. This follows from [AMS08, Equation (3.39)] and $g_U^{\mathrm{SpSt}}(\operatorname{grad}_f^g(U),\Delta) = \mathrm{d}f_U(\Delta) = \operatorname{tr}((\nabla f_U)^T\Delta)$ for all $\Delta \in \mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$, as well as the fact that $\operatorname{grad}_f^g(U) \in \mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$.

Proposition 4.4.9. Let $U \in \operatorname{SpSt}(2n,2k)$ and $\Delta \in \mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$. The lifted symplectic Grassmann geodesic from U in direction Δ is given by

$$\operatorname{Exp}_U^{\mathrm{SpGr}}(t\Delta) = \exp_{\mathrm{m}}(t(\bar\Omega - \bar\Omega^T))\exp_{\mathrm{m}}(t\bar\Omega^T)U, \tag{4.4.11}$$

where $\bar{\Omega}$ is given by (4.3.15).

Proof. We need to show that the tangent vector $\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Exp}_U^{\mathrm{SpGr}}(t\Delta)$ is horizontal for every t. Then, the claim follows from [ONe83, Cor. 7.46]. Since $\Delta \in \mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$ is equivalent to $J_{2n}UJ_{2k}^TU^T\bar{\Omega}(\Delta)J_{2n}UJ_{2k}^TU^T = 0$, the proof follows in the same fashion as the one of Lemma 4.3.10. □

Since $\bar{\Omega}^T = UH^+ - HU^+$, with H from (4.4.10), we can reduce (4.4.11) with (4.3.10) to the matrix exponentials of a $8k \times 8k$ and $4k \times 4k$ matrix, respectively.

Proposition 4.4.10. Let $U \in \operatorname{SpSt}(2n,2k)$ and $\Delta \in \mathsf{Hor}_U^{\rho;g}\operatorname{SpSt}(2n,2k)$. Define H as in (4.4.10),

$$X := \begin{bmatrix} J_{2n}^T H J_{2k} & -J_{2n}^T U J_{2k} & -U & H \end{bmatrix} \in \mathbb{R}^{2n\times 8k}$$

and

$$Y := \begin{bmatrix} U & H & J_{2n}^T H J_{2k} & J_{2n}^T U J_{2k} \end{bmatrix} \in \mathbb{R}^{2n\times 8k}.$$

Then

$$\operatorname{Exp}_U^{\mathrm{SpGr}}(t\Delta) = X \exp_{\mathrm{m}}(tY^TX)\begin{bmatrix} 0_{4k} \\ -I_{4k} \end{bmatrix} \exp_{\mathrm{m}}\left(t\begin{bmatrix} 0 & -H^+H \\ I_{2k} & 0 \end{bmatrix}\right)\begin{bmatrix} I_{2k} \\ 0 \end{bmatrix}.$$

Proof. It holds that $\bar{\Omega} - \bar{\Omega}^T = XY^T$ and $\bar{\Omega}^T = UH^+ - HU^+$. By (4.3.10),

$$\begin{aligned} \exp_{\mathrm{m}}(\bar{\Omega} - \bar{\Omega}^T)\begin{bmatrix} U & -H \end{bmatrix} &= \exp_{\mathrm{m}}(XY^T)\begin{bmatrix} U & -H \end{bmatrix} \\ &= (I_{2n} + X(\exp_{\mathrm{m}}(Y^TX) - I_{8k})(Y^TX)^{-1}Y^T)\begin{bmatrix} U & -H \end{bmatrix}. \end{aligned}$$

Since

$$(Y^TX)^{-1}Y^T\begin{bmatrix} U & -H \end{bmatrix} = (Y^TX)^{-1}Y^TX\begin{bmatrix} 0_{4k} \\ -I_{4k} \end{bmatrix},$$

it follows that $\exp_{\mathrm{m}}(\bar{\Omega} - \bar{\Omega}^T)\begin{bmatrix} U & -H \end{bmatrix} = X\exp_{\mathrm{m}}(Y^TX)\left[\begin{smallmatrix} 0_{4k} \\ -I_{4k} \end{smallmatrix}\right]$. Together with

$$\exp_{\mathrm{m}}(\bar{\Omega}^T)U = \begin{bmatrix} U & -H \end{bmatrix}\exp_{\mathrm{m}}\left(t\begin{bmatrix} 0 & -H^+H \\ I_{2k} & 0 \end{bmatrix}\right)\begin{bmatrix} I_{2k} \\ 0 \end{bmatrix}$$

as in Proposition 4.3.8, this shows the claim. □

4.5 Retractions and computational issues

Calculating the matrix exponential of an $n \times n$ matrix is computationally expensive if n is large. Furthermore, numerical experiments show that even though the matrix exponential of a Hamiltonian matrix is theoretically guaranteed to yield a symplectic matrix as an output, this is not necessarily the case in practice, where one needs to rely on numerical tools to compute the standard matrix exponential. While there are specialized algorithms for the matrix exponential of a Hamiltonian matrix [KLS19], there is another alternative: The Cayley map. In this section, we propose the use of the Cayley map for approximating the pseudo-Riemannian geodesics in order to define retractions on the symplectic Stiefel and Grassmann manifold. Furthermore, these retractions turn out to be invertible in closed form on both manifolds, which can for example be used for interpolation and optimization purposes and for defining local coordinates. In the experiments of Section 4.6, the Cayley-based retraction turns out to be computationally cheaper and to retain the manifold structure to a much higher numerical accuracy.

A *retraction* [AMS08] on a smooth manifold M with tangent bundle TM is a smooth mapping $R\colon TM \to M$ such that for any $x \in M$,

1. $R_x(0) = x$,
2. $\mathrm{d}(R_x)_0 = \mathrm{id}$,

where R_x is the restriction of R to T_xM.

The Cayley transformation

$$\mathrm{cay}(X) := (I_n + X)(I_n - X)^{-1}, \quad X \in \mathbb{R}^{n\times n}$$

is widely used as a standard approximation of the matrix exponential $\exp_{\mathrm{m}}(2X)$. Of special interest in the present context is the property that cay maps from $\mathfrak{sp}(2n, \mathbb{R})$ to $\mathrm{Sp}(2n, \mathbb{R})$ [AG01]. This was also exploited in [Fio11]. The inverse of the Cayley transform is given by [AG01]

$$\mathrm{cay}^{-1}(M) = (M - I_n)(I_n + M)^{-1}.$$

4.5.1 Cayley retraction on the real symplectic Stiefel manifold

Replacing the matrix exponential in the pseudo-Riemannian exponential (4.3.9) on the symplectic Stiefel manifold with the Cayley transform leads to the Cayley retraction defined in [Gao+21b, Definition 5.2]. Yet note that the Cayley retraction in the aforementioned reference was found unaware of the pseudo-Riemannian geodesics by transferring the Cayley retraction on the classical Stiefel manifold $\mathrm{St}(n, k)$ to the symplectic case.

Proposition 4.5.1. Let $U \in \mathrm{SpSt}(2n, 2k)$ and $\Delta \in T_U\mathrm{SpSt}(2n, 2k)$. For $\Delta = \tilde{\Omega}(U, \Delta)U$, with $\tilde{\Omega}(U, \Delta)$ as in (4.3.7), the map

$$\mathcal{R}_U^{\mathrm{SpSt}}(\Delta) := \mathrm{cay}\left(\frac{1}{2}\tilde{\Omega}(U, \Delta)\right) U \tag{4.5.1}$$

is a retraction. The derivative of the curve $\gamma(t) := \mathcal{R}_U^{\mathrm{SpSt}}(t\Delta) = \mathrm{cay}(\frac{t}{2}\tilde{\Omega}(U, \Delta))U$ is given by

$$\dot{\gamma}(t) = \frac{1}{2}\left(\left(I_{2n} + \frac{t}{2}\tilde{\Omega}(U,\Delta)\right)^{-1} + \left(I_{2n} - \frac{t}{2}\tilde{\Omega}(U,\Delta)\right)^{-1}\right)\tilde{\Omega}(U,\Delta)\gamma(t)$$

Proof. The fact that $\mathcal{R}^{\mathrm{SpSt}}$ is a retraction is shown in [Gao+21b, Prop. 5.3]. The formula for $\dot{\gamma}(t)$ follows from a straightforward calculation, making use of the fact that $\tilde{\Omega}(U,\Delta)$ commutes with $\mathrm{cay}(\frac{t}{2}\tilde{\Omega}(U,\Delta))$. □

In [Gao+21b, Proposition 5.5], it was proposed to use the Sherman-Morrison-Woodbury formula

$$(A + XY^T)^{-1} = A^{-1} - A^{-1}X(I + Y^T A^{-1} X)^{-1} Y^T A^{-1},$$

where $A \in \mathbb{R}^{n\times n}$, $X, Y \in \mathbb{R}^{n\times k}$, to reduce the matrix inverse in (4.5.1) from $2n \times 2n$ to $4k \times 4k$. We show that we can even reduce it to a matrix inversion of dimensions $2k \times 2k$.

Proposition 4.5.2. Let $U \in \mathrm{SpSt}(2n, 2k)$ and $\Delta \in T_U\mathrm{SpSt}(2n, 2k)$. Define $A := U^+\Delta$ and $H := \Delta - UA$. Then

$$\mathcal{R}_U^{\mathrm{SpSt}}(t\Delta) = -U + (tH + 2U)\left(\frac{t^2}{4}H^+H - \frac{t}{2}A + I_{2k}\right)^{-1}. \tag{4.5.2}$$

Proof. For $t = 0$, the equality holds. In the following, assume $t \neq 0$. Similarly to the proof of Proposition 4.3.8, define $X = \begin{bmatrix}\frac{1}{2}UA + H & -U\end{bmatrix} \in \mathbb{R}^{2n\times 4k}$ and $Y^T = \begin{bmatrix} U^+ \\ \Delta^+(I_{2n} - \frac{1}{2}UU^+)\end{bmatrix} \in \mathbb{R}^{4k\times 2n}$. Then again $\tilde{\Omega}(U,\Delta) = XY^T$ as in (4.3.7), and

$$Y^T X = \begin{bmatrix} \frac{1}{2}A & -I_{2k} \\ H^+H - \frac{1}{4}A^2 & \frac{1}{2}A \end{bmatrix}.$$

By definition

$$\begin{aligned}
\mathcal{R}_U^{\mathrm{SpSt}}(t\Delta) &= \mathrm{cay}\left(\frac{t}{2}\tilde{\Omega}(U,\Delta)\right)U = \left(I_{2n} + \frac{t}{2}XY^T\right)\left(I_{2n} - \frac{t}{2}XY^T\right)^{-1}U \\
&= \left(I_{2n} + \frac{t}{2}XY^T\right)\left(I_{2n} + \frac{t}{2}X(I_{4k} - \frac{t}{2}Y^TX)^{-1}Y^T\right)U \\
&= U + tX(I_{4k} - \frac{t}{2}Y^TX)^{-1}Y^TU.
\end{aligned}$$

It holds that

$$I_{4k} - \frac{t}{2}Y^TX = \begin{bmatrix} I_{2k} - \frac{t}{4}A & \frac{t}{2}I_{2k} \\ -\frac{t}{2}(H^+H - \frac{1}{4}A^2) & I_{2k} - \frac{t}{4}A \end{bmatrix}.$$

Block-matrix inversion via the Schur complement yields

$$(I_{4k} - \frac{t}{2}Y^TX)^{-1} = \begin{bmatrix} -\frac{1}{2}\Theta^{-1}(\frac{t}{2}A - 2I_{2k}) & -\frac{t}{2}\Theta^{-1} \\ -\frac{1}{2t}(\frac{t}{2}A - 2I_{2k})\Theta^{-1}(\frac{t}{2}A - 2I_{2k}) + \frac{2}{t}I_{2k} & -\frac{1}{2}(\frac{t}{2}A - 2I_{2k})\Theta^{-1} \end{bmatrix}$$

with $\Theta = \frac{t^2}{4}H^+H - \frac{t}{2}A + I_{2k} \in \mathbb{R}^{2k\times 2k}$. Writing $Y^TU = \begin{bmatrix} I_{2k} \\ -\frac{1}{2}A \end{bmatrix}$ it follows that

$$(I_{4k} - \frac{t}{2}Y^TX)^{-1}Y^TU = \begin{bmatrix} \Theta^{-1} \\ \frac{1}{t}(\frac{t}{2}A - 2I_{2k})\Theta^{-1} + \frac{2}{t}I_{2k} \end{bmatrix}.$$

Putting everything together, we obtain

$$\begin{aligned} \mathcal{R}_U^{\mathrm{SpSt}}(t\Delta) &= U + t(\frac{1}{2}UA + H - \frac{1}{t}U(\frac{t}{2}A - 2I_{2k} + \frac{t^2}{2}H^+H - tA + 2I_{2k}))\Theta^{-1} \\ &= -U + (tH + 2U)(\frac{t^2}{4}H^+H - \frac{t}{2}A + I_{2k})^{-1}, \end{aligned}$$

which shows the claim. □

Unlike the pseudo-Riemannian exponential (4.3.11) or the Riemannian exponential (4.3.19), we can invert the Cayley retraction (4.5.2) in closed form. Apart from interpolation, this facilitates the calculation of local coordinates on $\mathrm{SpSt}(2n, 2k)$.

Proposition 4.5.3. Let $U, V \in \mathrm{SpSt}(2n, 2k)$. If $(I_{2k} + U^+V)^{-1}$ and $(I_{2k} + V^+U)^{-1}$ exist, it holds for

$$A = 2((I_{2k} + V^+U)^{-1} - (I_{2k} + U^+V)^{-1}) \in \mathfrak{sp}(2k, \mathbb{R})$$

and

$$H = 2((V + U)(I_{2k} + U^+V)^{-1} - U) \in \mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n, 2k),$$

that

$$\mathcal{L}_U^{\mathrm{SpSt}}(V) := UA + H \in T_U\mathrm{SpSt}(2n, 2k) \tag{4.5.3}$$

fulfills $\mathcal{R}_U^{\mathrm{SpSt}}(\mathcal{L}_U^{\mathrm{SpSt}}(V)) = V$.

Proof. Since $A^+ = -A$, it holds that $A \in \mathfrak{sp}(2k, \mathbb{R})$ and $U^+H = 2(U^+V + I_{2k})(I_{2k} + U^+V)^{-1} - I_{2k}) = 0$ implies $H \in \mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n, 2k)$. Therefore, $\mathcal{L}_U^{\mathrm{SpSt}}(V)$ is a valid tangent vector. Since

$$\frac{1}{4}H^+H = (U^+V + I_{2k})^{-1} + (V^+U + I_{2k})^{-1} - I_{2k},$$

it holds that

$$\frac{1}{4}H^+H - \frac{1}{2}A + I_{2k} = 2(U^+V + I_{2k})^{-1}.$$

Therefore

$$\mathcal{R}_U^{\mathrm{SpSt}}(\mathcal{L}_U^{\mathrm{SpSt}}(V)) = -U + (H+2U)(\frac{1}{4}H^+H - \frac{1}{2}A + I_{2k})^{-1}$$
$$= -U + \frac{1}{2}H(U^+V + I_{2k}) + U(U^+V + I_{2k}) = V,$$

which shows the claim. □

4.5.2 Cayley retraction on the real symplectic Grassmann manifold

With the quotient manifold approach to the symplectic Stiefel manifold and the definition of the symplectic Grassmann manifold, we can show an additional property of $\mathcal{R}^{\mathrm{SpSt}}$: It maps horizontal tangent vectors (with respect to the pseudo-Riemannian metric h^{SpSt} from (4.3.8)) to curves with horizontal tangent vectors everywhere. We can therefore use it to calculate approximations of the pseudo-Riemannian symplectic Grassmann geodesics lifted to the symplectic Stiefel manifold.

Proposition 4.5.4. Let $U \in \mathrm{SpSt}(2n,2k)$ and $\Delta \in \mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n,2k)$. Furthermore, let $\tilde{\Omega}(U,\Delta)$ is as in (4.3.7). For

$$\gamma(t) := \mathcal{R}_U^{\mathrm{SpSt}}(t\Delta) = \mathrm{cay}\left(\frac{t}{2}\tilde{\Omega}(U,\Delta)\right)U,$$

it holds that $\dot{\gamma}(t) \in \mathsf{Hor}_{\gamma(t)}^{\rho,h}\,\mathrm{SpSt}(2n,2k)$ for all t.

Proof. We suppress the dependence of $\tilde{\Omega}$ on U and Δ for better legibility. We have to show that $\gamma(t)^+\dot{\gamma}(t) = 0$ for all t. It holds that $\mathrm{cay}(\frac{t}{2}\tilde{\Omega})$ commutes with $\tilde{\Omega}$ and with $\left(I_{2n} \pm \frac{t}{2}\tilde{\Omega}\right)^{-1}$, respectively, and $\mathrm{cay}(-\frac{t}{2}\tilde{\Omega})\,\mathrm{cay}(\frac{t}{2}\tilde{\Omega}) = I_{2n}$. Since $\gamma(t)^+ = U^+\,\mathrm{cay}(-\frac{t}{2}\tilde{\Omega})$, it follows that

$$\gamma(t)^+\dot{\gamma}(t) = \frac{1}{2}U^+\left(\left(I_{2n} + \frac{t}{2}\tilde{\Omega}\right)^{-1} + \left(I_{2n} - \frac{t}{2}\tilde{\Omega}\right)^{-1}\right)\tilde{\Omega}U.$$

Furthermore $\gamma(t)^+\dot{\gamma}(t) = 0$ is equivalent to $U\gamma(t)^+\dot{\gamma}(t) = 0$. It holds that $\tilde{\Omega} \in \mathfrak{sp}_{UU^+}(2n,\mathbb{R})$, since $\Delta \in \mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n,2k)$, which means $\tilde{\Omega} = \tilde{\Omega}UU^+ + UU^+\tilde{\Omega}$. Then

$$UU^+(I_{2n} \pm \frac{t}{2}\tilde{\Omega}) = (I_{2n} \mp \frac{t}{2}\tilde{\Omega})UU^+ \pm \frac{t}{2}\tilde{\Omega},$$

which implies

$$UU^+(I_{2n} \pm \frac{t}{2}\tilde{\Omega})^{-1} = (I_{2n} \mp \frac{t}{2}\tilde{\Omega})^{-1}UU^+ \mp \frac{t}{2}(I_{2n} \mp \frac{t}{2}\tilde{\Omega})^{-1}\tilde{\Omega}(I_{2n} \pm \frac{t}{2}\tilde{\Omega})^{-1}.$$

Therefore

$$U\gamma(t)^{+}\dot{\gamma}(t) = \frac{1}{2}UU^{+}\left(\left(I_{2n} + \frac{t}{2}\tilde{\Omega}\right)^{-1} + \left(I_{2n} - \frac{t}{2}\tilde{\Omega}\right)^{-1}\right)\tilde{\Omega}U$$
$$= \frac{1}{2}\left(\left(I_{2n} - \frac{t}{2}\tilde{\Omega}\right)^{-1} + \left(I_{2n} + \frac{t}{2}\tilde{\Omega}\right)^{-1}\right)UU^{+}\tilde{\Omega}U = 0,$$

because $UU^{+}\tilde{\Omega}U = \tilde{\Omega}(I_{2n} - UU^{+})U = 0$. □

Projecting the retraction from Proposition 4.5.4 to the symplectic Grassmann manifold leads to a retraction on $\mathrm{SpGr}(2n, 2k)$.

Proposition 4.5.5. Let $P \in \mathrm{SpGr}(2n, 2k)$ and $\Gamma \in T_P\mathrm{SpGr}(2n, 2k)$. Then

$$\mathcal{R}_P^{\mathrm{SpGr}}(\Gamma) := \mathrm{cay}\left(\frac{1}{2}[\Gamma, P]\right) P \,\mathrm{cay}\left(-\frac{1}{2}[\Gamma, P]\right)$$

defines a retraction on $\mathrm{SpGr}(2n, 2k)$. For $U \in \rho^{-1}(P) \subset \mathrm{SpSt}(2n, 2k)$, it fulfills $\mathcal{R}_P^{\mathrm{SpGr}}(\Gamma) = \rho(\mathcal{R}_U^{\mathrm{SpSt}}(\Gamma_U^{\mathsf{hor}}))$. The curve $\gamma(t) := \mathcal{R}_P^{\mathrm{SpGr}}(t\Gamma)$ fulfills

$$\dot{\gamma}(t) = \left[\frac{1}{2}[\Gamma, P]\left(\left(I_{2n} - \frac{t}{2}[\Gamma, P]\right)^{-1} + \left(I_{2n} + \frac{t}{2}[\Gamma, P]\right)^{-1}\right), \gamma(t)\right].$$

Proof. The first retraction property $\mathcal{R}_P^{\mathrm{SpGr}}(0) = P$ is immediate. The formula for $\dot{\gamma}(t)$ follows from a direct calculation, whence $\mathrm{d}(\mathcal{R}_P^{\mathrm{SpGr}})_0(\Gamma) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{R}_P^{\mathrm{SpGr}}(t\Gamma)|_{t=0} = \dot{\gamma}(0) = [[\Gamma, P], P] = \Gamma$. This implies the second retraction property. □

Similarly to Proposition 4.4.7, we can invert the retraction on the symplectic Grassmann manifold in closed form. As in the symplectic Stiefel case, this defines local coordinates on $\mathrm{SpGr}(2n, 2k)$. In the following results, let $\mathrm{sqrt_m}(\cdot)$ denote the principal matrix square root.

Proposition 4.5.6. Let $P, F \in \mathrm{SpGr}(2n, 2k)$. If for

$$\tilde{\Omega} := 2\,\mathrm{cay}^{-1}\left(\mathrm{sqrt_m}\left((I_{2n} - 2F)(I_{2n} - 2P)\right)\right), \tag{4.5.4}$$

it holds that $\tilde{\Omega} \in \mathfrak{sp}_P(2n)$, then $F = \mathcal{R}_P^{\mathrm{SpGr}}([\tilde{\Omega}, P])$.

Proof. It holds that $\mathrm{cay}\left(\frac{1}{2}\tilde{\Omega}\right)^2 = (I_{2n} - 2F)(I_{2n} - 2P)$. Since $(I_{2n} - 2P)^2 = I_{2n}$ and

$$(I_{2n} - 2P)\,\mathrm{cay}\left(\frac{1}{2}\tilde{\Omega}\right)(I_{2n} - 2P) = \mathrm{cay}\left(-\frac{1}{2}\tilde{\Omega}\right),$$

it follows that

$$I_{2n} - 2F = \operatorname{cay}\left(\frac{1}{2}\tilde{\Omega}\right)^2 (I_{2n} - 2P) = \operatorname{cay}\left(\frac{1}{2}\tilde{\Omega}\right)(I_{2n} - 2P)\operatorname{cay}\left(-\frac{1}{2}\tilde{\Omega}\right)$$
$$= I_{2n} - 2\operatorname{cay}\left(\frac{1}{2}\tilde{\Omega}\right) P \operatorname{cay}\left(-\frac{1}{2}\tilde{\Omega}\right),$$

which implies the claimed result. □

We can also directly invert $\mathcal{R}_P^{\mathrm{SpGr}}(\Gamma)$ on symplectic Stiefel representatives.

Proposition 4.5.7. Let $U, V \in \mathrm{SpSt}(2n, 2k)$. If

$$N := (U^+V)^{-1}\operatorname{sqrt_m}(U^+VV^+U) \in \mathrm{Sp}(2k, \mathbb{R})$$

and

$$H := 2(VN + U)(U^+VN + I_{2k})^{-1} - 2U \in \mathsf{Hor}_U^{\rho,h}\,\mathrm{SpSt}(2n, 2k)$$

are well-defined, it holds that

$$\mathcal{R}_U^{\mathrm{SpSt}}(H) = VN.$$

Proof. Since $NN^+ = (U^+V)^{-1}(U^+VV^+U)(V^+U)^{-1} = I_{2k}$, it holds that $N \in \mathrm{Sp}(2k, \mathbb{R})$. Furthermore

$$N^+V^+U = (U^+VN)^+ = (\operatorname{sqrt_m}(U^+VV^+U))^+ = \operatorname{sqrt_m}(U^+VV^+U) = U^+VN.$$

Therefore $\frac{1}{4}H^+H = 2(I_{2k} + U^+VN)^{-1} - I_{2k}$, which implies

$$\mathcal{R}_U^{\mathrm{SpSt}}(H) = -U + (H + 2U)(\frac{1}{4}H^+H + I_{2k})^{-1} = -U + VN + U = VN.$$

□

The difference between the connecting curves from Proposition 4.5.3 and Proposition 4.5.7 is visualized in Figure 4.2.

4.6 Numerical Experiments

In this section, we study the feasibility of different retractions on $\mathrm{SpSt}(2n, 2k)$ and investigate optimization problems via gradient descent on $\mathrm{SpSt}(2n, 2k)$ and $\mathrm{SpGr}(2n, 2k)$, respectively. All experiments are conducted with MATLAB version R2021a on a laptop with Ubuntu 18.04, an Intel® Core™ i7-8850H CPU and 16GB RAM. We generate random Hamiltonian matrices via $\Omega = \begin{bmatrix} A & B \\ C & -A^T \end{bmatrix}$, where A, B and C are generated by `randn(n,n)`, and then B and C are symmetrized. For reproducability, all random matrices are constructed with the random stream `s = RandStream('mt19937ar')`.

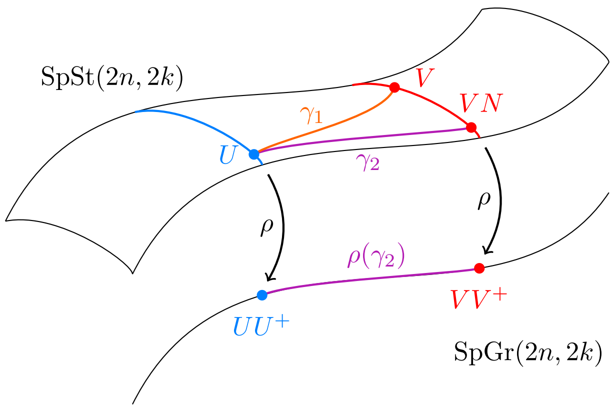

Figure 4.2: Connecting curves according to the inverse retractions Proposition 4.5.3 (γ_1) and Proposition 4.5.7 (γ_2).

4.6.1 Feasibility of different retractions on the symplectic Stiefel manifold

We compare the numerical feasibility of the Riemannian geodesic (4.3.19) with respect to g^{SpSt}, the Cayley-retraction (4.5.2), the pseudo-Riemannian geodesic (4.3.11) and the quasi-geodesic retraction (4.3.12) from [Gao+21b]. To this end, we generate a (pseudo) random point on $\mathrm{SpSt}(2n, 2k)$ via $U = \mathrm{cay}(\Omega)E$, where $\Omega \in \mathfrak{sp}(2n, \mathbb{R})$ is scaled to $||\Omega||_F = 1$. We furthermore generate a (pseudo) random tangent vector $\Delta \in T_U\mathrm{SpSt}(2n, 2k)$, also scaled to $||\Delta||_F = 1$. For the chosen retractions $\mathcal{R}$, we calculate $U(t) = \mathcal{R}_U(t\Delta)$ with $t \in [0, 10^3]$ and plot the feasibility $||U(t)^+U(t) - I_{2n}||_F$. The average over 10 runs is shown in Figure 4.3 for $n = 1000, k = 20$ (left) and $n = 1000, k = 200$ (right). It can be seen that the Riemannian geodesic, the pseudo-Riemannian geodesic and the quasi-geodesic retraction, which all rely on the matrix exponential, fail numerically to stay on $\mathrm{SpSt}(2n, 2k)$ for tangent vectors of Frobenius-norm larger than $\mathcal{O}(10^2)$. The Cayley-Retraction, while less feasible at some points, fulfills the manifold condition up to an error of about 10^{-8} for tangent vectors of any tested size on $\mathrm{SpSt}(2000, 40)$ and up to an error of about 10^{-4} on $\mathrm{SpSt}(2000, 400)$.

4.6.2 Gradient descent on the real symplectic Stiefel manifold

We tackle an academic instance of the 'nearest symplectic matrix' problem

$$\min_{U \in \mathrm{SpSt}(2n,2k)} ||U - A||_F^2$$

via a Riemannian gradient descent. For this, we set `A = randn(2*n,2*k)` and then normalize `A = A/norm(A,2)`, as in [Gao+21b]. The initial point for starting the optimization procedure is set to be `U0 = cay(X/2)E`, where $X \in \mathfrak{sp}(2n, \mathbb{R})$ is a random

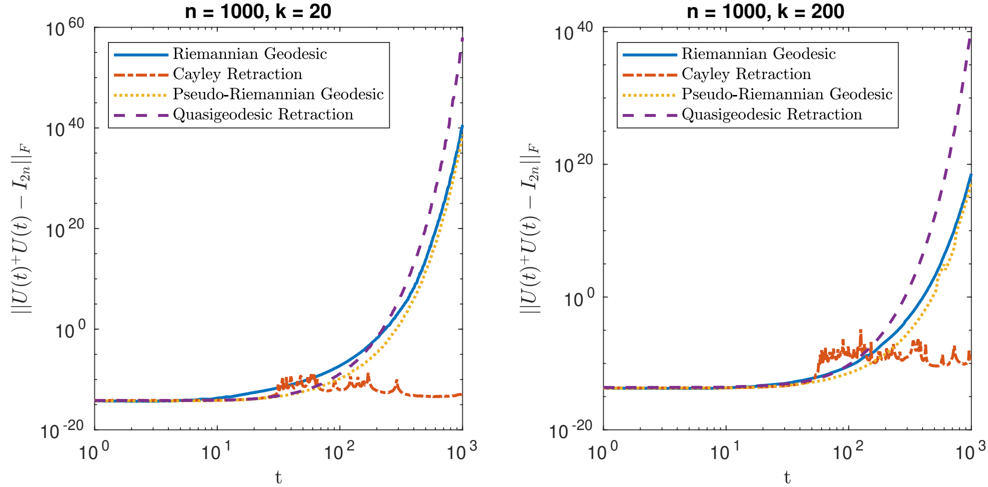

Figure 4.3: (cf. Subsection 4.6.1) Comparison of the feasibility $\|U(t)^+U(t) - I_{2n}\|_F$ of different retractions $\mathcal{R}$ on $\mathrm{SpSt}(2000, 40)$ and $\mathrm{SpSt}(2000, 400)$, where $U(t) = \mathcal{R}_U(t\Delta)$ and $\|\Delta\|_F = 1$. The data represent an average over 10 runs, $U(t)$ is evaluated at 500 logarithmically spaced steps.

$2n \times 2n$ Hamiltonian matrix, scaled by `X = X/norm(X,'fro')`. As gradient descent algorithm we use [Gao+21b, Alg. 1] with monotone line search and stopping criterion [Gao+21b, Eqs. (6.1) and (6.2)]. For the reader's convenience, we restate the procedure here as Algorithm 4.6.1 in the precise form in which we use it. As the trial step size γ_k, we use the alternating BB method γ_k^{ABB} [Gao+21b, Equation (6.4)], with the respective gradient. The other method parameters are set to $\delta = 0.1, \beta = 10^{-4}, \gamma_{\min} = 10^{-15}$ and $\gamma_{\max} = 10^{15}$, as in [Gao+21b, Subsection 6.1]. The step parameters are set to $h_{\min} = 0$ and $h_{\max} = 5$, and the tolerance parameters to $\epsilon = 10^{-6}$, $\epsilon_x = 10^{-6}$ and $\epsilon_f = 10^{-12}$, respectively.

We compare gradient descent for the Riemannian metric g^{SpSt} with geodesic stepping and Cayley stepping, respectively, with gradient descent from [Gao+21b] with Cayley stepping. For the gradient descent according to [Gao+21b], we choose the optimal settings stated in this reference, i.e., the canonical-like metric g_ρ with $\rho = \frac{1}{2}$ and gradient (I), according to [Gao+21b, Subsection 6.2.2]. In the actual implementation of all methods included in this comparison, care has been taken that the action of large matrices like J_{2n} and I_{2n} is applied directly, so that these matrices are never formed explicitly.

Figure 4.4 displays the objective function value versus the iteration count (left) and the convergence history according to the gradient norm (right), respectively. For comparison purposes, all methods are run for a fixed number of 60 iterations. It can be seen that the algorithms deliver similar results in regard of the convergence by itera-

Algorithm 4.6.1 Gradient descent algorithm [Gao+21b, Alg. 1]

Input: $U_0 \in \mathrm{SpSt}(2n, 2k)$, $f\colon \mathrm{SpSt}(2n, 2k) \to \mathbb{R}$, retraction $\mathcal{R}$, $\beta, \delta \in (0, 1)$, $0 < \gamma_{\min} < \gamma_{\max}$, initial step size $\gamma_0^{ABB} = f(U_0)$, maximal iterations $N \in \mathbb{N}$, Riemannian metric $\langle \cdot, \cdot \rangle_U$ with gradient grad_f, step parameters $h_{\min} < h_{\max} \in \mathbb{Z}$, tolerance parameters $\epsilon, \epsilon_x, \epsilon_f > 0$

for $0 \le k \le N$ **do**
  $\Delta_k = -\mathrm{grad}_f(U_k)$
  if $k > 0$ **then**
    $S_k = U_k - U_{k-1}$ and $Y_k = \mathrm{grad}_f(U_k) - \mathrm{grad}_f(U_{k-1})$
    if k is odd **then**
      $\gamma_k^{ABB} = \frac{\langle S_k, S_k \rangle}{|\langle S_k, Y_k \rangle|}$, where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product.
    else
      $\gamma_k^{ABB} = \frac{|\langle S_k, Y_k \rangle|}{\langle Y_k, Y_k \rangle}$, where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product.
    end if
  end if
  $\gamma_k = \max(\gamma_{\min}, \min(\gamma_k^{ABB}, \gamma_{\max}))$.
  for $h_{\min} \le h \le h_{\max}$ **do**
    $t_k = \gamma_k \delta^h$
    if $f(\mathcal{R}_{U_k}(t_k \Delta_k)) \le f(U_k) - \beta t_k \langle \Delta_k, \Delta_k \rangle_{U_k}$ **then**
      Break
    end if
  end for
  $U_{k+1} = \mathcal{R}_{U_k}(t_k \Delta_k)$
  if $\|\mathrm{grad}_f(U_k)\|_F < \epsilon$ **then**
    if $\frac{|f(U_k) - f(U_{k+1})|}{|f(U_k)| + 1} < \epsilon_f$ and $\frac{\|U_k - U_{k+1}\|_F}{\sqrt{2n}} < \epsilon_x$ **then**
      Break as converged.
    end if
  end if
end for

Output: Iterates U_k.

tions, depending on the chosen tolerance. The run time however differs: In Table 4.1, we compare the three methods and state the average iterations and run time until numerical convergence over 10 runs. We furthermore denote the relative deviation from the respective minimum over all three methods after convergence. It can be seen that for $\mathrm{SpSt}(2000, 40)$ and $\mathrm{SpSt}(2000, 400)$, gradient descent with Cayley stepping is the fastest method regarding run time, while Geodesic descent is the slowest. For $\mathrm{SpSt}(2000, 400)$, the run time for geodesic stepping increases drastically, since (4.3.19) requires the matrix exponential of both a $8k \times 8k$ and a $4k \times 4k$ matrix.

In Figure 4.5, we compare the convergence over time for one optimizer run on

Table 4.1: Numerical performance for the cases considered in Section 4.6.2, taking averages over 10 runs. The (pseudo-)random data are generated for $n = 1000$, and either $k = 20$ or $k = 200$, respectively. The minimum is the respective minimum over all three methods after convergence.

	$k = 20$	200	20	200	20	200
Method	rel. deviation from minimum		iterations		run time (s)	
g^{SpSt}, Geodesic	$1.6614 \cdot 10^{-15}$	$5.4229 \cdot 10^{-16}$	25.5	46.5	0.4091 s	46.4378 s
g^{SpSt}, Cayley	$3.8021 \cdot 10^{-15}$	$1.1783 \cdot 10^{-15}$	25.4	48.6	0.2008 s	9.1229 s
g_ρ from [Gao+21b]	$6.8733 \cdot 10^{-14}$	$1.1445 \cdot 10^{-14}$	32.1	42.5	0.3098 s	11.3064 s

$\mathrm{SpSt}(2000, 40)$. For each step, the run time is measured over one full iteration of the outer for-loop in lines 1 to 24 in Algorithm 4.6.1. It can be seen that gradient descent with respect to the Riemannian metric g^{SpSt} with Cayley stepping converges the fastest in terms of the run time. The iteration count for Cayley and geodesic stepping with respect to g^{SpSt} are comparable.

For Figure 4.6, we repeat the experiment from Figure 4.5 with the setting featured in [Gao+21b, Figure 6], i.e., we scale A to `A = 2*A/norm(A,2)`. In this case, the iteration count until convergence stays approximately the same for gradient descent with respect to the quotient metric g^{SpSt}, while it increases considerably for the canonical-like metric g_ρ.

4.6.3 Gradient descent on the real symplectic Grassmann manifold

In this subsection, we consider optimization via gradient descent on the real symplectic Grassmann manifold. More precisely, we search for the optimal symplectic subspace for representing a given data matrix $S \in \mathbb{R}^{2n\times 2n}$, i.e.,

$$\min_{U\in \mathrm{SpSt}(2n,2k)} ||S - UU^+S||_F^2. \tag{4.6.1}$$

This problem is associated with computing a *proper symplectic decomposition*, a task that is central in Hamiltonian model order reduction [PM16]. Here, we work in an academic setting, where the target matrix S is generated as a random symplectic subspace representative plus an error term, i.e.

$$S = AA^+ + \mathcal{E},$$

where $A \in \mathrm{SpSt}(2n, 2k)$ is a random symplectic Stiefel matrix found in the same manner as the initial point U_0, and $\mathcal{E}$ is a random $n \times n$-matrix, divided by its 2-norm. The parameters for the gradient descent algorithm are the same as in Subsection 4.6.2. The

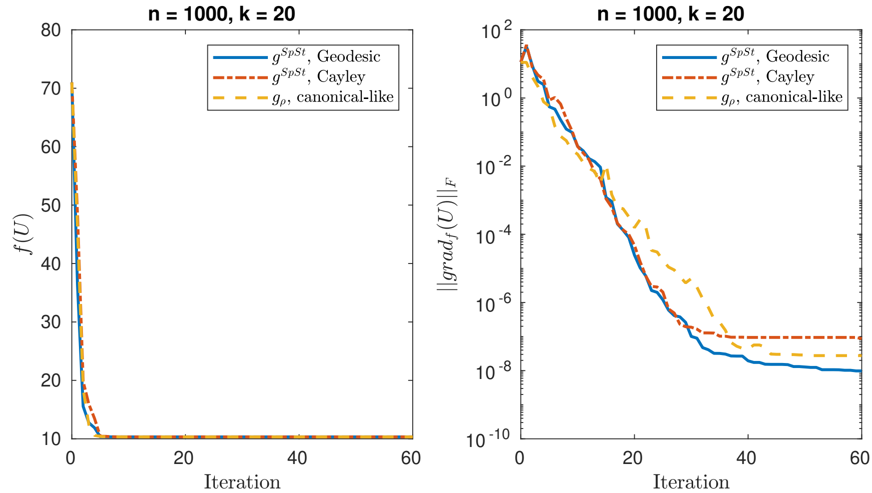

Figure 4.4: (cf. Subsection 4.6.2) Comparison of Riemannian gradient descent on $\mathrm{SpSt}(2000, 40)$ to find the symplectic Stiefel matrix closest to a random matrix A. The data represent an average over 10 runs. Here, g_ρ denotes the canonical-like metric from [Gao+21b] with Cayley stepping.

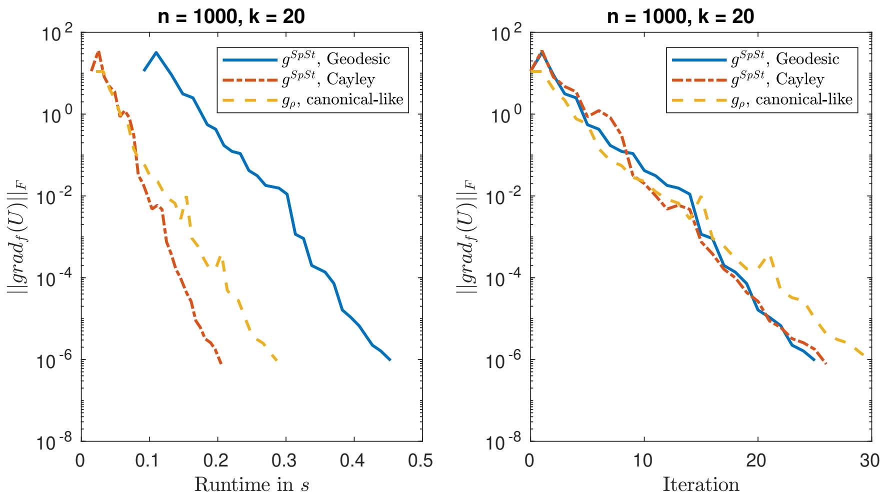

Figure 4.5: (cf. Subsection 4.6.2) Comparison of Riemannian gradient descent on $\mathrm{SpSt}(2000, 40)$ to find the symplectic Stiefel matrix closest to a random matrix A versus time. Here, g_ρ denotes the canonical-like metric from [Gao+21b] with Cayley stepping.

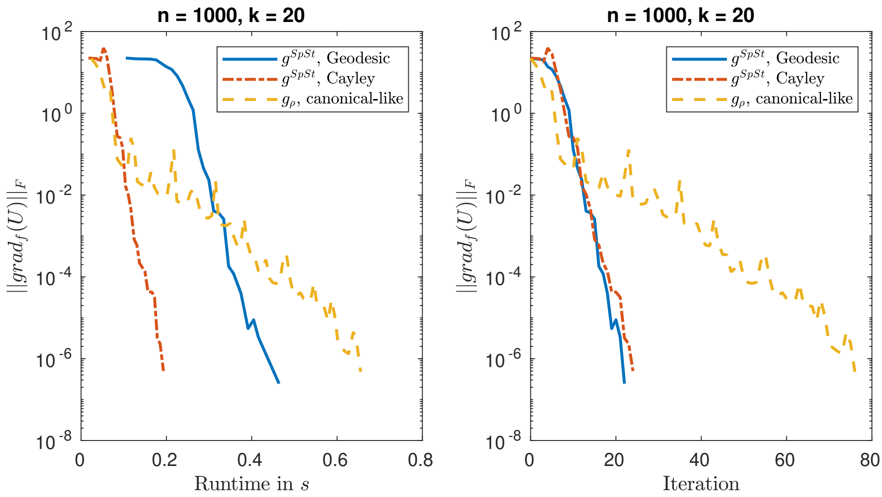

Figure 4.6: (cf. Subsection 4.6.2) Analogous to Figure 4.5 but with the target matrix A scaled to $||A||_2 = 2$. Here, g_ρ denotes the canonical-like metric from [Gao+21b] with Cayley stepping.

resulting average of the function value and the convergence history over 10 runs with a fixed number of 40 iterations is shown in Figure 4.7. It can be seen that gradient descent for all methods produces similar results in regards of the iteration count. For the gradient descent from [Gao+21b] and for the gradient descent according to g^{SpSt} with Cayley stepping, we ignore the quotient structure and treat (4.6.1) as a minimization problem on $\mathrm{SpSt}(2n, 2k)$. The run time and iteration count of the methods is compared in Table 4.2, similarly to Subsection 4.6.2. We also compare the convergence over run time for a single optimizer run in Figure 4.8. It can be seen that gradient descent with Cayley stepping according to g^{SpSt} or g^{SpGr} converges fastest and both methods perform comparable to the method of [Gao+21b]. As is to be expected, for $k = 200$, geodesic stepping is again considerably slower. Note however that for all methods, processing the $2n \times 2n$ input matrix S requires a high base level run time.

Remark. It is also possible to tackle optimization problems with pseudo-Riemannian methods [Fio11; GLY18]. For our experiments however, we achieved better result with the Riemannian methods. Nevertheless, pseudo-Riemannian optimization might prove beneficial in some settings.

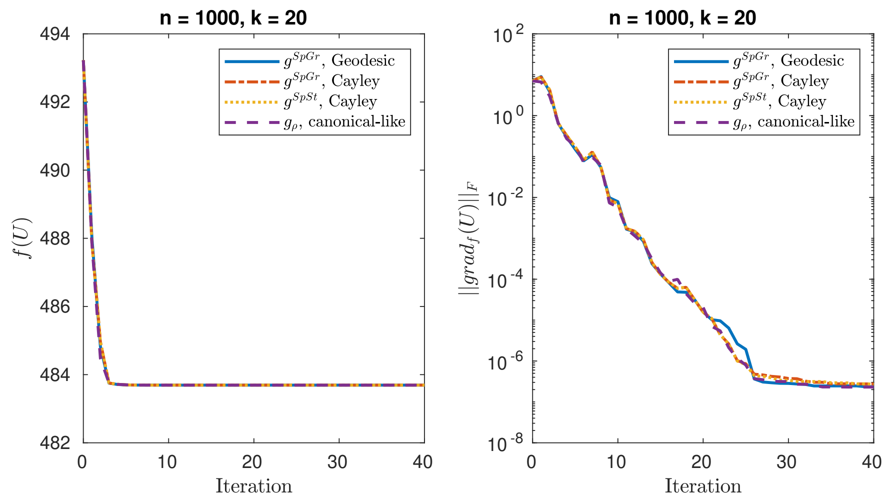

Figure 4.7: (cf. Subsection 4.6.3) Comparison of Riemannian gradient descent on $\mathrm{SpGr}(2000, 40)$ to find the optimal subspace representing a matrix S. The data represent an average over 10 runs. Here, g_ρ denotes the canonical-like metric from [Gao+21b] with Cayley stepping.

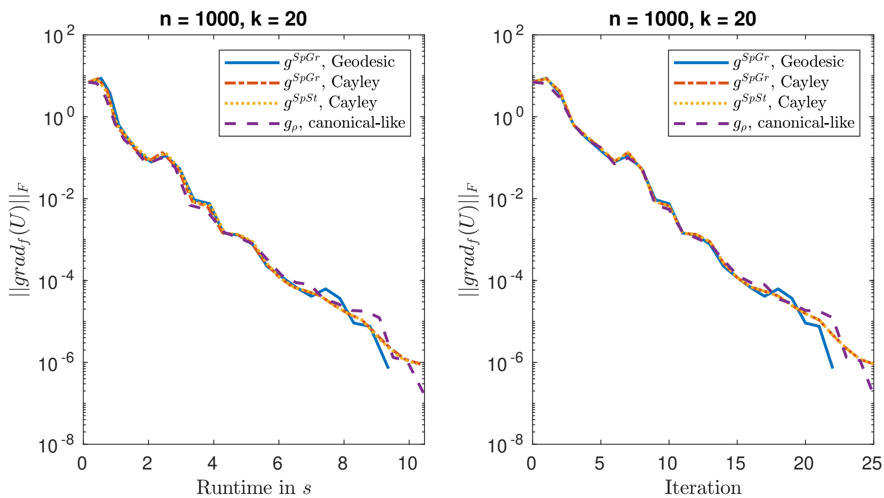

Figure 4.8: (cf. Subsection 4.6.3) Comparison of Riemannian gradient descent on $\mathrm{SpGr}(2000, 40)$ over time to find the optimal subspace representing a matrix S. Here, g_ρ denotes the canonical-like metric from [Gao+21b] with Cayley stepping.

Table 4.2: Numerical performance for the cases considered in Section 4.6.3, taking averages over 10 runs. The (pseudo-)random data are generated for $n = 1000$, and either $k = 20$ or $k = 200$, respectively. The minimum is the respective minimum over all five methods after convergence.

	$k = 20$	200	20	200	20	200
Method	rel. deviation from minimum		iterations		run time (s)	
g^{SpGr}, Geod.	$7.0799 \cdot 10^{-17}$	$6.9878 \cdot 10^{-17}$	25.2	33.7	11.4656	61.9820
g^{SpGr}, Cayley	$2.1071 \cdot 10^{-16}$	$1.0422 \cdot 10^{-16}$	23.9	35.4	10.5513	36.5796
g^{SpSt}, Geod.	$1.0617 \cdot 10^{-16}$	$1.9201 \cdot 10^{-16}$	25.2	34.1	11.4330	64.3123
g^{SpSt}, Cayley	$2.1071 \cdot 10^{-16}$	$1.0422 \cdot 10^{-16}$	23.9	35.4	10.5545	35.3957
g_ρ from [Gao+21b]	$3.5375 \cdot 10^{-17}$	$1.9164 \cdot 10^{-16}$	24.8	34.7	10.9243	37.2282

4.7 Conclusion

We introduced a novel pseudo-Riemannian framework for the real symplectic Stiefel manifold $\mathrm{SpSt}(2n, 2k)$. In analogy to the classical Stiefel and Grassmann manifolds, we introduced the real symplectic Grassmann manifold $\mathrm{SpGr}(2n, 2k)$. For a natural pseudo-Riemannian metric, we derived the corresponding geodesics. With the formulas at hand, we explained the Cayley retraction as an approximation of the pseudo-Riemannian geodesics and found an efficiently computable expression for the retraction, which turned out to be invertible in closed form.

Secondly, we introduced a new Riemannian framework for both $\mathrm{SpSt}(2n, 2k)$ and $\mathrm{SpGr}(2n, 2k)$, coming from a right-invariant Riemannian metric on $\mathrm{Sp}(2n, \mathbb{R})$, and derived the corresponding Riemannian geodesics. Since to the best of the authors' knowledge, the Riemannian geodesics for no other Riemannian metric on $\mathrm{SpSt}(2n, 2k)$ are known, this opens up new possibilities for theoretical studies and applications.

In the experiments, we showed that gradient descent with the Riemannian geodesics or optimized Cayley retraction outperforms the state-of-the-art method from [Gao+21b] in some cases and delivers comparable results in others. Cayley stepping with respect to the Riemannian metric g^{SpSt} on $\mathrm{SpSt}(2n, 2k)$ converges in general the fastest among all methods, regarding the run time.

The invertible retractions provide local coordinates on the manifolds $\mathrm{SpSt}(2n, 2k)$ and $\mathrm{SpGr}(2n, 2k)$, respectively. This renders it possible to apply tangent space methods, e.g. for interpolation purposes. A potential area of application of such tangent space interpolation is parametric model order reduction of Hamiltonian systems. The proposed coordinate transformations allow to approach this problem analogously to parametric model order reduction of general dynamical systems [BGW15; Zim21].

References

[AG01] V. I. Arnol'd and A. B. Givental'. "Symplectic Geometry". In: *Dynamical Systems IV: Symplectic Geometry and its Applications*. Ed. by V. I. Arnold and S. P. Novikov. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 1–138. ISBN: 978-3-662-06791-8. DOI: 10.1007/978-3-662-06791-8_1.

[AH17] B. M. Afkham and J. S. Hesthaven. "Structure preserving model reduction of parametric Hamiltonian systems". In: *SIAM Journal on Scientific Computing* 39.6 (2017), A2616–A2644.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, New Jersey: Princeton University Press, 2008. URL: http://press.princeton.edu/titles/8586.html.

[Bat+15] E. Batzies et al. "Geometric mean and geodesic regression on Grassmannians". In: *Linear Algebra Appl.* 466 (2015), pp. 83–101. ISSN: 0024-3795. DOI: 10.1016/j.laa.2014.10.003.

[BBH19] P. Buchfink, A. Bhatt, and B. Haasdonk. "Symplectic model order reduction with non-orthonormal bases". In: *Mathematical and Computational Applications* 24.2 (2019), p. 43.

[BCC20] P. Birtea, I. Caşu, and D. Comănescu. "Optimization on the real symplectic group". In: *Monatshefte für Mathematik* 191.3 (2020), pp. 465–485.

[BGW15] P. Benner, S. Gugercin, and K. Willcox. "A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems". In: *SIAM Review* 57.4 (2015), pp. 483–531. DOI: 10.1137/130932715.

[BJ15] R. Bhatia and T. Jain. "On symplectic eigenvalues of positive definite matrices". In: *Journal of Mathematical Physics* 56.11 (2015), p. 112201. DOI: 10.1063/1.4935852.

[BZA20] T. Bendokat, R. Zimmermann, and P.-A. Absil. *A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects*. 2020. arXiv: 2011.13699.

[CI00] E. Celledoni and A. Iserles. "Approximating the exponential from a Lie algebra to a Lie group". In: *Mathematics of Computation* 69 (2000), pp. 1457–1480. DOI: 10.1090/S0025-5718-00-01223-0.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith. "The Geometry of Algorithms with Orthogonality Constraints". In: *SIAM Journal on Matrix Analysis and Applications* 20.2 (Apr. 1998), pp. 303–353. ISSN: 0895-4798. DOI: 10.1137/S0895479895290954.

[Fio11] S. Fiori. "Solving minimal-distance problems over the manifold of real-symplectic matrices". In: *SIAM Journal on Matrix Analysis and Applications* 32.3 (2011), pp. 938–968.

[Gao+21a] B. Gao et al. "Geometry of the Symplectic Stiefel Manifold Endowed with the Euclidean Metric". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Cham: Springer International Publishing, 2021, pp. 789–796.

[Gao+21b] B. Gao et al. "Riemannian Optimization on the Symplectic Stiefel Manifold". In: *SIAM Journal on Optimization* 31.2 (2021), pp. 1546–1575. DOI: 10.1137/20M1348522.

[GLY18] T. Gao, L.-H. Lim, and K. Ye. *Semi-Riemannian Manifold Optimization*. 2018. arXiv: 1812.07643 [math.OC].

[HLW06] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*. Second. Vol. 31. Springer Series in Computational Mathematics. Berlin: Springer-Verlag, 2006, pp. xviii+644. ISBN: 3-540-30663-3; 978-3-540-30663-4. DOI: 10.1007/3-540-30666-8.

[KLS19] Y.-C. Kuo, W.-W. Lin, and S.-F. Shieh. "A structure preserving flow for computing Hamiltonian matrix exponential". In: *Numerische Mathematik* 143.3 (2019), pp. 555–582.

[KM02] A. Khvedelidze and D. Mladenov. "Generalized Calogero–Moser–Sutherland models from geodesic motion on GL+(n,R) group manifold". In: *Physics Letters A* 299.5 (2002), pp. 522–530. ISSN: 0375-9601. DOI: 10.1016/S0375-9601(02)00708-9.

[Lee12] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York, 2012. ISBN: 9781441999825. DOI: 10.1007/978-1-4419-9982-5.

[Lee18] J. M. Lee. *Introduction to Riemannian Manifolds*. Vol. 176. Graduate Texts in Mathematics. Springer, Cham, 2018, pp. xiii+437. ISBN: 978-3-319-91754-2; 978-3-319-91755-9. DOI: 10.1007/978-3-319-91755-9.

[ONe83] B. O'Neill. *Semi-Riemannian geometry - With applications to relativity*. Vol. 103. Pure and Applied Mathematics. Academic Press, New York, 1983, pp. xiii+468. ISBN: 0-12-526740-1.

[PM16] L. Peng and K. Mohseni. "Symplectic model reduction of Hamiltonian systems". In: *SIAM Journal on Scientific Computing* 38.1 (2016), A1–A27.

[Sed19] M. Sedano-Mendoza. *Isometry groups of generalized Stiefel manifolds*. 2019. arXiv: 1901.10630 [math.DG].

[Son+21] N. T. Son et al. *Symplectic eigenvalue problem via trace minimization and Riemannian optimization*. 2021. arXiv: 2101.02618 [math.OC].

[VAV13] B. Vandereycken, P.-A. Absil, and S. Vandewalle. "A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank". English. In: *IMA Journal of Numerical Analysis* 33.2 (2013), pp. 481–514.

[WSF18] J. Wang, H. Sun, and S. Fiori. "A Riemannian-steepest-descent approach for optimization on the real symplectic group". In: *Mathematical Methods in the Applied Sciences* 41.11 (2018), pp. 4273–4286.

[Xu03] H. Xu. "An SVD-like matrix decomposition and its applications". In: *Linear Algebra and its Applications* 368 (2003), pp. 1–24. ISSN: 0024-3795. DOI: 10.1016/S0024-3795(03)00370-7.

[Zim21] R. Zimmermann. "Manifold interpolation". In: *System- and Data-Driven Methods and Algorithms*. Ed. by P. Benner et al. Vol. 1. Model Order Reduction. Boston: De Gruyter, 2021.

Bibliography

Bibliography

[AB15] M. M. Alexandrino and R. G. Bettiol. *Lie Groups and Geometric Aspects of Isometric Actions*. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-16613-1. DOI: 10.1007/978-3-319-16613-1_2.

[AEK06] P.-A. Absil, A. Edelman, and P. Koev. "On the largest principal angle between random subspaces". In: *Linear Algebra and its Applications* 414.1 (2006), pp. 288–294. ISSN: 0024-3795. DOI: 10.1016/j.laa.2005.10.004.

[AF08] D. Amsallem and C. Farhat. "Interpolation Method for Adapting Reduced-Order Models and Application to Aeroelasticity". In: *AIAA Journal* 46.7 (2008), pp. 1803–1813. DOI: 10.2514/1.35374.

[Afk+18] B. M. Afkham et al. "Symplectic model-reduction with a weighted inner product". In: *arXiv preprint arXiv:1803.07799* (2018).

[AG01] V. I. Arnol'd and A. B. Givental'. "Symplectic Geometry". In: *Dynamical Systems IV: Symplectic Geometry and its Applications*. Ed. by V. I. Arnold and S. P. Novikov. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 1–138. ISBN: 978-3-662-06791-8. DOI: 10.1007/978-3-662-06791-8_1.

[AH17] B. M. Afkham and J. S. Hesthaven. "Structure preserving model reduction of parametric Hamiltonian systems". In: *SIAM Journal on Scientific Computing* 39.6 (2017), A2616–A2644.

[AMS04] P.-A. Absil, R. Mahony, and R. Sepulchre. "Riemannian Geometry of Grassmann Manifolds with a View on Algorithmic Computation". English. In: *Acta Applicandae Mathematica* 80.2 (2004), pp. 199–220. ISSN: 0167-8019. DOI: 10.1023/B:ACAP.0000013855.14971.91.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton, New Jersey: Princeton University Press, 2008. URL: http://press.princeton.edu/titles/8586.html.

[ATV13] B. Afsari, R. Tron, and R. Vidal. "On the Convergence of Gradient Descent for Finding the Riemannian Center of Mass". In: *SIAM Journal on Control and Optimization* 51.3 (2013), pp. 2230–2260. DOI: 10.1137/12086282X.

[BA15] N. Boumal and P.-A. Absil. "Low-rank matrix completion via preconditioned optimization on the Grassmann manifold". In: *Linear Algebra and its Applications* 475 (2015), pp. 200–239. DOI: 10.1016/j.laa.2015.02.027.

[Bat+15] E. Batzies et al. "Geometric mean and geodesic regression on Grassmannians". In: *Linear Algebra Appl.* 466 (2015), pp. 83–101. ISSN: 0024-3795. DOI: 10.1016/j.laa.2014.10.003.

[BBH19] P. Buchfink, A. Bhatt, and B. Haasdonk. "Symplectic model order reduction with non-orthonormal bases". In: *Mathematical and Computational Applications* 24.2 (2019), p. 43.

[BCC20] P. Birtea, I. Caşu, and D. Comănescu. "Optimization on the real symplectic group". In: *Monatshefte für Mathematik* 191.3 (2020), pp. 465–485.

[Ber97] S. Berceanu. "On the geometry of complex Grassmann manifold, its noncompact dual and coherent states". In: *Bull. Belg. Math. Soc. Simon Stevin* 4.2 (1997), pp. 205–243. ISSN: 1370-1444. DOI: 10.36045/bbms/1105731655.

[BG18] R. Bergmann and P.-Y. Gousenbourger. "A Variational Model for Data Fitting on Manifolds by Minimizing the Acceleration of a Bézier Curve". In: *Frontiers in Applied Mathematics and Statistics* 4 (2018), p. 59. ISSN: 2297-4687. DOI: 10.3389/fams.2018.00059.

[BGW15] P. Benner, S. Gugercin, and K. Willcox. "A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems". In: *SIAM Review* 57.4 (2015), pp. 483–531. DOI: 10.1137/130932715.

[BJ15] R. Bhatia and T. Jain. "On symplectic eigenvalues of positive definite matrices". In: *Journal of Mathematical Physics* 56.11 (2015), p. 112201. DOI: 10.1063/1.4935852.

[BKL17] J. Batista, K. Krakowski, and F. S. Leite. "Exploring quasi-geodesics on Stiefel manifolds in order to smooth interpolate between domains". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. 2017, pp. 6395–6402.

[BN91] A. A. Borisenko and Y. A. Nikolaevskiĭ. "Grassmann manifolds and Grassmann image of submanifolds". In: *Uspekhi Mat. Nauk* 46.2(278) (1991), pp. 41–83, 240. ISSN: 0042-1316. DOI: 10.1070/RM1991v046n02ABEH002742.

[Bou14] N. Boumal. "Optimization and estimation on manifolds". PhD thesis. Université catholique de Louvain, 2014.

[BW15] L. Balzano and S. J. Wright. "Local Convergence of an Algorithm for Subspace Identification from Partial Data". In: *Foundations of Computational Mathematics* 15.5 (2015), pp. 1279–1314. DOI: 10.1007/s10208-014-9227-7.

[BZ21a] T. Bendokat and R. Zimmermann. "Efficient Quasi-Geodesics on the Stiefel Manifold". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Cham: Springer International Publishing, 2021, pp. 763–771. ISBN: 978-3-030-80209-7. DOI: 10.1007/978-3-030-80209-7_82.

[BZ21b] T. Bendokat and R. Zimmermann. *The real symplectic Stiefel and Grassmann manifolds: metrics, geodesics and applications*. 2021. arXiv: 2108.12447 [math.DG].

[BZA20] T. Bendokat, R. Zimmermann, and P.-A. Absil. *A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects*. 2020. arXiv: 2011.13699.

[Car92] M. P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, 1992. ISBN: 9780817634902.

[CI00] E. Celledoni and A. Iserles. "Approximating the exponential from a Lie algebra to a Lie group". In: *Mathematics of Computation* 69 (2000), pp. 1457–1480. DOI: 10.1090/S0025-5718-00-01223-0.

[CV19] R. Chakraborty and B. C. Vemuri. "Statistics on the Stiefel manifold: Theory and applications". In: *The Annals of Statistics* 47.1 (2019), pp. 415–438. DOI: 10.1214/18-AOS1692.

[DE99] L. Dieci and T. Eirola. "On Smooth Decompositions of Matrices". In: *SIAM Journal on Matrix Analysis and Applications* 20.3 (1999), pp. 800–819. DOI: 10.1137/S0895479897330182.

[DVW10] J. Degroote, J. Vierendeels, and K. Willcox. "Interpolation among reduced-order matrices to obtain parameterized models for design, optimization and probabilistic analysis". In: *International Journal for Numerical Methods in Fluids* 63.2 (2010), pp. 207–230. ISSN: 1097-0363. DOI: 10.1002/fld.2089. URL: http://dx.doi.org/10.1002/fld.2089.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith. "The Geometry of Algorithms with Orthogonality Constraints". In: *SIAM Journal on Matrix Analysis and Applications* 20.2 (Apr. 1998), pp. 303–353. ISSN: 0895-4798. DOI: 10.1137/S0895479895290954.

[Fio11] S. Fiori. "Solving minimal-distance problems over the manifold of real-symplectic matrices". In: *SIAM Journal on Matrix Analysis and Applications* 32.3 (2011), pp. 938–968.

[Fle13] P. T. Fletcher. "Geodesic regression and the theory of least squares on Riemannian manifolds". In: *International journal of computer vision* 105.2 (2013), pp. 171–185. DOI: 10.1007/s11263-012-0591-y.

[Gal+03] K. A. Gallivan et al. "Efficient algorithms for inferences on Grassmann manifolds". In: *IEEE Workshop on Statistical Signal Processing*. 2003, pp. 315–318. DOI: 10.1109/SSP.2003.1289408.

[Gal11] J. H. Gallier. *Geometric Methods and Applications: For Computer Science and Engineering*. Texts in Applied Mathematics. New York: Springer, 2011. ISBN: 0-387-95044-3. DOI: 10.1007/978-1-4419-9961-0.

[Gao+21a] B. Gao et al. "Geometry of the Symplectic Stiefel Manifold Endowed with the Euclidean Metric". In: *Geometric Science of Information*. Ed. by F. Nielsen and F. Barbaresco. Cham: Springer International Publishing, 2021, pp. 789–796.

[Gao+21b] B. Gao et al. "Riemannian Optimization on the Symplectic Stiefel Manifold". In: *SIAM Journal on Optimization* 31.2 (2021), pp. 1546–1575. DOI: 10.1137/20M1348522.

[GLY18] T. Gao, L.-H. Lim, and K. Ye. *Semi-Riemannian Manifold Optimization*. 2018. arXiv: 1812.07643 [math.OC].

[GMA19] P.-Y. Gousenbourger, E. Massart, and P.-A. Absil. "Data fitting on manifolds with composite Bézier-like curves and blended cubic splines". In: *Journal of Mathematical Imaging and Vision* 61.5 (2019), pp. 645–671.

[GR17] R. Godement and U. Ray. *Introduction to the Theory of Lie Groups*. Universitext. Springer International Publishing, 2017. ISBN: 9783319543734. DOI: 10.1007/978-3-319-54375-8.

[Hal15] B. C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. 2nd. Springer Graduate texts in Mathematics. New York – Berlin – Heidelberg: Springer–Verlag, 2015. DOI: 10.1007/978-3-319-13467-3.

[HBP09] A. Hay, J. T. Borggaard, and D. Pelletier. "Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition". In: *Journal of Fluid Mechanics* 629 (2009), pp. 41–72. DOI: 10.1017/S0022112009006363.

[Hel01] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. Crm Proceedings & Lecture Notes. American Mathematical Society, 2001. ISBN: 9780821828489.

[HHT07] U. Helmke, K. Hüper, and J. Trumpf. *Newton's method on Graßmann manifolds*. 2007. eprint: 0709.2205 (math.OC).

[Hig08] N. J. Higham. *Functions of Matrices: Theory and Computation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2008, pp. xx+425. ISBN: 978-0-898716-46-7. DOI: 10.1137/1.9780898717778.

[HJ91] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991, pp. viii+607. ISBN: 0-521-30587-X. DOI: 10.1017/CBO9780511840371.

[HLW06] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*. Second. Vol. 31. Springer Series in Computational Mathematics. Berlin: Springer-Verlag, 2006, pp. xviii+644. ISBN: 3-540-30663-3; 978-3-540-30663-4. DOI: 10.1007/3-540-30666-8.

[HM94] U. Helmke and J. B. Moore. *Optimization and Dynamical Systems*. Communications & Control Engineering. London: Springer–Verlag, 1994. ISBN: 978-1-4471-3469-5. DOI: 10.1007/978-1-4471-3467-1.

[JMS19] V. Jurdjevic, I. Markina, and F. Silva Leite. "Extremal Curves on Stiefel and Grassmann Manifolds". English. In: *The Journal of Geometric Analysis* (2019).

[KL07] O. Koch and C. Lubich. "Dynamical Low-Rank Approximation". In: *SIAM J. Matrix Analysis Applications* 29 (2007), pp. 434–454. DOI: 10.1137/050639703.

[KLS19] Y.-C. Kuo, W.-W. Lin, and S.-F. Shieh. "A structure preserving flow for computing Hamiltonian matrix exponential". In: *Numerische Mathematik* 143.3 (2019), pp. 555–582.

[KM02] A. Khvedelidze and D. Mladenov. "Generalized Calogero–Moser–Sutherland models from geodesic motion on GL+(n,R) group manifold". In: *Physics Letters A* 299.5 (2002), pp. 522–530. ISSN: 0375-9601. DOI: 10.1016/S0375-9601(02)00708-9.

[KN96] S. Kobayashi and K. Nomizu. *Foundations of differential geometry. Vol. I & II.* Wiley Classics Library. Reprint of the 1963 original. John Wiley & Sons, Inc., New York, 1996.

[Koz01] E. S. Kozlov. "Geometry of Real Grassmann Manifolds. Part I - VI". In: *Journal of Mathematical Sciences* (2000 – 2001).

[Kra+17] K. A. Krakowski et al. "A modified Casteljau algorithm to solve interpolation problems on Stiefel manifolds". In: *Journal of Computational and Applied Mathematics* 311 (2017), pp. 84–99. ISSN: 0377-0427.

[Lee12] J. M. Lee. *Introduction to Smooth Manifolds.* Graduate Texts in Mathematics. Springer New York, 2012. ISBN: 9781441999825. DOI: 10.1007/978-1-4419-9982-5.

[Lee18] J. M. Lee. *Introduction to Riemannian Manifolds.* Vol. 176. Graduate Texts in Mathematics. Springer, Cham, 2018, pp. xiii+437. ISBN: 978-3-319-91754-2; 978-3-319-91755-9. DOI: 10.1007/978-3-319-91755-9.

[Lei61] K. Leichtweiss. "Zur Riemannschen Geometrie in Grassmannschen Mannigfaltigkeiten". In: *Math. Z.* 76 (1961), pp. 334–366. ISSN: 0025-5874. DOI: 10.1007/BF01210982.

[Li+19] C. Li et al. *Convergence Analysis of Gradient Algorithms on Riemannian Manifolds Without Curvature Constraints and Application to Riemannian Mass.* 2019. eprint: 1910.02280 (math.OC).

[LLY20] Z. Lai, L.-H. Lim, and K. Ye. *Simpler Grassmannian optimization.* 2020. eprint: 2009.13502 (math.OC).

[Man12] Y. Man Lui. "Advances in matrix manifolds for computer vision". In: *Image and Vision Computing* 30.6–7 (2012), pp. 380–388. ISSN: 0262-8856. DOI: 10.1016/j.imavis.2011.08.002.

[MLB20] L. Machado, F. S. Leite, and E. Batzies. “Geometric Algorithm to Generate Interpolating Splines on Grassmann and Stiefel Manifolds”. English. In: *CONTROLO 2020*. Cham: Springer International Publishing, 2020, pp. 180–189. ISBN: 1876-1100.

[MM16] 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. Cham: Springer International Publishing, 2016. ISBN: 3319450255, 9783319450254. DOI: 10.1007/978-3-319-45026-1.

[MS85] A. Machado and I. Salavessa. “Grassmannian manifolds as subsets of Euclidean spaces”. In: *Differential geometry (Santiago de Compostela, 1984)*. Vol. 131. Res. Notes in Math. Pitman, Boston, MA, 1985, pp. 85–102.

[NA05] Y. Nishimori and S. Akaho. “Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold”. In: *Neurocomputing* 67 (2005), pp. 106–135. ISSN: 0925-2312.

[Ngu13] T. S. Nguyen. “A real time procedure for affinely dependent parametric model order reduction using interpolation on Grassmann manifolds”. In: *International Journal for Numerical Methods in Engineering* 93.8 (2013), pp. 818–833. DOI: 10.1002/nme.4408.

[NS15] T. Nguyen and T. Stykel. “Model order reduction of parameterized circuit equations based on interpolation”. In: *Adv. Comput. Math* 41 (2015), pp. 1321–1342. DOI: 10.1007/s10444-015-9418-z.

[ONe83] B. O’Neill. *Semi-Riemannian geometry - With applications to relativity*. Vol. 103. Pure and Applied Mathematics. Academic Press, New York, 1983, pp. xiii+468. ISBN: 0-12-526740-1.

[PM16] L. Peng and K. Mohseni. “Symplectic model reduction of Hamiltonian systems”. In: *SIAM Journal on Scientific Computing* 38.1 (2016), A1–A27.

[QZL05] L. Qiu, Y. Zhang, and C.-K. Li. “Unitarily Invariant Metrics on the Grassmann Space”. In: *SIAM Journal on Matrix Analysis and Applications* 27.2 (2005), pp. 507–25. DOI: 10.1137/040607605.

[Rah+05] I. U. Rahman et al. “Multiscale Representations for Manifold-Valued Data”. In: *SIAM Journal on Multiscale Modeling and Simulation* 4.4 (2005), pp. 1201–1232. DOI: 10.1137/050622729.

[Ren13] Q. Rentmeesters. “Algorithms for data fitting on some common homogeneous spaces”. PhD thesis. Louvain, Belgium: Université Catholique de Louvain, 2013.

[Ros06] W. Rossmann. *Lie Groups: An Introduction Through Linear Groups*. Oxford Graduate Texts in Mathematics. Oxford University Press, 2006. ISBN: 9780199202515.

[Saa92] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Algortihms and Architectures for Advanced Scientific Computing. Manchester, UK: Manchester University Press, 1992.

[Sak77] T. Sakai. “On cut loci of compact symmetric spaces”. In: *Hokkaido Math. J.* 6.1 (1977), pp. 136–161. ISSN: 0385-4035. DOI: 10.14492/hokmj/1381758555.

[Sak96] T. Sakai. *Riemannian Geometry*. Fields Institute Communications. American Mathematical Soc., 1996. ISBN: 9780821889565. DOI: 10.1090/mmono/149.

[Sed19] M. Sedano-Mendoza. *Isometry groups of generalized Stiefel manifolds*. 2019. arXiv: 1901.10630 [math.DG].

[SL05] A. Srivastava and X. Liu. “Tools for application-driven linear dimension reduction”. In: *Neurocomputing* 67 (2005). Geometrical Methods in Neural Networks and Learning, pp. 136–160. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2004.11.036.

[Son+21] N. T. Son et al. *Symplectic eigenvalue problem via trace minimization and Riemannian optimization*. 2021. arXiv: 2101.02618 [math.OC].

[ST15] A. Srivastava and P. K. Turaga. *Riemannian computing in computer vision*. English (US). Springer International Publishing, 2015. ISBN: 978-3-319-22957-7. DOI: 10.1007/978-3-319-22957-7.

[UM14] K. Usevich and I. Markovsky. “Optimization on a Grassmann manifold with application to system identification”. In: *Automatica* 50.6 (2014), pp. 1656–1662. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2014.04.010.

[VAV13] B. Vandereycken, P.-A. Absil, and S. Vandewalle. “A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank”. English. In: *IMA Journal of Numerical Analysis* 33.2 (2013), pp. 481–514.

[WC88] G. L. Wu and W. H. Chen. “A matrix inequality and its geometric applications”. In: *Acta Math. Sinica* 31.3 (1988), pp. 348–355. ISSN: 0583-1431.

[WLL12] S. F. Walter, L. Lehmann, and R. Lamour. “On evaluating higher-order derivatives of the QR decomposition of tall matrices with full column rank in forward and reverse mode algorithmic differentiation”. In: *Optimization Methods and Software* 27.2 (2012), pp. 391–403. DOI: 10.1080/10556788.2011.610454.

[Won67] Y.-C. Wong. “Differential Geometry of Grassmann Manifolds”. In: *Proceedings of the National Academy of Sciences of the United States of America* 57 (1967), pp. 589–594. DOI: 10.1073/pnas.57.3.589.

[Won68a] Y.-C. Wong. "Conjugate loci in Grassmann manifolds". In: *Bull. Amer. Math. Soc.* 74 (1968), pp. 240–245. ISSN: 0002-9904. DOI: 10.1090/S0002-9904-1968-11903-2.

[Won68b] Y.-C. Wong. "Sectional Curvatures of Grassmann Manifolds". In: *Proceedings of the National Academy of Sciences of the United States of America* 60.1 (1968), pp. 75–79. ISSN: 00278424. DOI: 10.1073/pnas.60.1.75.

[WSF18] J. Wang, H. Sun, and S. Fiori. "A Riemannian-steepest-descent approach for optimization on the real symplectic group". In: *Mathematical Methods in the Applied Sciences* 41.11 (2018), pp. 4273–4286.

[Xu03] H. Xu. "An SVD-like matrix decomposition and its applications". In: *Linear Algebra and its Applications* 368 (2003), pp. 1–24. ISSN: 0024-3795. DOI: 10.1016/S0024-3795(03)00370-7.

[YL16] K. Ye and L.-H. Lim. "Schubert varieties and distances between subspaces of different dimensions". In: *SIAM J. Matrix Anal. Appl.* 37.3 (2016), pp. 1176–1197. ISSN: 0895-4798. DOI: 10.1137/15M1054201.

[ZB16] D. Zhang and L. Balzano. "Global Convergence of a Grassmannian Gradient Descent Algorithm for Subspace Estimation". In: Cadiz, Spain: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS, May 2016, pp. 1460–1468.

[ZD17] R. Zimmermann and K. Debrabant. "Parametric model reduction via interpolating orthonormal bases". In: *European Conference on Numerical Mathematics and Advanced Applications*. Springer. 2017, pp. 683–691.

[Zim14] R. Zimmermann. "A Locally Parametrized Reduced Order Model for the Linear Frequency Domain Approach to Time-Accurate Computational Fluid Dynamics". In: *SIAM Journal on Scientific Computing* 36.3 (2014), B508–B537. DOI: 10.1137/130942462.

[Zim17] R. Zimmermann. "A Matrix-Algebraic Algorithm for the Riemannian Logarithm on the Stiefel Manifold under the Canonical Metric". In: *SIAM J. Matrix. Anal. Appl.* 38.2 (2017), pp. 322–342.

[Zim19] R. Zimmermann. *Manifold interpolation and model reduction*. 2019. arXiv: 1902.06502 [math.NA].

[Zim20] R. Zimmermann. "Hermite Interpolation and Data Processing Errors on Riemannian Matrix Manifolds". In: *SIAM Journal on Scientific Computing* 42.5 (2020), A2593–A2619. DOI: 10.1137/19M1282878.

[Zim21] R. Zimmermann. "Manifold interpolation". In: *System- and Data-Driven Methods and Algorithms*. Ed. by P. Benner et al. Vol. 1. Model Order Reduction. Boston: De Gruyter, 2021.

[ZPW18] R. Zimmermann, B. Peherstorfer, and K. Willcox. “Geometric subspace updates with applications to online adaptive nonlinear model reduction”. In: *SIAM Journal on Matrix Analysis and Application* 39.1 (2018), pp. 234–261. DOI: 10.1137/17M1123286.