Manifold interpolation

Zimmermann, Ralf

Published in:
Model Order Reduction

DOI:
10.1515/9783110498967-007

Publication date:
2021

Document version:
Final published version

Ralf Zimmermann

7 Manifold interpolation

Abstract: One approach to parametric and adaptive model reduction is via the interpolation of orthogonal bases, subspaces or positive definite system matrices. In all these cases, the sampled inputs stem from matrix sets that feature a geometric structure and thus form so-called matrix manifolds. This chapter reviews the numerical treatment of the most important matrix manifolds that arise in the context of model reduction. Moreover, the principal approaches to data interpolation and Taylor-like extrapolation on matrix manifolds are outlined and complemented by algorithms in pseudo-code.

Keywords: parametric model reduction, matrix manifold, interpolation, Riemannian computing, Riemannian normal coordinates

MSC 2010: 15-01, 15A16, 15B10, 15B48, 53-04, 65F60, 41-01, 41A05, 65F99, 93A15, 93C30

7.1 Introduction & motivation

This chapter addresses interpolation approaches for parametric model reduction. This includes techniques for

- computing trajectories of parameterized subspaces,
- computing trajectories of parameterized reduced orthogonal bases,
- structure-preserving interpolation.

Mathematically, this requires data processing on nonlinear matrix manifolds. The exposition at hand intends to be an introduction and a reference guide to numerical procedures with matrix manifold-valued data. As such it addresses practitioners and scientists new to the field. It covers the essentials of those matrix manifolds that arise most frequently in practical problems in model reduction. The main purpose is not to discuss concrete model reduction applications, but rather to provide the essential tools, building blocks and background theory to enable the reader to devise her/his own approaches for such applications.

The text was designed such that it works as a commented formula collection, meanwhile giving sufficient context, explanations and, not least, precise references to enable the interested reader to immerse further in the topic.

Acknowledgement: We are grateful to ECOST for pushing us into this.

Ralf Zimmermann, SDU Odense, Odense, Denmark

https://doi.org/10.1515/9783110498967-007

7.1.1 Parametric model reduction via manifold interpolation: an introductory example

The basic objective in model reduction is to emulate a large-scale dynamical system with very few degrees of freedom such that its input/output behavior is preserved as well as possible. While classical model reduction techniques aim at producing an accurate low-order approximation to the autonomous behavior of the original system, parametric model reduction (pMOR) tries to account for additional system parameters. If we look for instance at aircraft aerodynamics, an important task is to solve the unsteady Navier–Stokes equations at various flight conditions, which are, amongst others, specified by the altitude, the viscosity of the fluid (i. e. the Reynolds number) and the relative velocity (i. e. the Mach number). We explain the objective of pMOR with the aid of a generic example in the context of proper orthogonal decomposition-based model reduction. Similar considerations apply to frequency domain approaches, Krylov subspace methods and balanced truncation, which are, e. g., discussed Chapters 2 and 3 of this volume and in [19, Chapter 1], [20, Chapter 3].

Consider a spatio-temporal dynamical system in semi-discrete form

$$\frac{\partial}{\partial t}x(t,\mu) = f(x(t,\mu);\mu), \quad x(t_0,\mu) = x_{0,\mu}, \tag{7.1}$$

where $x(t,\mu) \in \mathbb{R}^n$ is the spatially discretized *state vector* of dimension n, the vector $\mu = (\mu_1,\dots,\mu_d) \in \mathbb{R}^d$ accounts for additional system parameters and $f(\cdot;\mu) : \mathbb{R}^n \to \mathbb{R}^n$ is the (possibly nonlinear, parameter-dependent) right hand side function. Projection-based MOR starts with constructing a suitable low-dimensional subspace that acts as a space of candidate solutions.

Subspace construction. One way to construct the required projection subspace is the proper orthogonal decomposition (POD) see [19, Chapter 2]. In its simplest form, the POD can be summarized as follows. For a fixed system parameter $\mu = \mu_0$, let $x^1 := x(t_1,\mu_0),\dots,x^m := x(t_m,\mu_0) \in \mathbb{R}^n$ be a set of state vectors satisfying (7.1) and let $\mathbb{S} := (x^1,\dots,x^m) \in \mathbb{R}^{n\times m}$. The state vectors x^i are called *snapshots* and the matrix $\mathbb{S}$ is called the associated *snapshot matrix*. POD is concerned with finding a subspace $\mathcal{V}$ of dimension $r \le m$ represented by a column-orthogonal matrix $\mathbb{V}_r \in \mathbb{R}^{n\times r}$ such that the error between the input snapshots and their orthogonal projection onto $\mathcal{V} = \mathrm{ran}(\mathbb{V}_r)$ is minimized:

$$\min_{V\in\mathbb{R}^{n\times r}, V^TV=I} \sum_k \left\|x^k - VV^Tx^k\right\|_2^2 \quad \left(\Leftrightarrow \min_{V\in\mathbb{R}^{n\times r}, V^TV=I} \left\|\mathbb{S} - VV^T\mathbb{S}\right\|_F^2\right).$$

The main result of POD is that, for any $r \le m$, the best r-dimensional approximation of $\mathrm{ran}(x^1,\dots,x^m)$ in the above sense is $\mathcal{V} = \mathrm{ran}(v^1,\dots,v^r)$, where $\{v^1,\dots,v^r\}$ are the eigenvectors of the matrix $\mathbb{S}\mathbb{S}^T$ corresponding to the r largest eigenvalues. The subspace $\mathcal{V}$ is called the *POD subspace* and the matrix $\mathbb{V}_r = (v^1,\dots,v^r)$ is the *POD basis matrix*. The same subspace is obtained via a compact singular value decomposition (SVD) of

the snapshot matrix $\mathbb{S} = \mathbb{V}\Sigma\mathbb{Z}^T$, truncated to the first $r \leq m$ columns of $\mathbb{V} \in \mathbb{R}^{n\times m}$ by setting $\mathcal{V} := \operatorname{ran}(\mathbb{V}_r)$. For more details, see, e. g., [18, §3.3]. In the following, we drop the index r and assume that $\mathbb{V}$ is already the truncated matrix $\mathbb{V} = (v^1, \dots, v^r) \in \mathbb{R}^{n\times r}$.

Since the input snapshots are supplied at a fixed system parameter vector μ_0, the POD subspace is considered to be an appropriate space of solution candidates $\mathcal{V}(\mu_0) = \operatorname{ran}(\mathbb{V}(\mu_0))$ at μ_0.

Projection. POD leads to a parameter decoupling

$$\tilde{x}(t, \mu_0) = \mathbb{V}(\mu_0) x_r(t). \tag{7.2}$$

In this way, the time trajectory of the reduced model is uniquely defined by the coefficient vector $x_r(t) \in \mathbb{R}^r$ that represents the reduced state vector with respect to the subspace $\operatorname{ran}(\mathbb{V}(\mu_0))$. Given a matrix $\mathbb{W}(\mu_0)$ such that the matrix pair $\mathbb{V}(\mu_0), \mathbb{W}(\mu_0)$ is bi-orthogonal, i. e. $\mathbb{W}(\mu_0)^T\mathbb{V}(\mu_0) = I$, the original system (7.1) can be reduced in dimension as follows. Substituting (7.2) in (7.1) and multiplying with $\mathbb{W}(\mu_0)^T$ from the left leads to

$$\frac{d}{dt} x_r(t) = \mathbb{W}^T(\mu_0) f\big(\mathbb{V}(\mu_0) x_r(t); \mu_0\big), \quad x_r(t_0) = \mathbb{V}^T(\mu_0) x_{0,\mu_0}. \tag{7.3}$$

This approach goes by the name of Petrov–Galerkin projection, if $\mathbb{W}(\mu_0) \neq \mathbb{V}(\mu_0)$ and Galerkin projection if $\mathbb{W}(\mu_0) = \mathbb{V}(\mu_0)$. There are various ways to proceed from (7.3) depending on the nature of the function f and many of them are discussed in other chapters of Model Order Reduction. If $f(\cdot; \mu_0)$ is linear, the reduced operator $\mathbb{W}^T(\mu_0) \circ f(\cdot; \mu_0) \circ \mathbb{V}(\mu_0)$ can be computed a priori (*'offline'*) and stays fixed throughout the time integration. If $f(\cdot; \mu_0)$ is affine, the same approach can be carried over to the affine building blocks of $f(\cdot; \mu_0)$; see e. g. [45]. For a nonlinear $f(\cdot; \mu_0)$, an affine approximation can be constructed via the empirical interpolation method (EIM, [14]). Other approaches that address nonlinearities include the discrete empirical interpolation method (DEIM, [30]) and the missing point estimation (MPE, [13, 105]).

For illustration purposes, we proceed with $\mathbb{W}(\mu_0) = \mathbb{V}(\mu_0)$ and assume that the right hand side function f splits into a linear and a nonlinear part: $f(x; \mu_0) = A(\mu_0)x + \mathbf{f}(x; \mu_0)$, where $A(\mu_0) \in \mathbb{R}^{n\times n}$ is, say, a symmetric and negative definite matrix to foster stability. Then (7.3) becomes

$$\frac{d}{dt} x_r(t) = \mathbb{V}^T(\mu_0) A(\mu_0) \mathbb{V}(\mu_0) x_r(t) + \mathbb{V}^T(\mu_0) \mathbf{f}\big(\mathbb{V}(\mu_0) x_r(t); \mu_0\big).$$

In the discrete empirical interpolation method (DEIM, [30]), the large-scale nonlinear term $\mathbf{f}(\mathbb{V}(\mu_0)x_r(t); \mu_0)$ is approximated via a mask matrix $P = (e_{i_1}, \dots, e_{i_s}) \in \mathbb{R}^{n\times s}$, where $\{i_1, \dots, i_s\} \subset \{1, \dots, n\}$ and $e_j = (\dots, \overset{j}{1}, \dots)^T \in \mathbb{R}^n$ is the jth canonical unit vector. The mask matrix P acts as an entry selector on a given n-vector via $P^T v = (v_{i_1}, \dots, v_{i_s})^T \in \mathbb{R}^s$. In addition, another POD basis matrix $\mathbb{U}(\mu_0) \in \mathbb{R}^{n\times s}$ is used, which is obtained from snapshots of the nonlinear term. The matrices P and $\mathbb{U}(\mu_0)$ are combined to form an

oblique projection of the nonlinear term onto the subspace $\operatorname{ran}(\mathbb{U}(\mu_0))$. This leads to the reduced model

$$\begin{aligned}\frac{d}{dt}x_r(t) &= \mathbb{V}^T(\mu_0)A(\mu_0)\mathbb{V}(\mu_0)x_r(t)\\ &\quad + \mathbb{V}^T(\mu_0)\mathbb{U}(\mu_0)\left(P^T\mathbb{U}(\mu_0)\right)^{-1}P^T\mathbf{f}\left(\mathbb{V}(\mu_0)x_r(t);\mu_0\right),\end{aligned} \tag{7.4}$$

whose computational complexity is formally independent of the full-order dimension n; see [30] for details. Mind that by assumption, $M(\mu_0) := -\mathbb{V}^T(\mu_0)A(\mu_0)\mathbb{V}(\mu_0)$ is symmetric positive definite and that both $\mathbb{V}(\mu_0)$ and $\mathbb{U}(\mu_0)$ are column-orthogonal. Moreover, for a fixed mask matrix P, coordinate changes of $\mathbb{V}(\mu_0)$ and $\mathbb{U}(\mu_0)$ do not affect the approximated state $\tilde{x}(t,\mu_0) = \mathbb{V}(\mu_0)x_r(t)$, so that essentially, the reduced system (7.4) depends only on the subspaces $\operatorname{ran}(\mathbb{V}(\mu_0))$ and $\operatorname{ran}(\mathbb{U}(\mu_0))$ rather than the matrices $\mathbb{V}(\mu_0)$ and $\mathbb{U}(\mu_0)$.[1]

Solving (7.3), (7.4) constitutes the *online stage* of model reduction. The main focus of this chapter is *not* on the efficient solution of the reduced systems (7.3) or (7.4) at a fixed μ_0, but on tackling parametric variations in μ. In view of the associated computational costs, it is important that this can be achieved *without* computing additional snapshots in the online stage.

A straightforward way to achieve this is to extend the snapshot sampling to the μ-parameter range to produce POD basis matrices that are to cover all input parameters. This is usually referred to as the "global approach". For nonlinear systems, the global approach may suffer from requiring a large number of snapshot samples. Moreover, the snapshot information is blurred in the global POD and features that occur only in a restricted regime affect the ROM predictions everywhere. Therefore, localized approaches are preferable; see e. g. the applications in and the numerical examples in [38, 100].

In this chapter, the focus is on constructing trajectories of functions in the system parameters μ on certain sets of structured matrix spaces. In the above example, these are the symmetric positive definite matrices $\{M \in \mathbb{R}^{r\times r} \mid M^T = M, v^TMv > 0\ \forall v \neq 0\}$, the orthonormal basis matrices $\{U \in \mathbb{R}^{n\times s} \mid U^TU = I\}$ or the associated s-dimensional subspaces $\mathcal{U} := \operatorname{ran}(U) \subset \mathbb{R}^n$:

$$\begin{aligned}\mu \mapsto -\mathbb{V}^T(\mu)A(\mu)\mathbb{V}(\mu) &\in \{M \in \mathbb{R}^{r\times r} \mid M^T = M, v^TMv > 0\ \forall v \neq 0\},\\ \mu \mapsto \mathbb{U}(\mu) &\in \{U \in \mathbb{R}^{n\times s} \mid U^TU = I\},\\ \mu \mapsto \mathcal{U}(\mu) = \operatorname{ran}\left(\mathbb{U}(\mu)\right) &\in \{\mathcal{U} \subset \mathbb{R}^n \mid \mathcal{U} \text{ subspace}, \dim(\mathcal{U}) = s\}.\end{aligned}$$

We outline generic methods for constructing such trajectories via interpolation. All the special sets of matrices considered above feature a differentiable structure that

1 Replacing $\mathbb{U}$ with $\mathbb{U}S$, $S \in \mathbb{R}^{s\times s}$ orthogonal, does not affect (7.4) at all. Replacing $\mathbb{V}$ with $\mathbb{V}R$, $R \in \mathbb{R}^{r\times r}$ orthogonal, induces a coordinate change on the reduced state $x_r = R\hat{x}_r$ but preserves the output $\tilde{x}(t) = \mathbb{V}x_r(t) = \mathbb{V}R\hat{x}_r(t)$.

allows one to consider them (directly or indirectly) as submanifolds of some Euclidean matrix space, referred to as matrix manifolds. The above example is not exhaustive. Other matrix manifolds may arise in model reduction applications.

To keep the exposition both general and modular, the interpolation techniques will be formulated for arbitrary submanifolds. For working examples that put these techniques into action, the reader is referred to Chapter 5 of Volume 2 and Chapter 9 of Volume 3 of Model Order Reduction. Model reduction literature on manifold interpolation problems includes [8, 9, 18, 32, 34, 67, 74, 76, 78, 94, 100].

7.1.2 Structure and organization

The text is constructed modular rather than consecutive, so that selected reading is enabled. Yet, this entails that the reader will encounter some repetition.

Section 7.2 covers the essential background from differential geometry. Section 7.3 contains generic methods for interpolation and extrapolation on matrix manifolds. In Section 7.4, the geometric and numerical aspects of the matrix manifolds that arise most frequently in the context of model reduction are discussed.

A practitioner that faces a problem in matrix manifold interpolation may skim through the recap on elementary differential geometry in Section 7.2 and then move on to the subsection of Section 7.4 that corresponds to the matrix manifold in the application. This provides the specific ingredients and formulas for conducting the generic interpolation methods of Section 7.3.

7.1.3 Notation & abbreviations

- w. r. t.: with respect to
- EVD: eigenvalue decomposition
- SVD: singular value decomposition
- POD: proper orthogonal decomposition
- LTI: linear time-invariant (system)
- ODE: ordinary differential equation
- PDE: partial differential equation
- ONB: orthonormal basis
- $\mathbb{R}^{n\times r}$: the set of real n-by-r matrices
- I_n: the n-by-n identity matrix; if dimensions are clear, written as I
- $\operatorname{ran}(A)$: the subspace spanned by the columns of $A \in \mathbb{R}^{n\times r}$
- $GL(n)$: the general linear group of real, invertible n-by-n matrices
- $\operatorname{sym}(n) = \{A \in \mathbb{R}^{n\times n} \mid A^T = A\}$: the set of real, symmetric n-by-n matrices
- $\operatorname{skew}(n) = \{A \in \mathbb{R}^{n\times n} \mid A^T = -A\}$: the set of real, skew-symmetric n-by-n matrices

- $\mathrm{SPD}(n) = \{A \in \mathrm{sym}(n) \mid x^T Ax > 0 \forall x \in \mathbb{R}^n \setminus \{0\}\}$: the set of real, symmetric positive definite n-by-n matrices
- $O(n) = \{Q \in \mathbb{R}^{n\times n} \mid Q^T Q = I_n = QQ^T\}$: the orthogonal group
- $SO(n) = \{Q \in O(n) \mid \det(Q) = 1\}$: the special orthogonal group
- $St(n,r) = \{U \in \mathbb{R}^{n\times r} \mid U^T U = I_r\}$: the (compact) Stiefel manifold, $r \leq n$
- $Gr(n,r)$: the Grassmann manifold of r-dimensional subspaces of $\mathbb{R}^n$, $r \leq n$
- $\mathcal{M}$: a differentiable manifold
- $\mathcal{D}_p \subset \mathcal{M}$: an open domain around the point p on a manifold $\mathcal{M}$
- $D_x \subset \mathbb{R}^n$: an open domain in the Euclidean space around a point $x \in \mathbb{R}^n$
- $T_p\mathcal{M}$: the tangent space of $\mathcal{M}$ at a location $p \in \mathcal{M}$
- $\langle A, B\rangle_0 = \mathrm{trace}(A^T B)$: the standard (Frobenius) inner product on $\mathbb{R}^{n\times r}$
- $\langle v, w\rangle_p^{\mathcal{M}}$: the Riemannian metric on $T_p\mathcal{M}$ (the superscript is often omitted)
- $\exp_m$: standard matrix exponential
- $\log_m$: standard (principal) matrix logarithm
- $\mathrm{Exp}_p^{\mathcal{M}}$: the Riemannian exponential of a manifold $\mathcal{M}$ at base point $p \in \mathcal{M}$
- $\mathrm{Log}_p^{\mathcal{M}}$: the Riemannian logarithm of a manifold $\mathcal{M}$ at base point $p \in \mathcal{M}$.

7.2 Basic concepts of differential geometry

This section provides the essentials on elementary differential geometry. Established textbook references on differential geometry include [35, 60, 61, 63, 65]; condensed introductions can be found in [49, Appendices C.3, C.4, C.5] and [39]. An account of differential geometry that is tailor-made to matrix manifold applications is given in [3].

The fundamental objects of study in differential geometry are *differentiable manifolds*. Differentiable manifolds are generalizations of curves (one-dimensional) and surfaces (two-dimensional) to arbitrary dimensions. Loosely speaking, an n-dimensional differentiable manifold $\mathcal{M}$ is a topological space that 'locally looks like $\mathbb{R}^n$' with certain smoothness properties. This concept is rendered precise by postulating 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 important 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 [39, Figure 18.2, p. 496] or [49, Figure 3.1, p. 342]. Note that the coordinate change $x \circ \tilde{x}^{-1}$ maps from an open domain of $\mathbb{R}^n$ to an open domain of $\mathbb{R}^n$, so that the standard concepts of multivariate calculus apply. For details, see [3, §3.1.1] or [39, §18.8].

Depending on the context, we will write $x(p)$ for the value of a coordinate chart at p and also $x \in \mathbb{R}^n$ for a point in $\mathbb{R}^n$.

Of special importance to numerical applications are *embedded submanifolds in the Euclidean space*.

Definition 7.1 (Submanifolds of $\mathbb{R}^{n+d}$). A *parameterization* is a bijective differentiable function $f : \mathbb{R}^n \supset D \to f(D) \subset \mathbb{R}^{n+d}$ with continuous inverse such that its Jacobi matrix $Df_x \in \mathbb{R}^{(n+d)\times n}$ has full rank n at every point $x \in D$.

A subset $\mathcal{M} \subset \mathbb{R}^{n+d}$ is called an *n-dimensional embedded submanifold of* $\mathbb{R}^{n+d}$, if for every $p \in \mathcal{M}$, there exists an open neighborhood $\Omega \subset \mathbb{R}^{n+d}$ such that $\mathcal{D}_p := \mathcal{M} \cap \Omega$ is the image of a parameterization

$$f : \mathbb{R}^n \supset D_x \to f(D_x) = \mathcal{D}_p = \mathcal{M} \cap \Omega \subset \mathbb{R}^{n+d}.$$

One can show that, if $f : D \to \mathcal{M} \cap \Omega$ and $\tilde{f} : \tilde{D} \to \mathcal{M} \cap \tilde{\Omega}$ are two parameterizations, say with $f(x_0) = \tilde{f}(\tilde{x}_0) = p \in \mathcal{M} \cap \Omega \cap \tilde{\Omega}$, then

$$(f^{-1} \circ \tilde{f}) : \tilde{f}^{-1}(\Omega \cap \tilde{\Omega}) \to f^{-1}(\Omega \cap \tilde{\Omega})$$

is a diffeomorphism (between open sets in $\mathbb{R}^n$). In this sense, parameterizations f are the inverses of coordinate charts x. In addition to coordinate charts and parameterizations, submanifolds can be characterized via equality constraints. This fact is due to the inverse function theorem of classical multivariate calculus [64, §I.5]. For details, see [39, Thm. 18.7, p. 497].

Theorem 7.1 ([39, 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 $\mathbb{R}^{n+d}$.

An obvious application of Theorem 7.1 to the function $h : \mathbb{R}^3 \to \mathbb{R}, (x_1, x_2, x_3) \mapsto x_1^2 + x_2^2 + x_3^2 - 1$ establishes the unit sphere $S^2 = h^{-1}(0)$ as a 2-dimensional submanifold of $\mathbb{R}^{2+1}$. As a more sophisticated example, we recognize the orthogonal group as a differentiable (sub)-manifold.

Example 7.1. Consider the orthogonal group $O(n) \subset \mathbb{R}^{n\times n} \simeq \mathbb{R}^{n^2}$ and the set of symmetric matrices $\mathrm{sym}(n) \simeq \mathbb{R}^{n(n+1)/2}$. Define $h : \mathbb{R}^{n\times n} \to \mathrm{sym}(n), A \mapsto A^TA - I$. Then $Dh_A(B) = A^TB + B^TA$. For $Q \in O(n)$, the differential is indeed surjective: For any $M \in \mathrm{sym}(n)$, we have $Dh_Q(\frac{1}{2}QM) = \frac{1}{2}Q^TQM + \frac{1}{2}M^TQ^TQ = M$. As a consequence, the orthogonal group $O(n)$ is a submanifold of dimension $n^2 - \frac{1}{2}(n(n+1)) = \frac{1}{2}(n(n-1))$ of the Euclidean matrix space $\mathbb{R}^{n\times n}$.

7.2.1 Intrinsic and extrinsic coordinates

As a rule, numerical data processing on manifolds requires calculations in explicit coordinates. For differentiable submanifolds, we distinguish between two types: *extrinsic* and *intrinsic coordinates*. Extrinsic coordinates address points on a submanifold $\mathcal{M} \subseteq \mathbb{R}^n$ with respect to their coordinates in the ambient space $\mathbb{R}^n$, while intrinsic coordinates are with respect to the local parameterizations. Hence, extrinsic coordinates are what an outside observer would see, while intrinsic coordinates correspond to the perspective of an observer that resides on the manifold. Let us exemplify these concepts on the two-dimensional unit sphere S^2, embedded in $\mathbb{R}^3$. As a point set, the sphere is defined by the equation

$$S^2 = \{(x_1, x_2, x_3)^T \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 1\}.$$

Any three-vector $(x_1, x_2, x_3)^T \in S^2$ specifies a point on the sphere in *extrinsic* coordinates. However, it is intuitively clear that S^2 is intrinsically a two-dimensional object. Indeed, S^2 can be parameterized via

$$f : \mathbb{R}^2 \supset [0, 2\pi)^2 \to S^2 \subset \mathbb{R}^3, \quad (\alpha, \beta) \mapsto \begin{pmatrix} \sin(\alpha)\cos(\beta) \\ \sin(\alpha)\sin(\beta) \\ \cos(\alpha) \end{pmatrix}.$$

The parameter vector $(\alpha, \beta) \in \mathbb{R}^2$ specifies a point on S^2 in *intrinsic* coordinates. Even though intrinsic coordinates directly reflect the dimension of the manifold at hand, they often cannot be calculated explicitly and extrinsic coordinates are the preferred choice in numerical applications [36, §2, p. 305]. Turning back to Example 7.1, we recall that the intrinsic dimension of the orthogonal group is $\frac{1}{2}n(n-1)$. Yet, in practice, one uses the extrinsic representation with $(n \times n)$-matrices Q, keeping the defining equation $Q^T Q = I$ in mind.

7.2.2 Tangent spaces

We need a few more fundamental concepts.

Definition 7.2 (Tangent space of a differentiable submanifold). Let $\mathcal{M} \subset \mathbb{R}^{n+d}$ be an n-dimensional submanifold of $\mathbb{R}^{n+d}$. The *tangent space* of $\mathcal{M}$ 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.,

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

Here, $J \subseteq \mathbb{R}$ is an arbitrarily small open interval with $t_0 \in J$.

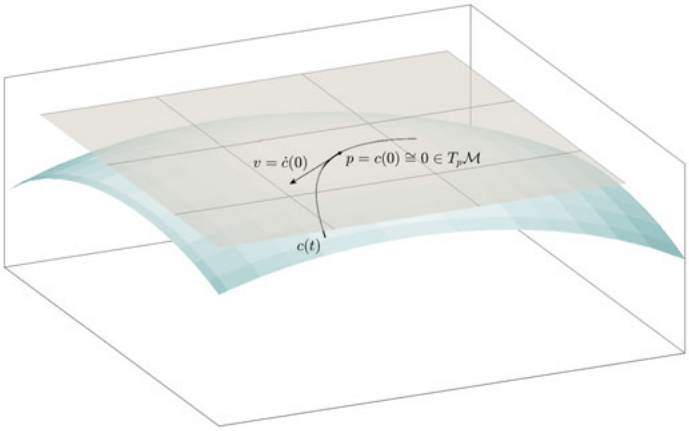


Figure 7.1: Visualization of a manifold (curved surface) with the tangent space $T_p\mathcal{M}$ attached. The tangent vector $v = \dot{c}(0) \in T_p\mathcal{M}$ is the velocity vector of a curve $c : t \mapsto c(t) \in \mathcal{M}$.

The concept is illustrated in Figure 7.1. It is straightforward to show that the tangent space is actually a vector space. Moreover, the tangent space can be characterized both with respect to intrinsic and extrinsic coordinates.

Theorem 7.2 (Tangent space, intrinsic characterization). *Let $\mathcal{M} \subset \mathbb{R}^{n+d}$ be an n-dimensional submanifold of $\mathbb{R}^{n+d}$ and let $f : \mathbb{R}^n \supseteq D \to f(D) \subseteq \mathcal{M}$ be a parameterization. Then, for $x \in D$ with $p = f(x) \in \mathcal{M}$, we have*

$$T_p\mathcal{M} = \operatorname{ran}(Df_x).$$

Theorem 7.3 (Tangent space, extrinsic characterization). *Let $h : \mathbb{R}^{n+d} \supset \Omega \to \mathbb{R}^d$ and $c_0 \in \mathbb{R}^d$ be as in Theorem 7.1 and let $\mathcal{M} := h^{-1}(c_0) \subset \mathbb{R}^{n+d}$. Then, for $p \in \mathcal{M}$, we have*

$$T_p\mathcal{M} = \ker(Dh_p).$$

Note that both Theorem 7.2 and Theorem 7.3 immediately show that the tangent space $T_p\mathcal{M}$ is a vector space of the same dimension n as the manifold $\mathcal{M}$.

Example 7.2. The tangent space of the orthogonal group $O(n)$ at a point Q_0 is

$$T_{Q_0}O(n) = \{\Delta \in \mathbb{R}^{n\times n} \mid \Delta^T Q_0 = -Q_0^T\Delta\}.$$

This fact can be established via considering a matrix curve $Q : t \mapsto Q(t)$ with $Q(0) = Q_0$ and velocity vector $\Delta = \dot{Q}(0) \in T_{Q_0}O(n)$. Then

$$0 = \frac{d}{dt}|_{t=0} I = \frac{d}{dt}|_{t=0} Q^T(t)Q(t) = \Delta^T Q_0 + Q_0^T\Delta.$$

(The claim follows by counting the dimension of the subspace $\{\Delta^T Q_0 = -Q_0^T\Delta\}$.) As an alternative, we can consider $h : \mathbb{R}^{n\times n} \to \operatorname{sym}(n), A \mapsto A^T A - I$ as in Example 7.1. Then $Dh_{Q_0}(\Delta) = Q_0^T\Delta + \Delta^T Q_0$ and $T_{Q_0}O(n) = \ker(Dh_{Q_0})$.

7.2.3 Geodesics and the Riemannian distance function

One of the most important problems in both general differential geometry and data processing on manifolds is to determine the shortest connection between two points on a given manifold. This requires one to measure the lengths of curves. Recall that the length of a curve $c : [a,b] \to \mathbb{R}^n$ in the Euclidean space is $L(c) = \int_a^b \|\dot{c}(t)\|\, dt$. In order to transfer this to the manifold setting, an inner product for tangent vectors is needed that is consistent with the manifold structure.

Definition 7.3 (Riemannian metrics). Let $\mathcal{M}$ be a differentiable submanifold of $\mathbb{R}^{n+d}$. 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 $\|v\|_p := \sqrt{\langle v, v\rangle_p}$.[2] 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 $\|\dot{c}(t)\|_{c(t)} \equiv v_0$ 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

$$\operatorname{dist}_{\mathcal{M}}(p,q) = \inf\{L(c) \mid c : [a,b] \to \mathcal{M} \text{ piecewise smooth}, c(a) = p, c(b) = q\}, \quad (7.5)$$

where, by convention, $\inf\{\emptyset\} = \infty$.

Hence, a shortest path between $p, q \in \mathcal{M}$ is a curve c that connects p and q such that $L(c) = \operatorname{dist}_{\mathcal{M}}(p,q)$. In general, shortest paths on $\mathcal{M}$ do not exist.[3] Yet, candidates for shortest curves between points that are sufficiently close to each other can be obtained via a variational principle: Given a parametric family of suitably regular curves $c_s : t \mapsto c_s(t) \in \mathcal{M}$, $s \in (-\varepsilon, \varepsilon)$ that connect the same fixed endpoints $c_s(a) = p$ and $c_s(b) = q$ for all s, one can consider the length functional $s \mapsto L(c_s)$. A curve $c = c_0$ is a first-order candidate for a shortest path between p and q, if it is a critical point of

2 This notation should not be confused with the classical p-norm $\sqrt[p]{\sum_i |v_i|^p}$.

3 Consider $\mathbb{R}^{2,*} = \mathbb{R}^2 \setminus \{(0,0)\}$ with the Euclidean inner product. There is no shortest connection from $(-1,0)$ to $(1,0)$ on $\mathbb{R}^{2,*}$. A sequence of curves that is in $\mathbb{R}^{2,*}$ and converges to the curve $c : [-1,1] \to \mathbb{R}^2, t \mapsto (t,0)$ is readily constructed. Hence, the Riemannian distance between $(-1,0)$ and $(1,0)$ is 2. Yet, every curve connecting these points must go around the origin. The length-minimizing curve of length 2 crosses the origin and is thus not an admissible curve on $\mathbb{R}^{2,*}$.

the length functional, i. e., if $\frac{d}{ds}|_{s=0} L(c_s) = 0$. Such curves are called *geodesics*. Differentiating the length functional leads to the so-called first variation formula [65, §6], which, in turn, leads to the characterizing equation for geodesics:

Definition 7.4 (Geodesics). A differentiable curve $c : [a, b] \to \mathcal{M}$ is called a *geodesic* (w. r. t. to a given Riemannian metric), if the *covariant derivative* of its velocity vector field vanishes, i. e.,

$$\frac{D\dot{c}}{dt}(t) = 0 \quad \forall t \in [a, b]. \tag{7.6}$$

Remark 7.1. If a starting point $c(0) = p \in \mathcal{M}$ and a starting velocity $\dot{c}(0) = v \in T_p\mathcal{M}$ are specified, then the geodesic equation (7.6) translates to an initial value problem of second order with guaranteed existence and uniqueness of local solutions, [3, p. 102].

An immediate consequence of (7.6) is that geodesics are constant-speed curves. A formal introduction of the covariant derivative $\frac{D}{dt}$ along a curve is beyond the scope of this contribution, and the interested reader is referred to, e. g., [65, §4, §5]. To get some intuition, we introduce this concept for embedded Riemannian submanifolds $\mathcal{M} \subset \mathbb{R}^{n+d}$, where the metric is the Euclidean metric of $\mathbb{R}^{n+d}$ restricted to the tangent bundle; see also [39, §20.12]:

A *vector field along a curve* $c : [a, b] \to \mathcal{M}$ is a differentiable map $v : [a, b] \to \mathbb{R}^{n+d}$ such that $v(t) \in T_{c(t)}\mathcal{M}$.[4] For every $p \in \mathcal{M}$, the ambient $\mathbb{R}^{n+d}$ decomposes into an orthogonal direct sum

$$\mathbb{R}^{n+d} = T_p\mathcal{M} \oplus T_p\mathcal{M}^\perp,$$

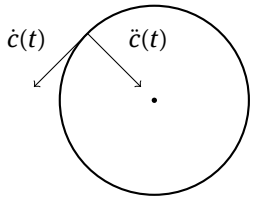
where $T_p\mathcal{M}^\perp$ is the orthogonal complement of $T_p\mathcal{M}$ and orthogonality is w. r. t. the standard Euclidean inner product on $\mathbb{R}^{n+d}$. Let $\Pi_p : \mathbb{R}^{n+d} \to T_p\mathcal{M}$ be the (base point-dependent) orthogonal projection onto the tangent space at p. In this setting (and only in this), the covariant derivative of a vector field $v(t)$ along a curve $c(t)$ is the tangent component of $\dot{v}(t)$, i. e., $\frac{Dv}{dt}(t) = \Pi_{c(t)}(\dot{v}(t))$. As a consequence,

$$\frac{D\dot{c}}{dt}(t) = \Pi_{c(t)}(\ddot{c}(t)) \tag{7.7}$$

and the geodesics on Riemannian submanifolds with the metric induced by the ambient Euclidean inner product are precisely the constant-speed curves with acceleration vectors orthogonal to the corresponding tangent spaces, i. e., $\ddot{c}(t) \in T_{c(t)}\mathcal{M}^\perp$.

Example 7.3. On the unit sphere $S^2 \subset \mathbb{R}^3$, the geodesics are great circles. When considered as curves in the ambient $\mathbb{R}^3$, their acceleration vector points directly to the origin and is thus orthogonal to the corresponding tangent space, see the cartoon below. When viewed as entities of S^2, these curves do not experience any acceleration at all.

4 The prime example for such a vector field is the curve's own velocity field $v(t) = \dot{c}(t)$.



Mind that a constant-speed curve in $\mathbb{R}^n$ changes its direction only, when it experiences a non-zero acceleration. In this sense, geodesics on manifolds are the counterparts to straight lines in the Euclidean 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. Of importance is the *Riemannian connection* or *Levi-Civita connection* that is compatible with a Riemannian metric [3, Thm 5.3.1], [65, Thm 5.4]. It is determined uniquely by the Koszul formula,

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

and is used to define the *Riemannian curvature tensor*

$$(X, Y, Z) \mapsto R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.^5$$

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 [65, Thm. 7.3]. Hence, 'flatness' depends on the Riemannian metric.

7.2.4 Normal coordinates

The local uniqueness and existence of geodesics allows us to map a tangent vector $v \in T_p\mathcal{M}$ to the endpoint of a geodesic that starts from $p \in \mathcal{M}$ with velocity v. Formalizing this principle gives rise to the *Riemannian exponential*,

$$\operatorname{Exp}_p^{\mathcal{M}} : T_p\mathcal{M} \supset B_\varepsilon(0) \to \mathcal{M}, \quad v \mapsto q := \operatorname{Exp}_p^{\mathcal{M}}(v) := c_{p,v}(1). \tag{7.8}$$

Here, $t \mapsto c_{p,v}(t)$ is the geodesic that starts from p with velocity v and $B_\varepsilon(0) \subset T_p\mathcal{M}$ is the open ball with radius ε and center 0 in the tangent space;[6] see Figure 7.2. Note

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

6 For technical reasons, $\varepsilon > 0$ must be chosen small enough such that $c_{p,v}(t)$ is defined on the unit interval $[0, 1]$.

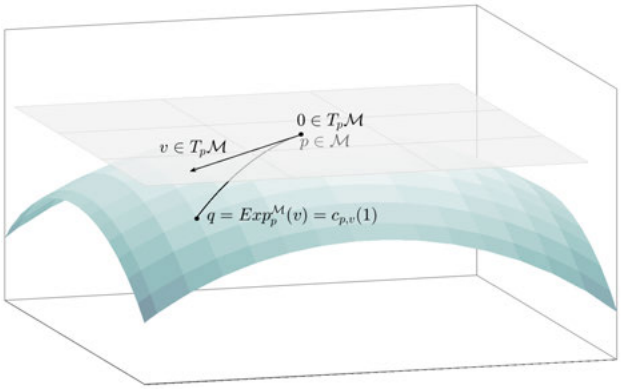


Figure 7.2: The Riemannian exponential sends tangent vectors to end point of geodesic curves.

that we can restrict the considerations to unit-speed geodesics via

$$\mathrm{Exp}_p^{\mathcal{M}}(v) := c_{p,v}(1) = c_{p,v/\|v\|}(t_v) = \mathrm{Exp}_p^{\mathcal{M}}\left(t_v \frac{v}{\|v\|}\right),$$

where $t_v = \|v\|$; see [65, §5., p. 72 ff.] for the details.

For $\varepsilon > 0$ small enough, the Riemannian exponential is a smooth diffeomorphism between $B_\varepsilon(0)$ and an open domain on $\mathcal{D}_p \subset \mathcal{M}$ around the point p. Hence, it is invertible. The smooth inverse map is called the *Riemannian logarithm* and is denoted by

$$\mathrm{Log}_p^{\mathcal{M}} : \mathcal{M} \supset \mathcal{D}_p \to B_\varepsilon(0) \subset T_p\mathcal{M}, \quad q \mapsto v := \left(\mathrm{Exp}_p^{\mathcal{M}}\right)^{-1}(q), \tag{7.9}$$

where v satisfies $c_{p,v}(1) = q$.

Thus, the Riemannian logarithm is associated with the *geodesic endpoint problem*: Given $p, q \in \mathcal{M}$, find a geodesic that connects p and q. The Riemannian exponential map establishes a local parametrization of a small region around a location $p \in \mathcal{M}$ in terms of coordinates of the flat vector space $T_p\mathcal{M}$. This is referred to as representing the manifold in *normal coordinates* [60, §III.8], [65, Lem. 5.10]. Normal coordinates are radially isometric in the sense that the Riemannian distance between p and $q = \mathrm{Exp}_p^{\mathcal{M}}(v)$ is exactly the same as the length of the tangent vector $\|v\|_p$ as measured in the metric on $T_p\mathcal{M}$, provided that v is contained in a neighborhood of $0 \in T_p\mathcal{M}$, where the exponential is invertible, [65, Lem. 5.10 & Cor. 6.11].

Mind that the definition of the Riemannian exponential depends on the geodesics, which, in turn, depend on the chosen Riemannian metric—via Definition 7.3. Different metrics lead to different geodesics and thus to different exponential and logarithm maps.

7.2.5 Matrix Lie groups and quotients by group actions

In general, a *Lie group* is a differentiable manifold $\mathcal{G}$ which also has a group structure, such that the group operations 'multiplication' and 'inversion',

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

are both smooth [39, 46, 41]. A *matrix Lie group* $\mathcal{G}$ is a subgroup of $GL(n, \mathbb{C})$ that is closed in $GL(n, \mathbb{C})$.[7] This definition already implies that $\mathcal{G}$ is an embedded submanifold of $\mathbb{C}^{n\times n}$ [46, Corollary 3.45]. Not all matrix groups are Lie groups and not all Lie groups are matrix Lie groups; see [46, §1.1 and §4.8]. However, matrix Lie groups are arguably the most important class of Lie groups when it comes to practical applications and this exposition is restricted to this subclass.

Let $\mathcal{G}$ be an arbitrary matrix Lie group. When endowed with the bracket operator or *matrix commutator* $[V, W] = VW - WV$, the tangent space $T_I\mathcal{G}$ at the identity is called the *Lie algebra* associated with the Lie group $\mathcal{G}$; see [46, §3]. As such, it is denoted by $\mathfrak{g} = T_I\mathcal{G}$. For any $A \in \mathcal{G}$, the function "left-multiplication with A" is a diffeomorphism $L_A : \mathcal{G} \to \mathcal{G}, L_A(B) = AB$; its differential at a point $M \in \mathcal{G}$ is the isomorphism $d(L_A)_M : T_M\mathcal{G} \to T_{L_A(M)}\mathcal{G}, d(L_A)_M(V) = AV$. Using this observation at $M = I$ shows that the tangent space at an arbitrary location $A \in \mathcal{G}$ is given by the translates (by left-multiplication) of the tangent space at the identity:

$$T_A\mathcal{G} = T_{L_A(I)}\mathcal{G} = A\mathfrak{g} = \{\Delta = AV \in \mathbb{R}^{n\times n} \mid V \in \mathfrak{g}\}, \tag{7.10}$$

[41, §5.6, p. 160]. The Lie algebra $\mathfrak{g} = T_I\mathcal{G}$ of $\mathcal{G}$ can equivalently be characterized as the set of all matrices Δ such that $\exp_m(t\Delta) \in \mathcal{G}$ for all $t \in \mathbb{R}$. The intuition behind this fact is that all tangent vectors are velocity vectors of smooth curves running on $\mathcal{G}$ (Definition 7.2) and that $c(t) = \exp_m(t\Delta)$ is a smooth curve starting from $c(0) = I$ with velocity $\dot{c}(0) = \Delta$; see [46, Def. 3.18 & Cor. 3.46] for the details. By definition, the exponential map[8] for a matrix Lie group is the matrix exponential restricted to the corresponding Lie algebra, i. e. the tangent space at the identity $\mathfrak{g} = T_I\mathcal{G}$, [46, §3.7],

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

In general, a Lie algebra is a vector space with a linear, skew-symmetric bracket operation, called *Lie bracket* $[\cdot, \cdot]$, that satisfies the *Jacobi identity*.

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

7 But not necessarily in $\mathbb{C}^{n\times n}$.

8 The exponential map of a Lie group must not be confused with the Riemannian exponential.

Quotients of Lie groups by closed subgroups

In many settings, it is important or sometimes even necessary to consider certain points p, q on a given differentiable manifold $\mathcal{M}$ as equivalent. Consider the following example.

Example 7.4. Let $U \in \mathbb{R}^{n\times r}$ feature orthonormal columns so that $U^T U = I_r$. We may extend the columns of $U = (u^1, \dots, u^r)$ to an orthogonal matrix $Q = (u^1, \dots, u^r, u^{r+1}, \dots, u^n) \in O(n)$. Define $I_r \times O(n-r) := \{\left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) \mid R \in O(n-r)\}$. This is actually a closed subgroup of $O(n)$, in symbols $(I_r \times O(n-r)) \leq O(n)$. The action $\tilde{Q} = Q\Phi$ with any orthogonal matrix $\Phi \in I_r \times O(n-r)$ preserves the first r columns of Q. Hence, we may identify U with the equivalence class $[Q] = \{Q\Phi \mid \Phi \in I_r \times O(n-r)\} \subset O(n)$. In Sections 7.4.4 and 7.4.5, we will see that this example establishes the Stiefel manifold of ONBs and eventually also the Grassmann manifold of subspaces as quotients of the orthogonal group $O(n)$.

Note that in the example, the equivalence relation is induced by actions of the Lie group $I_r \times O(n-r)$. Quotients that arise from such group actions are important examples of *quotient manifolds*. Theorems 7.4 and 7.5 cover this example as well as all other cases of quotient manifolds that are featured in this chapter. First, group actions need to be formalized.

Definition 7.5 (Cf. [66, p. 162,163]). Let $\mathcal{G}$ be a Lie group, $\mathcal{M}$ be a smooth manifold, and let $\mathcal{G} \times \mathcal{M} \to \mathcal{M}, (g, p) \mapsto g \cdot p$ be a *left action of $\mathcal{G}$ on $\mathcal{M}$*.[9] The *orbit relation* on $\mathcal{M}$ induced by $\mathcal{G}$ is defined by

$$p \simeq q :\Leftrightarrow \exists g \in \mathcal{G} : \quad g \cdot p = q.$$

The equivalence classes are the *$\mathcal{G}$-orbits* $[p] := \mathcal{G}p := \{g \cdot p \mid g \in \mathcal{G}\}$. The *orbit space* is denoted by $\mathcal{M}/\mathcal{G} := \{[p] \mid p \in \mathcal{M}\}$. The *quotient map* sends a point to its $\mathcal{G}$-orbit via $\Pi : \mathcal{M} \to \mathcal{M}/\mathcal{G}, p \mapsto [p]$. The action is *free*, if every isotropy group $\mathcal{G}_p := \{g \in \mathcal{G} \mid g \cdot p = p\}$ is trivial, $\mathcal{G}_p = \{e\}$.

Theorem 7.4 (Quotient manifold theorem, cf. [66, Thm. 21.10]). *Suppose $\mathcal{G}$ is a Lie group acting* smoothly, freely, and properly *on a smooth manifold $\mathcal{M}$. Then the orbit space $\mathcal{M}/\mathcal{G}$ is a manifold of dimension* $\dim \mathcal{M} - \dim \mathcal{G}$*, and has a unique smooth structure such that the quotient map $\Pi : \mathcal{M} \to \mathcal{M}/\mathcal{G}, p \mapsto [p]$ is a smooth submersion.*[10] *In this context, $\mathcal{M}$ is called the total space and $\mathcal{M}/\mathcal{G}$ is the quotient (space).*

A special case is Lie groups under actions of Lie subgroups.

---

9 The theory for right actions is analogous. In all cases considered in this chapter, $\mathcal{M}$ is a matrix manifold so that "$\cdot$" is the usual matrix product.

10 I. e. a smooth surjective mapping such that the differential is surjective at every point.

Definition 7.6. [66, §21, p. 551] Let $\mathcal{G}$ be a Lie group and $\mathcal{H} \leq \mathcal{G}$ be a Lie subgroup. For $g \in \mathcal{G}$, a subset of $\mathcal{G}$ of the form $[g] := g\mathcal{H} = \{g \cdot h \mid h \in \mathcal{H}\}$ is called a *left coset of* $\mathcal{H}$. The left cosets form a partition of $\mathcal{G}$, and the quotient space determined by this partition is called the *left coset space of* $\mathcal{G}$ *modulo* $\mathcal{H}$, and is denoted by $\mathcal{G}/\mathcal{H}$.

Coset spaces of Lie groups are again smooth manifolds.

Theorem 7.5 (Cf. [66, Thm 21.17, p. 551]). *Let $\mathcal{G}$ be a Lie group and let $\mathcal{H}$ be a closed subgroup of $\mathcal{G}$. The left coset space $\mathcal{G}/\mathcal{H}$ is a manifold of dimension* $\dim \mathcal{G} - \dim \mathcal{H}$ *with a unique differentiable structure such that the quotient map $\Pi : \mathcal{G} \to \mathcal{G}/\mathcal{H}, g \mapsto [g]$ is a smooth submersion.*

In general, if $\pi : \mathcal{M} \to \mathcal{N}$ is a surjective submersion between two manifolds $\mathcal{M}$ and $\mathcal{N}$, then for any $q \in \mathcal{N}$, the preimage $\pi^{-1}(q) \subset \mathcal{M}$ is called the *fiber over* q, and is denoted by $\mathcal{M}_q$. Each fiber $\mathcal{M}_q$ is itself a closed, embedded submanifold by the implicit function theorem. If $\mathcal{M}$ has a Riemannian metric $\langle \cdot, \cdot \rangle_p^{\mathcal{M}}$, then at each point $p \in \mathcal{M}$, the tangent space $T_p\mathcal{M}$ decomposes into an *orthogonal direct sum* $T_p\mathcal{M} = T_p\mathcal{M}_{\pi(p)} \oplus (T_p\mathcal{M}_{\pi(p)})^{\perp}$. The tangent space of the fiber $T_p\mathcal{M}_{\pi(p)} =: V_p$ is the called the *vertical space*, its orthogonal complement $H_p := V_p^{\perp}$ is the *horizontal space*. The vertical space is the kernel $V_p = \ker(d\pi_p)$ of the differential $d\pi_p : T_p\mathcal{M} \to T_{\pi(p)}\mathcal{N}$; the horizontal space is isomorphic to $T_{\pi(p)}\mathcal{N}$. This allows one to identify $H_p \cong T_{\pi(p)}\mathcal{N}$; see [3, Figure 3.8., p. 44] for an illustration. This construction helps to compute tangent spaces of quotients, if the tangent space of the total space is known.

If $\mathcal{G}/\mathcal{H}$ is a quotient as in Theorem 7.4 or 7.5 and if $\Pi : \mathcal{G} \to \mathcal{G}/\mathcal{H}$ is the corresponding quotient map, then Π is a local diffeomorphism. A Riemannian metric on the quotient can be defined by

$$\langle v, w \rangle_{[g]}^{\mathcal{G}/\mathcal{H}} := \left\langle (d\Pi_g)^{-1}(v), (d\Pi_g)^{-1}(w) \right\rangle_g^{\mathcal{G}}, \quad v, w \in T_{[g]}(\mathcal{G}/\mathcal{H}). \tag{7.11}$$

For this (and only this) metric, the quotient map is a *local isometry*.

In fact, Theorem 7.5 additionally establishes $\mathcal{G}/\mathcal{H}$ as a *homogeneous space*, i. e. a smooth manifold $\mathcal{M}$ endowed with a *transitive smooth action* by a Lie group (cf. [66, §21, p. 550]). In the setting of the theorem, the group action is given by the left action of $\mathcal{G}$ on $\mathcal{G}/\mathcal{H}$ given by $g_1 \cdot [g_2] := [g_1 \cdot g_2]$. A transitive action allows us to transport a location $p \in \mathcal{M}$ to any other location $q \in \mathcal{M}$.

7.3 Interpolation on non-flat manifolds

When working with matrix manifolds, the data is usually given in extrinsic coordinates; see Section 7.2. For example, data on the compact Stiefel manifold $St(n, r) = \{U \in \mathbb{R}^{n \times r} \mid U^T U = I_r\}$, $r \leq n$, is given in form of n-by-r matrices. These matrices

feature nr entries while the intrinsic number of degrees of freedom, i. e., the intrinsic dimension is $nr - \frac{1}{2}r(r+1)$; see Section 7.4.4. Essentially, the practical obstacle associated with data interpolation on matrix manifolds arises from this fact. Given, say, k matrices on $St(n,r)$ in extrinsic coordinates, interpolating entry-by-entry will most certainly lead to interpolants that do not feature orthogonal columns and thus are *not* points on the Stiefel manifold. Likewise, entry-by-entry interpolation of positive definite matrices is not guaranteed to produce another positive definite matrix.

There are essentially two different approaches to address this issue: Performing the interpolation on the tangent space of the manifold and using the Riemannian barycenter or Riemannian center of mass as an interpolant. Both will be explained in more detail in the next two subsections.[11]

7.3.1 Interpolation in normal coordinates

As outlined in Section 7.2, every location $p \in \mathcal{M}$ on an n-dimensional differentiable manifold features a small neighborhood $\mathcal{D}_p$ that is the domain of a coordinate chart $x : \mathcal{M} \supset \mathcal{D}_p \to D_{x(p)} \subset \mathbb{R}^n$ that maps bijectively onto an open set $D_{x(p)} \subset \mathbb{R}^n$. Therefore, for a sample data set $\{p_1, \dots, p_k\} \subset \mathcal{D}_p$ that is completely contained in the domain of a single coordinate chart x, interpolation can be performed as follows:

1. Map the data set to $D_{x(p)}$: Calculate $v_1 = x(p_1), \dots, v_k = x(p_k) \in D_{x(p)}$.
2. Interpolate in $D_{x(p)}$ to produce the interpolant $v^* \in D_{x(p)}$.
3. Map back to manifold: compute $p^* = x^{-1}(v^*) \in \mathcal{D}_p$.

In principle, any coordinate chart may be applied. In practice, the challenge is to find a suitable coordinate chart that can be evaluated efficiently. Moreover, it is desirable that the chosen chart preserves the geometry of the original data set as well as possible.[12] The standard choice is to use *normal coordinates* as introduced in Section 7.2.4. This means that the Riemannian logarithm is used as the coordinate chart

$$\mathrm{Log}_p^{\mathcal{M}} : \mathcal{M} \supset \mathcal{D}_p \to B_\varepsilon(0) \subset T_p\mathcal{M}$$

with the Riemannian exponential

$$\mathrm{Exp}_p^{\mathcal{M}} : T_p\mathcal{M} \supset B_\varepsilon(0) \to \mathcal{D}_p \subset \mathcal{M}$$

as the corresponding parameterization. The general procedure of data interpolation via the tangent space is formulated as Algorithm 7.1.

11 The German speaking reader may find an introduction that addresses a general scientific audience in [90].

12 There are no isometric coordinate charts on a non-flat manifold; see [65, Thm 7.3].

Algorithm 7.1: Interpolation in normal coordinates.

Input: Data set $\{p_1, \dots, p_k\} \subset \mathcal{M}$.

1: Choose $p_i \in \{p_1, \dots, p_k\}$ as a base point.
2: Check that $\operatorname{Log}_{p_i}^{\mathcal{M}}(p_j)$ is well-defined for all $j = 1, \dots, k$.
3: **for** $j = 1, \dots, k$ **do**
4: Compute $v_j := \operatorname{Log}_{p_i}^{\mathcal{M}}(p_j) \in T_p\mathcal{M}$.
5: **end for**
6: Compute v^* via Euclidean interpolation of $\{v_1, \dots, v_k\}$.
7: Compute $p^* := \operatorname{Exp}_{p_i}^{\mathcal{M}}(v^*)$

Output: $p^* \in \mathcal{M}$.

Remark 7.2. There are a few facts that the practitioner needs to be aware of:

1. The interpolation procedure of Algorithm 7.1 depends on which sample point is selected to act as the base point. Different choices may lead to different interpolants.[13]
2. For *matrix manifolds*, the tangent space is often also given in *extrinsic coordinates*. This means that an entry-by-entry interpolation of the matrices that represent the tangent vectors may lead to an interpolant that is not in the tangent space. As an illustrative example, consider the Grassmannian $Gr(n, r)$. Matrices $\Delta_1, \dots, \Delta_k \in T_{[U]}Gr(n, r)$ are characterized by $U^T\Delta_j = 0$. Entry-by-entry interpolation in the tangent space may potentially result in a matrix Δ^* that is not orthogonal to the base point U, i. e. $U^T\Delta^* \neq 0$; see [100, §2.4].
   In general, because of the vector space structure of the tangent space of any manifold $\mathcal{M}$, it is sufficient to use an interpolation method that expresses the interpolant in $T_p\mathcal{M}$ as a weighted linear combination of the sampled tangent vectors $v_1, \dots, v_k \in T_p\mathcal{M}$,

$$v^* = \sum_{j=1}^{k} \omega_j v_j.$$

   Amongst others, *linear interpolation*, *Lagrange and Hermite interpolation*, *spline interpolation* and interpolation via *radial basis functions* fulfill this requirement. As an aside, the interpolation procedure is computationally less expensive, since it works on the weight coefficients ω_j rather than on every single entry.

Quasi-linear interpolation of trajectories via geodesics

In this paragraph, we address applications, where the sampled manifold data features a univariate parametric dependency. The setting is as follows. Let $\mathcal{M}$ be a Riemannian

13 In the practical applications considered in [8], it was observed that the base point selection has only a minor impact on the final result.

manifold and suppose that there is a trajectory

$$c : [a, b] \to \mathcal{M}, \quad \mu \mapsto c(\mu)$$

on $\mathcal{M}$ that is sampled at k instants $\mu_1, \dots, \mu_k \in [a, b]$. Then an interpolant $\hat{c}$ for c can be computed via Algorithm 7.2.

Algorithm 7.2: Geodesic interpolation.

Input: Data set $\{c(\mu_1), \dots, c(\mu_k)\} \subset \mathcal{M}$ sampled from a curve $c : \mu \to c(\mu)$, unsampled instant $\mu^* \in [\mu_j, \mu_{j+1}]$.

1: Compute $v_{j+1} := \operatorname{Log}^{\mathcal{M}}_{c(\mu_j)}(c(\mu_{j+1})) \in T_{c(\mu_j)}\mathcal{M}$.

2: Compute $\hat{c}(\mu^*) := \operatorname{Exp}^{\mathcal{M}}_{c(\mu_j)}(\frac{\mu^* - \mu_j}{\mu_{j+1} - \mu_j} v_{j+1})$

Output: $\hat{c}(\mu^*) \in \mathcal{M}$ interpolant of $c(\mu^*)$.

The interpolants at $\mu \in [\mu_j, \mu_{j+1}]$ that are output by Algorithm 7.2 lie on the unique geodesic connection between the points $c(\mu_j)$ and $c(\mu_{j+1})$. Hence, it is the straightforward manifold analogue of linear interpolation and is base-point independent.

The generic formulation of Algorithm 7.1 allows one to employ higher-order interpolation methods. However, this does not necessarily lead to more accurate results: the overall error depends not only on the interpolation error within the tangent space but also on the distortion caused by mapping the data to a selected (fixed) tangent space; see Figure 7.3.

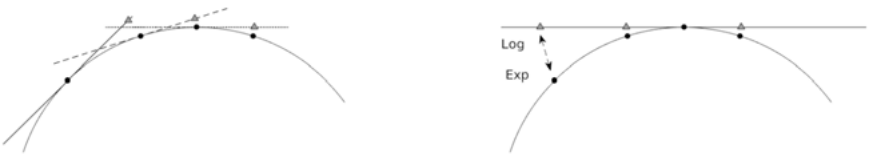


Figure 7.3: Illustration of the course of action of Algorithms 7.1 and 7.2. Algorithm 7.1 (right) first maps all data points to a selected fixed tangent space. In Algorithm 7.2 (left), two points $p_j = c(\mu_j)$ and $p_{j+1} = c(\mu_{j+1})$ are connected by a geodesic line, then the base is shifted to point p_{j+1} and the procedure is repeated.

Algorithms 7.1 and 7.2 can be applied in practical applications, where the Riemannian exponential and logarithm mappings are known in explicit form. Applications in parametric model reduction that consider matrix manifolds include [34] ($GL(n)$-data), [8, 76, 100] (Grassmann data), [104] (Stiefel data) and [9, 82] (SPD(n) data).

7.3.2 Interpolation via the Riemannian center of mass

As pointed out in Remark 7.2, interpolation of manifold data via the back and forth mapping of a complete data set of sample points between the manifold and its tan-

gent space depends on the chosen base point. As a consequence, sample points may experience an uneven distortion under the projection onto the tangent space; see Figure 7.3 (right). An approach that avoids this issue is to interpret interpolation as the task of finding suitably weighted Riemannian centers of mass. This concept was introduced in the context of geodesic finite elements in [44, 91].

The idea is as follows: The Riemannian center of mass[14] or Fréchet mean of a sample data set $\{p_1, \dots, p_k\} \in \mathcal{M}$ on a manifold with respect to the scalar weights $w_i \geq 0$, $\sum_{i=0}^{k} w_i = 1$ is defined as the minimizer(s) of the Riemannian objective function

$$\mathcal{M} \ni q \mapsto f(q) = \frac{1}{2} \sum_{i=1}^{k} w_i \operatorname{dist}(q, p_i)^2 \in \mathbb{R},$$

where $\operatorname{dist}(q, p_i)$ is the Riemannian distance of (7.5). This definition generalizes the notion of the barycentric mean in Euclidean spaces. However, on curved manifolds, the global center might not be unique. Moreover, local minimizers may appear. For more details, see [58] and [4], which also give uniqueness criteria.

Interpolation is now performed by computing weighted Riemannian centers. More precisely, let $\mu_1, \dots, \mu_k \subset \mathbb{R}^d$ be sampled parameter locations and let $p_i = p(\mu_i) \in \mathcal{M}$, $i = 1, \dots, k$ be the corresponding sample locations on $\mathcal{M}$. Interpolation is within the convex hull $\operatorname{conv}\{\mu_1, \dots, \mu_k\} \subset \mathbb{R}^d$ of the samples.

Let $\{\varphi_i : \mu \mapsto \varphi_i(\mu) \mid i = 1, \dots, k\}$ be a suitable set of interpolation functions with $\varphi_i(\mu_j) = \delta_{ij}$, say Lagrangians [91], splines [44] or radial basis functions [26]. Then the interpolant $p^* \approx p(\mu^*) \in \mathcal{M}$ at an unsampled parameter location $\mu^* \in \operatorname{conv}\{\mu_1, \dots, \mu_k\}$ is defined as the minimizer of

$$p^* = \operatorname*{arg\,min}_{q \in \mathcal{M}} f(q) = \frac{1}{2} \sum_{i=1}^{k} \varphi_i(\mu^*) \operatorname{dist}(q, p_i)^2. \tag{7.12}$$

At a sample location μ_j, one has indeed that $\sum_{i=1}^{k} \varphi_i(\mu_j) \operatorname{dist}(q, p_i)^2 = \sum_{i=1}^{k} \delta_{ij} \operatorname{dist}(q, p_i)^2 = \operatorname{dist}(q, p_j)^2$, which has the unique global minimum at $q = p_j$.

Computing p^* requires one to solve a Riemannian optimization problem. The simplest approach is a gradient descent method [3, 4]. The gradient of the objective function f in (7.12) is

$$\nabla f_q = -\sum_{i=1}^{k} \varphi_i(\mu^*) \operatorname{Log}_q^{\mathcal{M}}(p_i) \in T_q\mathcal{M}, \tag{7.13}$$

see [58, Thm 1.2], [4, §2.1.5], [91, eq. (2.4)]. Hence, just like interpolation in the tangent space, the interpolation via the Riemannian center can be pursued only in applications, where the Riemannian logarithm can be computed. A generic gradient descent

14 Here, we introduce this for discrete data sets; for centers w. r. t. a general mass distribution; see the original paper [58], Section 1.

algorithm to compute the barycentric interpolant for a function $p : \mathbb{R}^d \ni \mu \mapsto p(\mu) \in \mathcal{M}$ reads as follows.

Algorithm 7.3: Interpolation via the weighted Riemannian center [4, 84].

Input: Sample data set $\{p_1 = p(\mu_1), \dots, p_k = p(\mu_k)\} \subset \mathcal{M}$, unsampled parameter location $\mu^* \in \operatorname{conv}(\mu_1, \dots, \mu_k) \subset \mathbb{R}^d$, initial guess q_0, convergence threshold τ.

1: $k := 0$
2: Compute ∇f_{q_k} according to (7.13)
3: **while** $\|\nabla f_{q_k}\|_q > \tau$ **do**
4: select a step size α_k
5: $q_{k+1} := \operatorname{Exp}^{\mathcal{M}}_{q_k}(-\alpha_k \nabla f_{q_k})$
6: $k := k + 1$
7: **end while**

Output: $p^* := q_k \in \mathcal{M}$ interpolant of $p(\mu^*)$.

An implementation of this (type of) method for finding the Karcher mean in $SO(3)$ is discussed in [84]. Of course, Riemannian analogues to more sophisticated nonlinear optimization methods may also be employed; see [3].

In the context of model reduction, the benefits of interpolation via weighted Riemannian centers and the computational costs of solving the associated Riemannian optimization problem must be juxtaposed.

7.3.3 Additional approaches

A large variety of sophisticated ideas and further manifold interpolation techniques exist in the literature: The acceleration-minimizing property of cubic splines in the Euclidean space can be generalized to Riemannian manifolds in the form of a variational problem [24, 27, 33, 57, 77, 88, 93]; see also [81] and the references therein. Moreover, the construction concepts of Bézier curves and the De Casteljau algorithm [15] can be transferred to Riemannian manifolds [1, 62, 81, 75, 89]. Bézier curves in Euclidean spaces are polynomial splines that rely on a number of so-called control points. To obtain the value of a Bézier curve at time t, a recursive sequence of straight-line convex combinations between pairs of control points must be computed. The transition of this technique to Riemannian manifolds is via replacing the inherent straight lines with geodesics [81]. Another option is to conduct the Bézier–De Casteljau algorithm in the tangent space and to transfer the results to the manifold via a geodesic averaging of the spline arcs that were constructed in the tangent spaces at the first and the last control point, respectively; see [43].

Derivative information may also be incorporated in manifold interpolation schemes. A Hermite-type method that is specifically tailored for interpolation prob-

lems on the Grassmann manifold is sketched in [7, §3.7.4]. General Hermitian manifold interpolation in compact, connected Lie groups with a bi-invariant metric has been considered in [55]. A practical approach to first-order Hermite interpolation of data on arbitrary Riemannian manifolds is discussed in [103].

7.3.4 Quasi-linear extrapolation on matrix manifolds

In application scenarios, where both snapshot data of the full-order model and derivative information are at hand, various approaches have been suggested to exploit the latter. On the one hand, derivatives can be used for improving the ROMs accuracy and approximation quality by constructing POD bases that incorporate snapshots and snapshot derivatives [28, 51, 54, 99]. On the other hand, snapshot derivatives enable to parameterize the ROM bases and subspaces or to perform sensitivity analyses [48, 47, 97, 101]. In this section, we outline an approach to transferring the idea of extrapolation and parameterization via local linearizations to manifold-valued functions. The underlying idea is comparable to the trajectory piece-wise linear (TPWL) method; see [85] and Chapter 3 of this volume. Yet, TPWL linearizes the full-order model prior to the ROM projection, whereas here we consider linearizing ROM building blocks like the reduced orthogonal bases, reduced subspaces or reduced system matrices.

A geometric first-order Taylor approximation

Any differentiable function $f : \mathbb{R}^n \to \mathbb{R}^n$ can be linearized via a first-order Taylor expansion. A step ahead of size t in direction $d \in \mathbb{R}^n$ gives $f(x_0 + td) = f(x_0) + tDf_{x_0}(d) + \mathcal{O}(t^2)$. When considering $t \mapsto c(t) := f(x_0 + td)$ as a curve, then the first-order Taylor approximant is the straight line $g : t \mapsto c(0) + \dot{c}(0)t$. Such a first-order linearization often serves for extrapolating a given nonlinear function in a neighborhood of a selected expansion point. For doing so, the starting point $c(0)$ and the starting velocity $\dot{c}(0)$ must be available. This procedure translates to the manifold setting, when straight lines are replaced with geodesics.

Let $\mu \in \mathbb{R}$ be a scalar parameter and let $c : \mu \mapsto c(\mu) \in \mathcal{M}$ be a curve on a submanifold $\mathcal{M}$. For given initial values $c(\mu_0) = p_0 \in \mathcal{M}$ and $\dot{c}(\mu_0) = v_0 \in T_{p_0}\mathcal{M}$, the corresponding unique geodesic c_{p_0,v_0} is expressed via the Riemannian exponential as

$$c_{p_0,v_0} : \mu \to \mathcal{M}, \quad \mu \mapsto \mathrm{Exp}_{p_0}^{\mathcal{M}}(\mu v_0).$$

Algorithm 7.4: Geodesic extrapolation.

Input: Scalar parameter $\mu_0 \in \mathbb{R}$, initial values $c(\mu_0) \in \mathcal{M}, \dot{c}(\mu_0) \in T_{c(\mu_0)}\mathcal{M}$ sampled from a curve $c : \mu \to c(\mu) \in \mathcal{M}$, parameter value $\mu^* > 0$.

1: Compute $\hat{c}(\mu_0 + \mu^*) := \mathrm{Exp}_{c(\mu_0)}^{\mathcal{M}}(\mu^* \dot{c}(\mu_0))$

Output: $\hat{c}(\mu_0 + \mu^*) \in \mathcal{M}$ extrapolant of $c(\mu_0 + \mu^*)$.

Example: extrapolating POD basis matrices

As outlined in Section 7.1.1, snapshot POD works by collecting state vector snapshots, $x^1 := x(t_1, \mu_0), \dots, x^m := x(t_m, \mu_0) \in \mathbb{R}^n$ followed by an SVD of the snapshot matrix $(x^1, \dots, x^m)(\mu_0) =: \mathbb{S}(\mu_0) = \mathbb{U}(\mu_0)\Sigma(\mu_0)\mathbb{Z}^T(\mu_0)$. Here, the matrix dimensions are $\mathbb{U}(\mu_0) \in \mathbb{R}^{n\times m}$, $\Sigma(\mu_0) \in \mathbb{R}^{m\times m}$, $\mathbb{Z}(\mu_0) \in \mathbb{R}^{m\times m}$. The objective is to approximate $\mathbb{U}(\mu_0 + \mu)$ for a small $\mu > 0$ based on the data $\mathbb{U}(\mu_0), \dot{\mathbb{U}}(\mu_0)$, where $\mathbb{U}(\mu_0)$ is a point on the Stiefel manifold $St(n, m)$ and $\dot{\mathbb{U}}(\mu_0)$ is a tangent vector; see Section 7.4.4.1.

Differentiating the SVD. If the snapshot matrix function $\mu \mapsto \mathbb{S}(\mu) \in \mathbb{R}^{n\times m}$ is smooth in the neighborhood of $\mu_0 \in \mathbb{R}$ and if the singular values of $\mathbb{S}(\mu_0)$ are mutually distinct,[15] then the singular values and both the left and the right singular vectors are differentiable in $\mu \in [\mu_0 - \delta\mu, \mu_0 + \delta\mu]$ for $\delta\mu$ small enough. For brevity, let $\dot{\mathbb{S}} = \frac{d\mathbb{S}}{d\mu}(\mu_0)$ denote the derivative with respect to μ evaluated in μ_0 and so forth. Let $\mu \mapsto \mathbb{S}(\mu) = \mathbb{U}(\mu)\Sigma(\mu)\mathbb{Z}(\mu)^T \in \mathbb{R}^{n\times m}$ and let $C(\mu) = (\mathbb{S}^T\mathbb{S})(\mu)$. Let u^j and v^j, $j = 1, \dots, m$, denote the columns of $\mathbb{U}(\mu_0)$ and $\mathbb{Z}(\mu_0)$, respectively. We have

$$\dot{\sigma}_j = \left(u^j\right)^T \dot{\mathbb{S}} v^j, \quad (j = 1, \dots, m), \tag{7.14}$$

$$\dot{\mathbb{Z}} = \mathbb{Z}A, \quad \text{where } A_{ij} = \begin{cases} \frac{\sigma_j (u^j)^T \dot{\mathbb{S}} v^i + \sigma_i (u^i)^T \dot{\mathbb{S}} v^j}{(\sigma_j + \sigma_i)(\sigma_j - \sigma_i)}, & i \neq j \\ 0, & i = j \end{cases} \quad (i, j = 1, \dots, m), \tag{7.15}$$

$$\dot{\mathbb{U}} = \dot{\mathbb{S}}\mathbb{Z}\Sigma^{-1} + \mathbb{S}\dot{\mathbb{Z}}\Sigma^{-1} + \mathbb{S}\mathbb{Z}\dot{\Sigma}^{-1} = \left(\dot{\mathbb{S}}\mathbb{Z} + \mathbb{U}(\Sigma A - \dot{\Sigma})\right)\Sigma^{-1}. \tag{7.16}$$

A proof can be found in [48]. Note that $\mathbb{U}^T(\mu_0)\dot{\mathbb{U}}(\mu_0)$ is skew-symmetric so that indeed $\dot{\mathbb{U}}(\mu_0) =: \Delta(\mu_0) \in T_{\mathbb{U}(\mu_0)}St(n, m)$. The above equations hold in approximative form for the truncated SVD. For convenience, assume that $\mathbb{U}(\mu_0) \in St(n, r)$ is now the truncated to $r \leq m$ columns.

Performing the Taylor extrapolation on $St(n, r)$. With $\mathbb{U}(\mu_0), \dot{\mathbb{U}}(\mu_0)$ at hand, $\mathbb{U}(\mu_0 + \mu)$ can be approximated using the Stiefel exponential: $\mathbb{U}(\mu_0 + \mu) \approx \hat{\mathbb{U}}(\mu_0 + \mu) := \mathrm{Exp}^{St}_{\mathbb{U}_0}(\mu\dot{\mathbb{U}}(\mu_0))$; see Algorithm 7.7. The process is illustrated in Figure 7.4.

Note that when the μ-dependency is real-analytic, then the Euclidean Taylor expansion

$$\mathbb{U}(\mu_0 + \mu) = \mathbb{U}(\mu_0) + \mu\dot{\mathbb{U}}(\mu_0) + \frac{\mu^2}{2}\ddot{\mathbb{U}}(\mu_0) + \mathcal{O}(\mu^3) \in St(n, r) \tag{7.17}$$

converges to an orthogonal matrix $\mathbb{U}(\mu_0 + \mu) \in St(n, r)$. Yet, when truncating the Taylor series, we leave the Stiefel manifold. In particular, the columns of the first-order approximation are not orthonormal, i. e. $\mathbb{U}(\mu_0) + \mu\dot{\mathbb{U}}(\mu_0) \notin St(n, r)$ for $\mu \neq 0$. By construction, the Stiefel geodesic features the same starting velocity $\dot{\mathbb{U}}(\mu_0)$ and thus matches the Taylor series up to terms of second order. In addition, it respects the geometric structure of the Stiefel manifold and thus preserves column-orthonormality for every μ.

15 This condition can be relaxed; see the results of [5, §7].

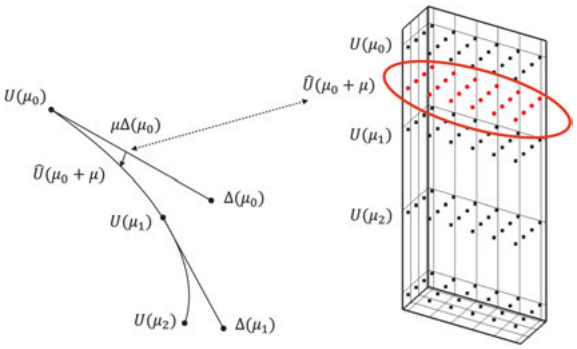


Figure 7.4: Extrapolation of matrix manifold data. Sketched on the right is the sample matrix data in $\mathbb{R}^{n\times r}$. The curved line on the left represents the nonlinear matrix manifold; the straight lines represent the tangent vectors in the tangent space. The matrix curve is linearized at $U(q_0)$, $U(q_1)$, etc.

7.4 Matrix manifolds of practical importance

In this section, we discuss the matrix manifolds that feature most often in practical applications in the context of model reduction. For each manifold under consideration, we recap, if applicable

- the representation of points/locations in numerical schemes.
- the representation of tangent vectors in numerical schemes.
- the most common Riemannian metrics.
- how to compute distances, geodesics and the Riemannian exponential and logarithm mappings.

7.4.1 The general linear group

This section is devoted to the general linear group $GL(n)$ of invertible square matrices. In model reduction, regular matrices appear for example as (reduced) system matrices in LTI and discretized PDE systems [9, 34, 78] and parameterizations have to be such that matrix regularity is preserved. In addition, the discussion of the seemingly simple matrix manifold $GL(n)$ is important, because it is the fundamental matrix Lie group from which all other matrix Lie groups are derived. Moreover, it provides the background for understanding quotient spaces of $GL(n)$; see Subsection 7.2.5 and also [23, 96]. A short summary on the Riemannian geometry of $GL(n)$ is given in [83, §6].

7.4.1.1 Introduction and data representation in numerical schemes

Because $GL(n) = \det^{-1}(\mathbb{R}\backslash\{0\}) = \{A \in \mathbb{R}^{n\times n} \mid \det(A) \neq 0\}$, $GL(n)$ is an open subset of the n^2-dimensional vector space $\mathbb{R}^{n\times n} \simeq \mathbb{R}^{n^2}$ and is thus an n^2-dimensional differentiable

manifold; see [66, Examples 1.22–1.27]. The matrix manifold $GL(n)$ is disconnected as it decomposes into two connected components, namely the regular matrices of positive determinant and the regular matrices of negative determinant.

Because $GL(n)$ is an open subset of the vector space $\mathbb{R}^{n\times n}$, the tangent space at a location $A \in GL(n)$ is simply $T_A GL(n) = \mathbb{R}^{n\times n}$. For $GL(n)$, the Lie algebra is $\mathfrak{gl}(n) = \mathbb{R}^{n\times n}$, so that the Lie group exponential is the standard matrix exponential $\exp_m : \mathbb{R}^{n\times n} = \mathfrak{gl}(n) \to GL(n)$. From the Lie group perspective (7.10), the tangent space at an arbitrary point $A \in GL(n)$ is to be considered as the set $T_A GL(n) = A\mathfrak{gl}(n) = A(\mathbb{R}^{n\times n})$, even though this set coincides with $\mathbb{R}^{n\times n}$.

7.4.1.2 Distances and geodesics

The obvious choice for a Riemannian metric on $GL(n)$ is to use the inner product from the ambient Euclidean matrix space, i. e.,

$$\langle \Delta, \tilde{\Delta} \rangle_A = \langle \Delta, \tilde{\Delta} \rangle_0 = \operatorname{trace}(\Delta^T \tilde{\Delta}),$$

for $A \in GL(n)$ and $\Delta, \tilde{\Delta} \in T_A GL(n) = \mathbb{R}^{n\times n}$.

In many applications, it is more appropriate to consider metrics with certain invariance properties.[16] A *left-invariant metric* can be obtained from the standard metric via

$$\langle \Delta, \tilde{\Delta} \rangle_A = \left\langle A^{-1}\Delta, A^{-1}\tilde{\Delta} \right\rangle_0, \quad A \in GL(n), \quad \Delta, \tilde{\Delta} \in T_A GL(n). \tag{7.18}$$

When formally considering $\Delta = AV, \tilde{\Delta} = A\tilde{V} \in T_A GL(n) = A\mathfrak{gl}(n)$ as left-translates of tangent vectors $V, \tilde{V} \in T_I GL(n) = \mathfrak{gl}(n)$, then this metric satisfies $\langle \Delta, \tilde{\Delta} \rangle_A = \langle V, \tilde{V} \rangle_0$. Alternatively, $\langle V, \tilde{V} \rangle_0 = \langle AV, A\tilde{V} \rangle_A$, which explains the name 'left-invariant'.

The Riemannian exponential and logarithm for the flat metric

When equipped with the Euclidean metric, $GL(n)$ is flat: since the tangent space is the full matrix space $\mathbb{R}^{n\times n}$, the geodesic equation (7.7) requires the acceleration of a geodesic curve to vanish completely. Hence, the geodesic that starts from $A \in GL(n)$ with velocity $\Delta \in \mathbb{R}^{n\times n}$ is the straight line $C(t) = A + t\Delta$. Note that the curve $t \mapsto C(t)$ may leave the manifold $GL(n)$ for some $t \in \mathbb{R}$ as it may hit a matrix with zero determinant. The formulas for the Riemannian exponential and logarithm mapping at a base point $A \in GL(n)$ are

$$\operatorname{Exp}_A^{GL} : T_A GL(n) \supset B_\varepsilon(0) \to GL(n), \quad \Delta \mapsto \tilde{A} := A + \Delta, \tag{7.19}$$

16 "Eulerian motion of a rigid body can be described as motion along geodesics in the group of rotations of three-dimensional euclidean space provided with a left-invariant Riemannian metric. A significant part of Euler's theory depends only upon this invariance, and therefore can be extended to other groups." [11, Appendix 2, p. 318].

$$\mathrm{Log}_A^{GL}: \quad GL(n) \to T_A GL(n), \qquad \tilde{A} \mapsto \Delta := (\tilde{A} - A). \tag{7.20}$$

In (7.19), $B_\varepsilon(0)$ denotes a suitably small open neighborhood around $0 \in T_A GL(n) \simeq \mathbb{R}^{n\times n}$ such that $A + \Delta \in GL(n)$ for all $\Delta \in B_\varepsilon(0)$.

The Riemannian exponential for the left-invariant metric on $GL(n)$

The left-invariant metric induces a non-flat geometry on $GL(n)$. Formulae for the covariant derivatives and the corresponding geodesics are derived in [10, Thm. 2.14]. The counterparts w. r. t. the right-invariant metrics can be found in [96]. Given a base point $A \in GL(n)$ and a starting velocity $\Delta = AV \in T_A GL(n) = A\mathfrak{gl}(n)$, the associated geodesic is

$$\Gamma_{A,\Delta}: t \mapsto A \exp_m(tV^T) \exp_m(t(V - V^T)). \tag{7.21}$$

The Riemannian exponential is

$$\begin{aligned} \mathrm{Exp}_M^{GL}(\Delta) = \Gamma_{A,\Delta}(1) &= A \exp_m(V^T) \exp_m(V - V^T) \\ &= A \exp_m((A^{-1}\Delta)^T) \exp_m((A^{-1}\Delta) - (A^{-1}\Delta)^T). \end{aligned} \tag{7.22}$$

The author is not aware of a closed formula for the inverse map, i. e., the Riemannian logarithm for the left-invariant metric; see also the discussion in [96, §4.5]. The thesis [83, §6.2] introduces a Riemannian shooting method for computing the Riemannian logarithm w. r. t. the left-invariant metric.

An important special case

For tangent vectors $\Delta = AV \in T_A GL(n)$ with *normal* $V \in \mathbb{R}^{n\times n}$, i. e., $VV^T = V^T V$, we have that the matrices V^T and $(V - V^T)$ commute. Therefore, according to (7.36), $A \exp_m(V^T) \exp_m(V - V^T) = A \exp_m(V^T + V - V^T) = A \exp_m(V)$ and the Riemannian exponential reduces to

$$\mathrm{Exp}_A^{GL}: T_A GL(n) \cap \{\Delta \mid A^{-1}\Delta \text{ normal}\} \to GL(n), \Delta \mapsto \tilde{A} = A \exp_m(A^{-1}\Delta).$$

The Riemannian logarithm is

$$\mathrm{Log}_A^{GL}: \mathcal{D}_A \cap \{\tilde{A} \mid A^{-1}\tilde{A} \text{ normal}\} \to T_A GL(n), \quad \tilde{A} \mapsto \Delta = A \log_m(A^{-1}\tilde{A}),$$

where $\mathcal{D}_A \subset GL(n)$ is a domain such that a suitable branch of the matrix logarithm is well-defined. These expressions are sometimes encountered in the literature as the Riemannian exponential and logarithm mappings. Yet, one should be aware of the fact that they hold under special circumstances.

7.4.2 The orthogonal group

This section is devoted to the *orthogonal group* $O(n) \subset \mathbb{R}^{n\times n}$ of orthogonal n-by-n matrices. In parametric model reduction, such matrices may appear as eigenvector matrices in symmetric EVD problems.

7.4.2.1 Introduction and data representation in numerical schemes

The orthogonal group is $O(n) = \{Q \in \mathbb{R}^{n\times n} \mid QQ^T = I = Q^TQ\}$. The manifold structure of $O(n)$ can be established via Theorem 7.1; see also Example 7.1. The orthogonal group decomposes into two connected components, namely the orthogonal matrices with determinant 1 and the orthogonal matrices with determinant –1. The former constitute the *special orthogonal group* $SO(n) = \{Q \in O(n) \mid \det(Q) = 1\}$. The orthogonal group is a closed subgroup of the Lie group $GL(n)$ and thus itself a Lie group (Section 7.2.5). The tangent space $T_IO(n)$ at the identity forms the Lie algebra associated with the Lie group $O(n)$. It coincides with the Lie algebra of $SO(n)$ and as such is denoted by $\mathfrak{so}(n) = T_ISO(n) = T_IO(n)$, [46, §3.3, 3.4]. The Lie algebra of $SO(n)$ is precisely the vector space of skew-symmetric matrices, $\mathfrak{so}(n) = \operatorname{skew}(n)$. According to (7.10), the tangent space at an arbitrary location Q is given by the translates (by left-multiplication) of the Lie algebra

$$T_QO(n) = Q\mathfrak{so}(n) = \{\Delta = QV \in \mathbb{R}^{n\times n} \mid V \in \operatorname{skew}(n)\},$$

which is the same as $\{\Delta \in \mathbb{R}^{n\times n} \mid Q^T\Delta = -\Delta^TQ\}$. The Lie exponential is

$$\exp_m|_{\mathfrak{so}(n)} : \mathfrak{so}(n) \to SO(n). \tag{7.23}$$

This restriction is a surjective map; see Appendix A. The dimensions of both $T_QO(n)$ and $O(n)$ are $\frac{1}{2}n(n-1)$.

7.4.2.2 Distances and geodesics

We follow up on the discussion in Section 7.4.1.1. For the orthogonal group, the Euclidean metric and the left-invariant metric coincide. Let $\Delta = QV, \tilde{\Delta} = Q\tilde{V} \in T_QO(n) = Q\mathfrak{so}(n)$. Then

$$\begin{aligned}\langle \Delta, \tilde{\Delta}\rangle_Q &= \langle Q^{-1}\Delta, Q^{-1}\tilde{\Delta}\rangle_0 = \langle V, \tilde{V}\rangle_0 \\ &= \operatorname{trace}(V^T\tilde{V}) = \operatorname{trace}(V^TQ^TQ\tilde{V}) = \langle \Delta, \tilde{\Delta}\rangle_I.\end{aligned}$$

In fact, the metric is also right-invariant, which makes it a *bi-invariant* metric; see [6, §2]. Bi-invariant metrics are important, because for Lie groups endowed with bi-invariant metrics, the Lie exponential map and the Riemannian exponential map at the identity coincide [6, Thm. 2.27, p. 40].

The Riemannian exponential and logarithm maps on $O(n)$

The Riemannian $O(n)$-exponential at a base point $Q \in O(n)$ sends a tangent vector $\Delta \in T_Q O(n)$ to the endpoint $\tilde{Q} \in O(n)$ of a geodesic that starts from Q with velocity vector Δ. Therefore, it provides at the same time an expression for the geodesic curves on $O(n)$. A formula for computing the Riemannian $O(n)$-exponential was derived in [36, §2.2.2]. Given $Q \in O(n)$, we have

$$\mathrm{Exp}_Q^{On} : T_Q O(n) \to O(n), \quad \Delta \mapsto \tilde{Q} := Q \exp_m(Q^T \Delta). \tag{7.24}$$

This result is also immediate from abstract Lie theory; see [6, Eq. (2.2) & Thm. 2.27].[17] The corresponding Riemannian logarithm on $O(n)$ is

$$\mathrm{Log}_Q^{On} : O(n) \supset \mathcal{D}_Q \to T_Q O(n), \quad \tilde{Q} \mapsto \Delta := Q \log_m(Q^T \tilde{Q}) \tag{7.25}$$

and is well defined on a neighborhood $\mathcal{D}_Q \subset O(n)$ around Q such that, for all $\tilde{Q} \in \mathcal{D}_p$, the orthogonal matrix $Q^T \tilde{Q}$ does not feature $\lambda = -1$ as an eigenvalue.

The Riemannian distance between orthogonal matrices

For given $Q, \tilde{Q} \in O(n)$ from the same connected component of $O(n)$, consider the EVD $Q^T \tilde{Q} = \Psi \Lambda \Psi^H$. Because $Q^T \tilde{Q}$ is orthogonal, we have $\Lambda = \mathrm{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$ and we assume that $\theta_1, \ldots, \theta_n \in (-\pi, \pi)$. The Riemannian distance is

$$\mathrm{dist}_{On}(Q, \tilde{Q}) = \| \mathrm{Log}_Q^{On}(\tilde{Q}) \|_Q = \| \log_m(\Lambda) \|_F = \left(\sum_{k=1}^{n} \theta_k^2 \right)^{\frac{1}{2}}.$$

The compact Lie group $SO(n)$ is a geodesically complete Riemannian manifold [6, Hopf–Rinow theorem, p. 31], and each two points of $SO(n)$ can be joined by a minimal geodesic.

7.4.3 The matrix manifold of symmetric positive definite matrices

This section is devoted to the matrix manifold SPD(n) of real, symmetric positive-definite n-by-n matrices. In model reduction, such matrices appear for example as

17 The Lie exponential is $\exp_m|_{\mathfrak{so}(n)} : \mathfrak{so}(n) \to SO(n)$, which is in the case at hand the Riemannian exponential at the identity, $\mathrm{Exp}_I^{SO} = \exp_m|_{\mathfrak{so}(n)}$. This translates to any other location via [6, Eq. (2.2)] as follows: Pick any $Q \in SO(n)$ and consider the mapping "left-multiplication by Q", i. e., $L_Q : SO(n) \to SO(n), P \mapsto QP$. Then the differential is $d(L_Q)_I : T_I SO(n) \to T_{L_Q(I)} SO(n), V \mapsto \Delta := QV$. Because L_Q is an isometry,

$$Q \,\mathrm{Exp}_I^{SO}(V) = L_Q\big(\mathrm{Exp}_I^{SO}(V)\big) = \mathrm{Exp}_{L_Q(I)}^{SO}\big(d(L_Q)_I(V)\big) = \mathrm{Exp}_Q^{SO}(QV),$$

which gives $\mathrm{Exp}_Q^{SO}(QV) = Q \,\mathrm{Exp}_I^{SO}(V) = Q \exp_m(Q^{-1}\Delta)$ and thus (7.24).

(reduced) system matrices in second-order parametric ODEs. For example, in linear structural or electrical dynamical systems, mass, stiffness and damping matrices are usually in SPD(n), [9, §4.2]. Moreover, positive definite matrices arise as Gramians of reachable and observable LTI systems in the context of balanced truncation; see Chapter 2 of this volume.

Related is the manifold of positive semi-definite matrices of fixed rank. It is investigated in [23, 96, 68]. An application in model reduction features in [67].

7.4.3.1 Introduction and data representation in numerical schemes

The set

$$\mathrm{SPD}(n) = \{A \in \mathrm{sym}(n) \mid x^T Ax > 0 \; \forall x \in \mathbb{R}^n \setminus \{0\}\}$$

is an open subset of the metric Hilbert space $(\mathrm{sym}(n), \langle\cdot,\cdot\rangle_0)$ of symmetric matrices. As such, it is a differentiable manifold [22, §6]. Moreover, it forms a *convex cone* [37, Example 2, p. 8], [71, §2.3], and can be realized as a quotient $\mathrm{SPD}(n) \simeq GL(n)/O(n)$. The latter is based on the fact that, for $A \in \mathrm{SPD}(n)$, matrix factorizations $A = ZZ^T$ with $Z \in GL(n)$ are invariant under orthogonal transformations $Z \mapsto ZQ$, $Q \in O(n)$, [23, §2, p.3].

Since SPD(n) is an open subset of the vector space sym(n), the tangent space is simply

$$T_A\mathrm{SPD}(n) = \mathrm{sym}(n). \tag{7.26}$$

The dimensions of both $T_A\mathrm{SPD}(n)$ and SPD(n) are $\frac{1}{2}n(n+1)$.

There is a smooth one-to-one correspondence between sym(n) and SPD(n). That is, every positive definite matrix can be written as the matrix exponential of a unique symmetric matrix, [39, Lem. 18.7, p. 472]. Put in different words, when restricted to sym(n), the standard matrix exponential

$$\exp_m : \mathrm{sym}(n) \to \mathrm{SPD}(n)$$

is a diffeomorphism, its inverse is the standard principal matrix logarithm

$$\log_m : \mathrm{SPD}(n) \to \mathrm{sym}(n);$$

see also [12, Thm. 2.8]. The group $GL(n)$ acts on SPD(n) via congruence transformations

$$g_X(A) = X^T AX, \quad X \in GL(n), A \in \mathrm{SPD}(n). \tag{7.27}$$

For additional background on SPD(n); see [72, 73, 79]. Applications in computer vision are presented in [31, 59].

7.4.3.2 Distances and geodesics

The literature knows a large variety of distance measures on SPD(n); see [56, Table 3.1, p. 56]. Yet, there are essentially two choices that are associated with inner products on the tangent space of SPD(n) and thus induce Riemannian metrics on the manifold SPD(n): the so-called *natural metric* and the *log-Euclidean metric*. Let $A \in \mathrm{SPD}(n)$ and let $\Delta, \tilde{\Delta} \in \mathrm{sym}(n)$ be two tangent vectors.

– The natural metric is

$$\langle \Delta, \tilde{\Delta} \rangle_A = \langle A^{-1/2} \Delta A^{-1/2}, A^{-1/2} \tilde{\Delta} A^{-1/2} \rangle_0 = \mathrm{trace}(A^{-1} \Delta A^{-1} \tilde{\Delta}),$$

see [22, §6, p. 201], [23]. It also goes by the name *trace matric*, [64, §XII.1, p.322]. In statistical applications, it is usually called the *affine-invariant metric* [70, 80].[18]

– The log-Euclidean metric is

$$\langle \Delta, \tilde{\Delta} \rangle_A = \langle D(\log_m)_A(\Delta), D(\log_m)_A(\tilde{\Delta}) \rangle_0;$$

see [12, eq. (3.5)].

For the natural metric, it is more appropriate to consider $\mathrm{sym}(n) = T_I\mathrm{SPD}(n)$ as the tangent space at the identity and the tangent space at an arbitrary location $A \in \mathrm{SPD}(n)$ as $T_A\mathrm{SPD}(n) = A^{1/2}(T_I\mathrm{SPD}(n))A^{1/2}$, which, of course, is nothing but a reparameterization of $\mathrm{sym}(n)$. From this perspective, we have for tangent vectors $\Delta = A^{1/2} V A^{1/2}, \tilde{\Delta} = A^{1/2} \tilde{V} A^{1/2}$

$$\langle \Delta, \tilde{\Delta} \rangle_A = \langle V, \tilde{V} \rangle_0.$$

The congruence transformations (7.27) are isometries of SPD(n) with respect to the natural metric, [64, Thm. XII.1.1, p. 324], [22, Lem. 6.1.1, p. 201]. See also the discussion in [80, §3].

By a standard pullback construction from differential geometry [35, Def. 2.2, Example 2.5], the log-Euclidean metric transfers the inner product $\langle \cdot, \cdot \rangle_0$ on $\mathrm{sym}(n)$ to SPD(n) via the matrix logarithm $\log_m : \mathrm{SPD}(n) \to \mathrm{sym}(n)$. In [12, eq. (3.5)], the authors take this construction one step further and use the $\exp_m$–$\log_m$-correspondence to define a multiplication that turns SPD(n) into a Lie group and, eventually, into a vector space. As such, it is a *flat* manifold, i. e. a Riemannian manifold with zero curvature. In this way, the computational challenges that come with dealing with data on nonlinear manifolds are circumvented.

---

18 The motivation is as follows: if $y = Ax + v_0$, $A \in GL(n)$ is an affine transformation of a random vector x, then the mean is transformed to $\bar{y} := A\bar{x} + v_0$ and the covariance matrix undergoes a congruence transformation $C_{yy} = E[(y - \bar{y})(y - \bar{y})^T] = AC_{xx}A^T$.

Which metric is to be preferred is problem-dependent; see the various contributions in [92] and [69]. Since the natural metric arises canonical both from the geometric approach, [64, §XII.1], and the matrix-algebraic approach [22, §6] and since staying with the standard matrix multiplication is consistent with the setting of solving dynamical systems in model reduction applications, we restrict the discussion of the Riemannian exponential and logarithm to the geometry that is based on the natural metric.

The SPD(n) exponential

The Riemannian SPD(n)-exponential at a base point $A \in \mathrm{SPD}(n)$ sends a tangent vector Δ to the endpoint $\tilde{A} \in \mathrm{SPD}(n)$ of a geodesic that starts from A with velocity vector Δ. Therefore, it provides at the same time an expression for the geodesic curves on SPD(n) with respect to the natural metric. Formulae for computing the SPD(n)-exponential can be found in [23], [80]. The reader preferring a matrix-analytic approach is referred to [22, §6].

Algorithm 7.5: Riemannian SPD(n)-exponential

Input: base point $A \in \mathrm{SPD}(n)$, tangent vector $\Delta \in T_A\mathrm{SPD}(n) = \mathrm{sym}(n)$
Output: $\tilde{A} := \mathrm{Exp}_A^{\mathrm{SPD}}(\Delta) = A^{\frac{1}{2}} \exp_m(A^{-\frac{1}{2}}\Delta A^{-\frac{1}{2}})A^{\frac{1}{2}}$.

Here, $A^{\frac{1}{2}}$ denotes the matrix square root of A; see Appendix A.

The SPD(n) logarithm

The Riemannian SPD(n)-logarithm at a base point $A \in \mathrm{SPD}(n)$ finds for another point $\tilde{A} \in \mathrm{SPD}(n)$ an SPD(n)-tangent vector Δ such that the geodesic that starts from A with velocity Δ reaches $\tilde{A}$ after an arc length of $\|\Delta\|_A = \sqrt{\langle \Delta, \Delta \rangle_A}$. Therefore, it provides for two given data points $A, \tilde{A} \in \mathrm{SPD}(n)$

- a solution to the geodesic endpoint problem: a geodesic that starts from A and ends at $\tilde{A}$;
- the Riemannian distance between the given points $A, \tilde{A}$.

Formulas for computing the SPD(n)-logarithm can be found in [23], [80].

Algorithm 7.6: Riemannian SPD(n)-logarithm.

Input: base point $A \in \mathrm{SPD}(n)$, location $\tilde{A} \in \mathrm{SPD}(n)$
Output: $\Delta := \mathrm{Log}_A^{\mathrm{SPD}}(\tilde{A}) = A^{\frac{1}{2}} \log_m(A^{-\frac{1}{2}}\tilde{A}A^{-\frac{1}{2}})A^{\frac{1}{2}}$.

Both Algorithms 7.5 and 7.6 require one to compute the spectral decomposition of n-by-n-matrices. The computational effort is $\mathcal{O}(n^3)$. In the context of parametric model reduction, the Riemannian exponential and logarithm maps are usually required for

reduced matrix operators [9]. If n denotes the dimension of the full state vectors and $r \ll n$ denotes the dimension of the reduced state vectors, then matrix exponentials for r-by-r-matrices are required, so that the computational effort reduces to $\mathcal{O}(r^3)$.

7.4.4 The Stiefel manifold

This section is devoted to the *Stiefel manifold* $St(n,r) \subset \mathbb{R}^{n\times r}$ of rectangular column-orthogonal n-by-r matrices, $r \leq n$. Points $U \in St(n,r)$ may be considered as orthonormal bases of cardinality r, or r-frames in $\mathbb{R}^n$. In model reduction, such matrices appear as orthogonal coordinate systems for low-order ansatz spaces that usually stem from a proper orthogonal decomposition or a singular value decomposition of given input solution data. Modeling data on the Stiefel manifold corresponds to data processing for orthonormal bases and thus allows for example for interpolation/parameterization of POD subspace bases. The most important use case in model reduction is where the Stiefel matrices are tall and skinny, i. e., $r \ll n$. Interpolation problems on the Stiefel manifold have not yet been considered in the model reduction context. The reference [62] discusses interpolation of Stiefel data; however, using quasi-geodesics rather than geodesics. Reference [103] includes numerical experiments for interpolating orthogonal frames on the Stiefel manifold that relies on the canonical Riemannian Stiefel logarithm [83, 102].

7.4.4.1 Introduction and data representation in numerical schemes

The *Stiefel manifold* is the compact, homogeneous matrix manifold of column-orthogonal rectangular matrices

$$St(n,r) := \{U \in \mathbb{R}^{n\times r} \mid U^T U = I_r\}.$$

The manifold structure can be directly established via Theorem 7.1 in a similar way as in Example 7.1. An alternative approach is via Example 7.4, where $St(n,r)$ is identified with the quotient space $St(n,r) \cong O(n)/(I_r \times O(n-r))$ under actions of the closed subgroup $I_r \times O(n-r) := \{\left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) \mid R \in O(n-r)\} \leq O(n)$. Two square orthogonal matrices in $O(n)$ are identified as the same point on $St(n,r)$, if their first r columns coincide; see [36, §2.4].

For any matrix representative $U \in St(n,r)$, the tangent space of $St(n,r)$ at U is represented by

$$T_U St(n,r) = \{\Delta \in \mathbb{R}^{n\times r} \mid U^T\Delta = -\Delta^T U\} \subset \mathbb{R}^{n\times r}.$$

Every tangent vector $\Delta \in T_U St(n,r)$ may be written as

$$\Delta = UA + (I - UU^T)T, \quad A \in \mathbb{R}^{r\times r} \text{ skew}, \quad T \in \mathbb{R}^{n\times r} \text{arbitrary}, \tag{7.28}$$

$$\Delta = UA + U^{\perp}B, \quad A \in \mathbb{R}^{r\times r} \text{ skew}, \quad B \in \mathbb{R}^{(n-r)\times r} \text{ arbitrary}, \tag{7.29}$$

where, in the latter case, $U^{\perp} \in St(n, n-r)$ is such that $(U, U^{\perp}) \in O(n)$ is a square orthogonal matrix. The dimension of both $T_U St(n,r)$ and $St(n,r)$ is $nr - \frac{1}{2}r(r+1)$. For additional background and applications, see [3, 21, 29, 36, 52, 95].

7.4.4.2 Distances and geodesics

Let $U \in St(n,r)$ be a point and let $\Delta = UA + U^{\perp}B$, $\tilde{\Delta} = U\tilde{A} + U^{\perp}\tilde{B} \in T_U St(n,r)$ be tangent vectors. There are two standard metrics on the Stiefel manifold.

- The *Euclidean metric* on $T_U St(n,r)$ is the one inherited from the ambient $\mathbb{R}^{n\times r}$:

$$\langle \Delta, \tilde{\Delta} \rangle_0 = \operatorname{trace}(\Delta^T \tilde{\Delta}) = \operatorname{trace} A^T \tilde{A} + \operatorname{trace} B^T \tilde{B}.$$

- The *canonical metric* on $T_U St(n,r)$

$$\langle \Delta, \tilde{\Delta} \rangle_U = \operatorname{trace}\left(\Delta^T \left(I - \frac{1}{2}UU^T\right)\tilde{\Delta}\right) = \frac{1}{2}\operatorname{trace} A^T \tilde{A} + \operatorname{trace} B^T \tilde{B}$$

  is derived from the quotient representation $St(n,r) = O(n)/(I_r \times O(n-r))$ of the Stiefel manifold.

The canonical metric counts the *independent* coordinates[19] of a tangent vector equally, when measuring the length $\sqrt{\langle \Delta, \Delta \rangle_U}$ of a tangent vector $\Delta = UA + U^{\perp}B$, while the Euclidean metric disregards the skew-symmetry of A [36, §2.4]. Recall that different metrics entail different measures for the lengths of curves and thus different formulae for geodesics.

The Stiefel exponential

The Riemannian Stiefel exponential at a base point $U \in St(n,r)$ sends a Stiefel tangent vector Δ to the endpoint $\tilde{U} \in St(n,r)$ of a geodesic that starts from U with velocity vector Δ. Therefore, it provides at the same time an expression for geodesic curves on $St(n,r)$.

A closed-form expression for the Stiefel exponential w. r. t. Euclidean metric is included in [36, §2.2.2],

$$\tilde{U} = \operatorname{Exp}^{St}_U(\Delta) = (U, \Delta) \exp_m \left(\begin{pmatrix} U^T\Delta & -\Delta^T\Delta \\ I_p & U^T\Delta \end{pmatrix} \right) \begin{pmatrix} I_p \\ 0 \end{pmatrix} \exp_m(-U^T\Delta).$$

In [53], an alternative formula is derived that features only matrix exponentials of skew-symmetric matrices. An efficient algorithm for computing the Stiefel exponential w. r. t. the canonical metric was derived in [36, §2.4.2]:

19 That is, the upper triangular entries of the skew-symmetric A and the entries of B of $\Delta = UA + U^{\perp}B$.

Algorithm 7.7: Stiefel exponential [36].

Input: base point $U \in St(n,r)$, tangent vector $\Delta \in T_U St(n,r)$

1: $A := U^T \Delta$ {horizontal component, skew}

2: $QR := \Delta - UA$ {(thin) qr-decomp. of normal component of Δ.}

3: $\begin{pmatrix} A & -R^T \\ R & 0 \end{pmatrix} = T \Lambda T^H \in \mathbb{R}^{2r \times 2r}$ {EVD of skew-symmetric matrix}

4: $\begin{pmatrix} M \\ N \end{pmatrix} := T \exp_m(\Lambda) T^H \begin{pmatrix} I_r \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{2r \times r}$

Output: $\tilde{U} := \mathrm{Exp}_U^{St}(\Delta) = UM + QN \in St(n,r)$

In applications, where $\mathrm{Exp}_U^{St}(\mu\Delta)$ needs to be evaluated for various parameters μ as in the example of Section 7.3.4, steps 1.–3. should be computed a priori (offline). Apart from elementary matrix multiplications, the algorithm requires one to compute the standard matrix exponential of a skew-symmetric matrix. This, however, is for a $2r$-by-$2r$-matrix and does not scale in the dimension n. With the usual assumption of model reduction that $n \gg p$, the computational effort is $\mathcal{O}(nr^2)$.

The Stiefel logarithm

The Riemannian Stiefel logarithm at a base point $U \in St(n,r)$ finds for another point $\tilde{U} \in St(n,r)$ a Stiefel tangent vector Δ such that the geodesic that starts from U with velocity Δ reaches $\tilde{U}$ after an arc length of $\|\Delta\|_U = \sqrt{\langle \Delta, \Delta \rangle_U}$. Therefore, it provides for two given data points $U, \tilde{U} \in St(n,r)$

- a solution to the geodesic endpoint problem: a geodesic that starts from U and ends at $\tilde{U}$;
- the Riemannian distance between the given points $U, \tilde{U}$.

An efficient algorithm for computing the Stiefel logarithm w. r. t. the canonical metric was derived in [102].

Algorithm 7.8: Stiefel logarithm [102].

Input: base point $U \in St(n,r)$, $\tilde{U} \in St(n,r)$ 'close' to base point, $\tau > 0$ convergence threshold

1: $M := U^T \tilde{U} \in \mathbb{R}^{r \times r}$

2: $QN := \tilde{U} - UM \in \mathbb{R}^{n \times r}$ {(thin) qr-decomp. of normal component of $\tilde{U}$}

3: $V_0 := \begin{pmatrix} M & X_0 \\ N & Y_0 \end{pmatrix} \in O(2r)$ {compute orth. completion of the block $\begin{pmatrix} M \\ N \end{pmatrix}$}

4: **for** $k = 0, 1, 2, \ldots$ **do**

5: $\quad \begin{pmatrix} A_k & -B_k^T \\ B_k & C_k \end{pmatrix} := \log_m(V_k)$ {matrix log of orth. matrix}

6: $\quad$ **if** $\|C_k\|_2 \leq \tau$ **then**

7: $\quad\quad$ break

8: **end if**
9: $\Phi_k := \exp_m(-C_k)$ {matrix exp of skew matrix}
10: $V_{k+1} := V_k W_k$, where $W_k := \begin{pmatrix} I_r & 0 \\ 0 & \Phi_k \end{pmatrix}$
11: **end for**
Output: $\Delta := \mathrm{Log}_U^{St}(\tilde{U}) = UA_k + QB_k \in T_U St(n,r)$

The analysis in [102] shows that the algorithm is guaranteed to converge if the input data points $U, \tilde{U}$ are at most a Euclidean distance of $d = \|U - \tilde{U}\|_2 \leq 0.09$ apart. In this case, the algorithm exhibits a linear rate of convergence that depends on d but is smaller than $\frac{1}{2}$. In practice, the algorithm seems to converge, whenever the initial V_0 is such that its standard matrix logarithm $\log_m(V_0)$ is well-defined. Note that two points on $St(n,r)$ can at most be a Euclidean distance of 2 away from each other.

Apart from elementary matrix multiplications, the algorithm requires one to compute the standard matrix logarithm of an orthogonal $2r$-by-$2r$-matrix and the standard matrix exponential of a skew-symmetric r-by-r-matrix at every iteration k. Yet, these operations are independent of the dimension n. With the usual assumption of model reduction that $r \ll n$, the computational effort is $\mathcal{O}(nr^2)$.

For the Stiefel manifold equipped with the Euclidean metric, methods for calculating the Stiefel logarithm are introduced in [25].

7.4.5 The Grassmann manifold

This section is devoted to the *Grassmann manifold* $Gr(n,r)$ of r-dimensional subspaces of $\mathbb{R}^n$ for $r \leq n$. Every point $\mathcal{U} \in Gr(n,r)$, i. e., every subspace may be represented by selecting a basis $\{u^1, \ldots, u^r\}$ with $\mathrm{ran}(u^1, \ldots, u^r) = \mathcal{U}$. In numerical schemes, we work exclusively with orthonormal bases. In this way, points $\mathcal{U}$ on the Grassmann manifold are to be represented by points $U \in St(n,r)$ on the Stiefel manifold via $\mathcal{U} = \mathrm{ran}(U)$. For details and theoretical background, see the references [2, 3, 36]. Modeling data on the Grassmann manifold corresponds to data processing for subspaces and thus allows, for example, for the interpolation/parameterization of POD subspaces see [19, Chapter 5], [19, Chapter 9]. The most important use case in model reduction is where the subspaces are of low dimension when compared to the surrounding state space, i. e., $n \gg p$. Grassmann interpolation problems in the context of projection-based parametric model reduction are considered in [8, 76, 100, 87]. Subspaces also feature in Krylov subspace approaches; see [20, Chapter 3].

7.4.5.1 Introduction and data representation in numerical schemes

The set of all r-dimensional subspaces $\mathcal{U} \subset \mathbb{R}^n$ forms the *Grassmann manifold*

$$Gr(n,r) := \{\mathcal{U} \subset \mathbb{R}^n \mid \mathcal{U} \text{ subspace}, \dim(\mathcal{U}) = r\}.$$

The Grassmann manifold is a quotient of $O(n)$ under the action of the Lie subgroup $O(r) \times O(n-r) = \{\left(\begin{smallmatrix} S & 0 \\ 0 & R \end{smallmatrix}\right) \mid S \in O(r), R \in O(n-r)\} \leq O(n)$. Two matrices $Q, \tilde{Q} \in O(n)$ are in the same $(O(r) \times O(n-r))$-orbit, if and only if the first r columns of Q and $\tilde{Q}$ span the same subspace and the tailing $n-r$ columns span the corresponding orthogonal complement subspace. Theorem 7.5 applies and shows that $Gr(n,r) = O(n)/(O(r) \times O(n-r))$ is a homogeneous manifold.

Alternatively, the Grassmann manifold can be realized as a quotient manifold of the Stiefel manifold with the help of Theorem 7.4,

$$Gr(n,r) = St(n,r)/O(r) = \{[U] \mid U \in St(n,r)\}, \tag{7.30}$$

where the $O(r)$-orbits are $[U] = \{UR \mid R \in O(r)\}$. A matrix $U \in St(n,r)$ is called a *matrix representative* of a subspace $\mathcal{U} \in Gr(n,r)$, if $\mathcal{U} = \text{ran}(U)$. The orbit $[U]$ and the subspace $\mathcal{U} = \text{ran}(U)$ are to be considered as the same object. For any matrix representative $U \in St(n,r)$ of $\mathcal{U} \in Gr(n,r)$ the tangent space of $Gr(n,r)$ at $\mathcal{U}$ is represented by

$$T_{\mathcal{U}}Gr(n,r) = \{\Delta \in \mathbb{R}^{n\times r} \mid U^T\Delta = 0\} \subset \mathbb{R}^{n\times r}.$$

Every tangent vector $\Delta \in T_{\mathcal{U}}Gr(n,r)$ may be written as

$$\Delta = (I - UU^T)T, \quad T \in \mathbb{R}^{n\times r} \text{ arbitrary, or,} \tag{7.31}$$

$$\Delta = U^{\perp}B, \quad B \in \mathbb{R}^{(n-r)\times r} \text{ arbitrary,} \tag{7.32}$$

where in the latter case, $U^{\perp} \in St(n,n-r)$ is such that $(U,U^{\perp}) \in O(n)$ is a square orthogonal matrix. The dimension of both $T_{\mathcal{U}}Gr(n,r)$ and $Gr(n,r)$ is $nr - r^2$.

7.4.5.2 Distances and geodesics

A metric on $T_{\mathcal{U}}Gr(n,r)$ can be obtained via making use of the fact that the Grassmannian is a quotient of the Stiefel manifold. Alternatively, one can restrict the standard inner matrix product $\langle A,B\rangle_0 = \text{trace}(A^TB)$ to the Grassmann tangent space. In the case of the Grassmannian, the two approaches lead to the same metric

$$\langle \Delta, \tilde{\Delta}\rangle_{\mathcal{U}} = \text{trace}(\Delta^T\tilde{\Delta}) = \langle \Delta, \tilde{\Delta}\rangle_0;$$

see [36, §2.5].

The Grassmann exponential

The Riemannian Grassmann exponential at a base point $\mathcal{U} \in Gr(n,r)$ sends a Grassmann tangent vector Δ to the endpoint $\tilde{\mathcal{U}} \in Gr(n,r)$ of a geodesic that starts from $\mathcal{U}$ with velocity vector Δ. Therefore, it provides at the same time an expression for the geodesic curves on $Gr(n,r)$. An efficient algorithm for computing the Grassmann exponential was derived in [36, §2.5.1]:

Algorithm 7.9: Grassmann exponential [36].

Input: base point $\mathcal{U} = [U] \in Gr(n,r)$, where $U \in St(n,r)$, tangent vector $\Delta \in T_U Gr(n,r)$

1: $Q\Sigma V^T \overset{\text{SVD}}{:=} \Delta$, with $Q \in St(n,r)$ {(thin) SVD of tangent vector}
2: $\tilde{U} := UV\cos(\Sigma)V^T + Q\sin(\Sigma)V^T$ {cos and sin act only on diag. entries.}

Output: $\tilde{\mathcal{U}} := \mathrm{Exp}^{Gr}_{\mathcal{U}}(\Delta) = [\tilde{U}] \in Gr(n,r)$.

Apart from elementary matrix multiplications, the algorithm requires one to compute the singular value decomposition of an n-by-r-matrix. The computational effort is $\mathcal{O}(nr^2)$.

The Grassmann logarithm

The Riemannian Grassmann logarithm at a base point $\mathcal{U} \in Gr(n,r)$ finds for another point $\tilde{\mathcal{U}} \in Gr(n,r)$ a Grassmann tangent vector Δ such that the geodesic that starts from $\mathcal{U}$ with velocity Δ reaches $\tilde{\mathcal{U}}$ after an arc length of $\|\Delta\|_{\mathcal{U}} = \sqrt{g^C_{\mathcal{U}}(\Delta,\Delta)}$. Therefore, it provides for two given data points $\mathcal{U}, \tilde{\mathcal{U}} \in Gr(n,r)$

- a solution to the geodesic endpoint problem: a geodesic that starts from $\mathcal{U}$ and ends at $\tilde{\mathcal{U}}$;
- the Riemannian distance between the given points $\mathcal{U}, \tilde{\mathcal{U}}$.

An algorithm for computing the Grassmann logarithm is stated implicitly in [2, §3.8, p. 210]. The reference [40] features expressions for the Grassmann exponential and the corresponding logarithm that formally work with Grassmann representatives in $SO(n)/(SO(r)\times SO(n-r))$ but also keep the computational effort $\mathcal{O}(nr^2)$. Reference [82, §4.3] gives the corresponding mappings after identifying subspaces with orthoprojectors; see also [16].

Algorithm 7.10: Grassmann Logarithm.

Input: base point $\mathcal{U} = [U] \in Gr(n,r)$ with $U \in St(n,r)$, $\tilde{\mathcal{U}} = [\tilde{U}] \in Gr(n,r)$ with $\tilde{U} \in St(n,r)$.

1: $M := U^T\tilde{U}$
2: $L := (I - UU^T)\tilde{U}M^{-1} = \tilde{U}M^{-1} - U$
3: $Q\Sigma V^T \overset{SVD}{:=} L$ {(thin) SVD }
4: $\Delta := Q\arctan(\Sigma)V^T$ {arctan acts only on diag. entries.}

Output: $\Delta = \mathrm{Log}^{Gr}_{\mathcal{U}}(\tilde{\mathcal{U}}) \in T_{\mathcal{U}}Gr(n,r)$

The composition $\mathrm{Exp}^{Gr}_{[U]} \circ \mathrm{Log}^{Gr}_{[U]}$ is the identity on $Gr(n,r)$, wherever it is defined. Yet, on the level of the actual matrix representatives, the operation

$$(\mathrm{Exp}^{Gr}_{[U]} \circ \mathrm{Log}^{Gr}_{[U]})([\tilde{U}_{in}]) = [\tilde{U}_{out}]$$

produces a matrix $\tilde{U}_{out} \neq \tilde{U}_{in}$. Directly recovering the input matrix can be achieved via a Procrustes-type preprocessing step, where $\tilde{U}$ is replaced with $\tilde{U}_* := \tilde{U}\Phi$, $\Phi = \arg\min_{\Phi \in O(r)} \|U - \tilde{U}\Phi\|$. This leads to the following.

Algorithm 7.11: Grassmann Logarithm: modified version.[20]

Input: base point $\mathcal{U} = [U] \in Gr(n,r)$ with $U \in St(n,r)$, $\tilde{\mathcal{U}} = [\tilde{U}] \in Gr(n,r)$ with $\tilde{U} \in St(n,r)$.

1: $\Psi S R^T \overset{\text{SVD}}{:=} \tilde{U}^T U$

2: $\tilde{U}_* := \tilde{U}(\Psi R^T)$ {'Transition to Procrustes representative'}

3: $L := (I - UU^T)\tilde{U}_*$

4: $Q\Sigma V^T \overset{\text{SVD}}{:=} L$ {(thin) SVD}

5: $\Delta := Q \arcsin(\Sigma) V^T$ {arcsin acts only on diagonal entries.}

Output: $\Delta = \text{Log}^{Gr}_{\mathcal{U}}(\tilde{\mathcal{U}}) \in T_{\mathcal{U}} Gr(n,r)$

An additional advantage of the modified Grassmann logarithm is that the matrix inversion $M^{-1} = (U^T\tilde{U})^{-1}$ is avoided. In fact, it is replaced by the SVD $\Psi S R^T = \tilde{U}^T U$ that is used to solve the Procrustes problem $\min_{\Phi \in O(r)} \|U - \tilde{U}\Phi\|$. The SVD exists also if $U^T\tilde{U}$ does not have full rank.

Distances between subspaces

The Riemannian logarithm provides the distance between two subspaces $\mathcal{U} = [U], \tilde{\mathcal{U}} = [\tilde{U}] \in Gr(n,r)$ as follows: First, compute $\Delta = \text{Log}^{Gr}_{\mathcal{U}}(\tilde{\mathcal{U}})$, then compute $\|\Delta\|_{\mathcal{U}} = \text{dist}_{Gr}(\mathcal{U}, \tilde{\mathcal{U}})$. In practice, however, this boils down to computing the singular values of the matrix $M = U^T\tilde{U}$, which can be seen as follows. By Algorithm 7.11, $\|\Delta\|^2_{\mathcal{U}} = \text{trace}(\Delta^T\Delta) = \sum_{k=1}^{p} \arcsin(\sigma_k)^2$, where the σ_k's are the singular values of $L = (I - UU^T)\tilde{U}_*$. These match precisely the square roots of the eigenvalues of L^TL. Using the SVD of the square matrix $\tilde{U}^TU = \Psi S R^T$ as in steps 1&2 of Algorithm 7.11, the eigenvalues of L^TL can be read off from

$$L^TL = \tilde{U}_*^T(I - UU^T)\tilde{U}_* = I - RS^2R^T = R(I - S^2)R^T,$$

so that $\sigma_k^2 = 1 - s_k^2$, when consistently ordered. As a consequence, $s_k = \sqrt{1 - \sigma_k^2} = \cos(\arcsin(\sigma_k))$, which implies

$$\text{dist}_{Gr}(\mathcal{U}, \tilde{\mathcal{U}}) = \left(\sum_{k=1}^{p} \arcsin(\sigma_k)^2 \right)^{\frac{1}{2}} = \left(\sum_{k=1}^{p} \arccos(s_k)^2 \right)^{\frac{1}{2}}, \tag{7.33}$$

where $\sigma_1, \ldots, \sigma_r$ and $s_1, \ldots, s_r$ are the singular values of L and $\tilde{U}^TU$, respectively.

The numerical linear algebra literature knows a variety of distance measures for subspaces. Essentially, all of them are based on the principal angles [36, §2.5.1, §4.3].

20 This is an original contribution to this chapter; for a detailed discussion, see [17, Section 5.2].

The *principal angles* (or canonical angles) $\theta_1, \dots, \theta_r \in [0, \frac{\pi}{2}]$ between two subspaces $[U], [\tilde{U}] \in Gr(n, r)$ are defined recursively by

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

The principal angles can be computed via $\theta_k := \arccos(s_k) \in [0, \frac{\pi}{2}]$, where s_k is the kth singular value of $U^T \tilde{U} \in \mathbb{R}^{r \times r}$ [42, §6.4.3]. Hence, the Riemannian subspace distance (7.33) expressed in terms of the principal angles is precisely

$$\operatorname{dist}([U], [\tilde{U}]) := \|\Theta\|_2, \quad \Theta = (\theta_1, \dots, \theta_r) \in \mathbb{R}^r. \tag{7.34}$$

In particular, (7.34) shows that any two points on $Gr(n, r)$ can be connected by a geodesic of length at most $\frac{\sqrt{r}}{2}\pi$; see also [98, Thm 8(b)].

7.5 Conclusion

Interpolation of structured matrices is a viable building block in parametric model reduction approaches. In order to preserve the characteristic features, the matrix sets in question are considered as geometric entities, so-called differentiable manifolds. In this chapter, we exposed how concepts from Riemannian geometry apply in designing manifold counterparts to Euclidean interpolation algorithms. As examples, the generic approach of interpolating in Riemannian normal coordinates, the quasi-linear, geodesic interpolation method and interpolation via the Riemannian center of mass are discussed. All the aforementioned methods share many of their constituent algorithmic units and acquaintance with these units allows one to adapt and modify the established approaches as needed or to design new ones. In this spirit, for a selection of matrix manifolds that feature frequently in practical applications, namely, the general linear group, the orthogonal group, the set of symmetric positive definite matrices, the Stiefel manifold and the Grassmann manifold, we have gathered the essential geometric concepts and formulas necessary to conduct Riemannian interpolation.

Appendix A

The matrix exponential and logarithm

The standard matrix exponential and matrix logarithm are defined via the power series

$$\exp_m(X) := \sum_{j=0}^{\infty} \frac{X^j}{j!}, \quad \log_m(X) := \sum_{j=1}^{\infty} (-1)^{j+1} \frac{(X - I)^j}{j}. \tag{7.35}$$

For $X \in \mathbb{R}^{n\times n}$, $\exp_m(X)$ is invertible with inverse $\exp_m(-X)$. The following restrictions of the exponential map are important:

$$\exp_m|_{\mathrm{sym}(n)} : \mathrm{sym}(n) \to \mathrm{SPD}(n), \quad \exp_m|_{\mathrm{skew}(n)} : \mathrm{skew}(n) \to SO(n).$$

The former is a diffeomorphism [79, Thm. 2.8], the latter is a differentiable, surjective map [41, §. 3.11, Thm. 9]. For additional properties and efficient methods for numerical computation, see [50, §10, 11].

A few properties of the exponential function for real or complex numbers carry over to the matrix exponential. However, since matrices do not commute, the standard exponential law is replaced by

$$\begin{aligned} \exp_m(Z(X,Y)) &= \exp_m(X)\exp_m(Y), \qquad (7.36)\\ Z(X,Y) &= X + Y + \frac{1}{2}[X,Y] \\ &\quad + \frac{1}{12}([X,[X,Y]] + [Y,[Y,X]]) - \frac{1}{24}[Y,[X,[X,Y]]]\ldots, \end{aligned}$$

where $[X,Y] = XY - YX$ is the commutator bracket, or Lie bracket. This is Dynkin's formula for the Baker–Campbell–Hausdorff series; see [86, §1.3, p. 22]. From a theoretical point of view, it is important that all terms in this series can be expressed in terms of the Lie bracket. A special case is

$$\exp_m(X+Y) = \exp_m(X)\exp_m(Y), \quad \text{if } [X,Y] = 0.$$

Matrix square roots and the polar decomposition

Every $S \in \mathrm{SPD}(n)$ has a unique matrix square root in $\mathrm{SPD}(n)$, i. e., a matrix denoted by $S^{\frac{1}{2}}$ with the property $S^{\frac{1}{2}}S^{\frac{1}{2}} = S$. This square root can be obtained via an EVD $S = Q\Lambda Q^T$ by setting

$$S^{\frac{1}{2}} := Q\sqrt{\Lambda}Q^T,$$

where $Q \in O(n)$, $\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ and $\lambda_i > 0$ are the eigenvalues of S. Every $A \in GL(n)$ can be uniquely decomposed into an orthogonal matrix times a symmetric positive definite matrix,

$$A = QP = Q\exp_m(X), \quad Q \in O(n), P \in \mathrm{SPD}(n), X \in \mathrm{sym}(n).$$

The polar factors can be constructed via taking the square root of the assuredly positive definite matrix A^TA and subsequently setting $P := (A^TA)^{\frac{1}{2}}$ and $Q := AP^{-1}$. Because the restriction of $\exp_m$ to the symmetric matrices is a diffeomorphism onto $\mathrm{SPD}(n)$, there is a unique $X \in \mathrm{sym}(n)$ with $P = \exp_m(X)$. For details, see [46, Thm. 2.18].

The Procrustes problem

Let $A, B \in \mathbb{R}^{n\times r}$. The Procrustes problem aims at finding an orthogonal transformation $R^* \in O(r)$ such that R^* is the minimizer of

$$\min_{R\in O(r)} \|A - BR\|_F.$$

The optimal R^* is $R^* = UV^T$, where $B^TA \overset{\text{SVD}}{=} U\Sigma V^T \in \mathbb{R}^{r\times r}$; see [42].

Bibliography

[1] P.-A. Absil, P.-Y. Gousenbourger, P. Striewski, and B. Wirth. Differentiable piecewise-Bézier surfaces on Riemannian manifolds. *SIAM J. Imaging Sci.*, 9(4):1788–1828, 2016.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.*, 80(2):199–220, 2004.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, New Jersey, 2008.

[4] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Control Optim.*, 51(3):2230–2260, 2013.

[5] D. Alekseevsky, A. Kriegl, P. W. Michor, and M. Losik. Choosing roots of polynomials smoothly. *Isr. J. Math.*, 105(1):203–233, 1998.

[6] M. M. Alexandrino and R. G. Bettiol. *Lie Groups and Geometric Aspects of Isometric Actions*. Springer, Cham, 2015.

[7] D. Amsallem. Interpolation on Manifolds of CFD-based Fluid and Finite Element-based Structural Reduced-order Models for On-line Aeroelastic Prediction. PhD thesis, Stanford University 2010.

[8] D. Amsallem and C. Farhat. Interpolation method for adapting reduced-order models and application to aeroelasticity. *AIAA J.*, 46(7):1803–1813, 2008.

[9] D. Amsallem and C. Farhat. An online method for interpolating linear parametric reduced-order models. *SIAM J. Sci. Comput.*, 33(5):2169–2198, 2011.

[10] E. Andruchow, G. Larotonda, L. Recht, and A. Varela. The left invariant metric in the general linear group. *J. Geom. Phys.*, 86:241–257, 2014.

[11] V. I. Arnol'd. *Mathematical Methods of Classical Mechanics*. *Graduate Texts in Mathematics*. Springer, New York, 1997.

[12] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 29(1):328–347, 2006.

[13] P. Astrid, S. Weiland, K. Willcox, and T. Backx. Missing points estimation in models described by proper orthogonal decomposition. *IEEE Trans. Autom. Control*, 53(10):2237–2251, 2008.

[14] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations. *C. R. Math. Acad. Sci. I*, 339:667–672, 2004.

[15] R. H. Bartels, J. C. Beatty, and B. A. Barsky. *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling*. *Morgan Kaufmann Series in Comp.* Elsevier Science, 1995.

[16] E. Batzies, K. Hüper, L. Machado, and F. Silva Leite. Geometric mean and geodesic regression on Grassmannians. *Linear Algebra Appl.*, 466:83–101, 2015.

[17] T. Bendokat, R. Zimmermann, and P.-A. P.-A.. A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects, 2020. arXiv preprint arXiv:2011.13699.
[18] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.*, 57(4):483–531, 2015.
[19] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. H. A. Schilders and L. M. Silveira. *Model Order Reduction. Volume 2: Snapshot-Based Methods and Algorithms*. De Gruyter, Berlin, 2020.
[20] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. H. A. Schilders and L. M. Silveira. *Model Order Reduction. Volume 3: Applications*. De Gruyter, Berlin, 2020.
[21] A. V. Bernstein and A. P. Kuleshov. Tangent bundle manifold learning via Grassmann & Stiefel eigenmaps, 2012. arXiv preprint arXiv:1212.6031.
[22] R. Bhatia. *Positive Definite Matrices*. *Princeton Series in Applied Mathematics*. Princeton University Press, Princeton, New Jersey, 2007.
[23] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.*, 31(3):1055–1070, 2009.
[24] N. Boumal and P.-A. Absil. A discrete regression method on manifolds and its application to data on SO(n). In *IFAC Proceedings Volumes, 18th IFAC World Congress*, volume 44(1), pages 2284–2289, 2011.
[25] D. Bryner. Endpoint geodesics on the Stiefel manifold embedded in Euclidean space. *SIAM J. Matrix Anal. Appl.*, 38(4):1139–1159, 2017.
[26] M. D. Buhmann. *Radial Basis Functions*. *Cambridge Monographs on Applied and Computational Mathematics*, volume 12. Cambridge University Press, Cambridge, UK, 2003.
[27] M. Camarinha, F. Silva Leite, and P. Crouch. On the geometry of Riemannian cubic polynomials. *Differ. Geom. Appl.*, 15(2):107–135, 2001.
[28] K. Carlberg and C. Farhat. A low-cost, goal-oriented 'compact proper orthogonal decomposition' basis for model reduction of state systems. *Int. J. Numer. Methods Eng.*, 86(3):381–402, 2011.
[29] R. Chakraborty and B. C. Vemuri. Statistics on the (compact) Stiefel manifold: Theory and applications. *Ann. Stat.*, 47(1):415–438, 2019.
[30] S. Chaturantabut and D. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM J. Sci. Comput.*, 32(5):2737–2764, 2010.
[31] A. Cherian and S. Sra. Positive definite matrices: Data representation and applications in computer vision. In H. Q. Minh and V. Murino, editors, *Algorithmic Advances in Riemannian Geometry and Applications: For Machine Learning, Computer Vision, Statistics, and Optimization*, pages 93–114. Springer, Cham, 2016.
[32] Y. Choi, D. Amsallem, and C. Farhat. Gradient-based constrained optimization using a database of linear reduced-order models *arXiv*, arXiv:1506.07849v1, pages 1–28, 2015.
[33] P. Crouch and F. Silva Leite. The dynamic interpolation problem: On Riemannian manifolds, Lie groups, and symmetric spaces. *J. Dyn. Control Syst.*, 1(2):177–202, 1995.
[34] J. Degroote, J. Vierendeels, and K. Willcox. Interpolation among reduced-order matrices to obtain parameterized models for design, optimization and probabilistic analysis. *Int. J. Numer. Methods Fluids*, 63(2):207–230, 2010.
[35] M. P. do Carmo. *Riemannian Geometry. Mathematics: Theory & Applications*. Birkhäuser, Boston, 1992.
[36] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998.
[37] J. Faraut and A. Koranyi. *Analysis on Symmetric Cones*. *Oxford Mathematical Monographs*. Oxford University Press, New York, 1994.
[38] T. Franz, R. Zimmermann, S. Görtz, and N. Karcher. Interpolation-based reduced-order modeling for steady transonic flows via manifold learning. *Int. J. Comput. Fluid Mech.*, 228:106–121, 2014. Special Issue on Reduced Order Modeling.

[39] J. H. Gallier. *Geometric methods and applications: for computer science and engineering*. *Texts in Applied Mathematics*. Springer, New York, 2011.
[40] K. A. Gallivan, A. Srivastava, X. Liu, and P. Van Dooren. Efficient algorithms for inferences on Grassmann manifolds. In *IEEE Workshop on Statistical Signal Processing*, pages 315–318, 2003.
[41] R. Godement and U. Ray. *Introduction to the Theory of Lie Groups*. *Universitext*. Springer, 2017.
[42] G. . H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, 4th edition, 2013.
[43] P.-Y. Gousenbourger, E. Massart, and P.-A. Absil. Data fitting on manifolds with composite Bézier-like curves and blended cubic splines. *J. Math. Imaging Vis.*, online:1–27, 2018.
[44] P. Grohs. Quasi-interpolation in Riemannian manifolds. *IMA J. Numer. Anal.*, 33(3):849–874, 2013.
[45] B. Haasdonk and M. Ohlberger. Efficient reduced models and a-posteriori error estimation for parametrized dynamical systems by offline/online decomposition. *Math. Comput. Model. Dyn. Syst.*, 17(2):145–161, 2011.
[46] B. C. Hall. *Lie Groups, Lie Algebras, and representations: An elementary introduction*. *Springer Graduate texts in Mathematics*. Springer, New York – Berlin – Heidelberg, 2nd edition, 2015.
[47] A. Hay, J. Borggaard, I. Akhtar, and D. Pelletier. Reduced-order models for parameter dependent geometries based on shape sensitivity analysis. *J. Comput. Phys.*, 229(4):1327–1352, 2010.
[48] A. Hay, J. T. Borggaard, and D. Pelletier. Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition. *J. Fluid Mech.*, 629:41–72, 2009.
[49] U. Helmke and J. B. Moore. *Optimization and Dynamical Systems*. *Communications & Control Engineering*. Springer, London, 1994.
[50] N. J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
[51] M. Hinze and S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In *Dimension Reduction of Large-Scale Systems*. *Lecture Notes in Computational Science and Engineering*, volume 45, pages 261–306. Springer, Berlin–Heidelberg, 2005.
[52] K. Hüper, M. Kleinsteuber, and F. Silva Leite. Rolling Stiefel manifolds. *Int. J. Syst. Sci.*, 39(9):881–887, 2008.
[53] K. Hüper and F. Ullrich. Real Stiefel manifolds: An extrinsic point of view. In *2018 13th APCA International Conference on Automatic Control and Soft Computing (CONTROLO)*, pages 13–18, 2018.
[54] K. Ito and S. S. Ravindran. A reduced-order method for simulation and control of fluid flows. *J. Comput. Phys.*, 143:403–425, 1998.
[55] J. Jakubiak, F. S. Leite, and R. Rodrigues. A two-step algorithm of smooth spline generation on Riemannian manifolds. *J. Comput. Appl. Math.*, 194:177–191, 2006.
[56] S. Jayasumana, R. Hartley, and M. Salzmann. Kernels on Riemannian manifolds. In A. Srivastava and P. K. Turaga, editors, *Riemannian computing in computer vision*, pages 45–67. Springer, 2015.
[57] H. Le, K. R. Kim, and I. L. Dryden. Smoothing splines on Riemannian manifolds, with applications to 3D shape space, 2018. arXiv:1801.04978v2.
[58] H. Karcher. Riemannian center of mass and mollifier smoothing. *Commun. Pure Appl. Math.*, 30(5):509–541, 1977.
[59] H. J. Kim, N. Adluru, B. B. Bendlin, S. C. Johnson, B. C. Vemuri, and V. Singh. Canonical correlation analysis on SPD(n) manifolds. In A. Srivastava and P. K. Turaga, editors, *Riemannian computing in computer vision*, pages 69–100. Springer, 2015.

[60] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry. Interscience Tracts in Pure and Applied Mathematics, no. 15*, volume I. John Wiley & Sons, New York – London – Sidney, 1963.

[61] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry. Interscience Tracts in Pure and Applied Mathematics, no. 15*, volume II. John Wiley & Sons, New York – London – Sidney, 1969.

[62] K. A. Krakowski, L. Machado, F. Silva Leite, and J. Batista. Solving interpolation problems on Stiefel manifolds using quasi-geodesics. In *Pré-Publicaçiões do Departamento de Matemática*, number 15–36, Universidade de Coimbra, 2015.

[63] W. Kühnel. *Differential Geometry: Curves – Surfaces – Manifolds. Student mathematical library*. American Mathematical Society, 2006.

[64] S. Lang. *Fundamentals of Differential Geometry. Graduate Texts in Mathematics*. Springer, New York, 2001.

[65] J. M. Lee. *Riemannian Manifolds: an Introduction to Curvature*. Springer, New York – Berlin – Heidelberg, 1997.

[66] J. M. Lee. *Introduction to Smooth Manifolds. Graduate Texts in Mathematics*. Springer, New York, 2012.

[67] E. Massart, P.-Y. Gousenbourger, T. S. Nguyen, T. Stykel, and P.-A. Absil. Interpolation on the manifold of fixed-rank positive-semidefinite matrices for parametric model order reduction: preliminary results. Technical Report UCL-INMA-2018.13, University of Louvain, 2018.

[68] E. Massart and P.-A. Absil. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *SIAM J. Matrix Anal. Appl.*, 41:171, 2020.

[69] H. Q. Minh and V. Murino. *Algorithmic Advances in Riemannian Geometry and Applications: For Machine Learning, Computer Vision, Statistics, and Optimization. Advances in Computer Vision and Pattern Recognition*. Springer, Cham, 2016.

[70] H. Q. Minh and V. Murino. From covariance matrices to covariance operators: Data representation from finite to infinite-dimensional settings. In H. Q. Minh and V. Murino, editors, *Algorithmic Advances in Riemannian Geometry and Applications: For Machine Learning, Computer Vision, Statistics, and Optimization*, pages 115–143. Springer, Cham, 2016.

[71] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 26(3):735–747, 2005.

[72] M. Moakher and P. G. Batchelor. Symmetric positive-definite matrices: From geometry to applications and visualization. In J. Weickert and H. Hagen, editors, *Visualization and Processing of Tensor Fields, Mathematics and Visualization*, pages 285–298. Springer, Berlin – Heidelberg, 2006.

[73] M. Moakher and M. Zéraï. The Riemannian geometry of the space of positive-definite matrices and its applications to the regularization of positive-definite matrix-valued data. *J. Math. Imaging Vis.*, 40(2):171–187, 2011.

[74] M. Morzyński, W. Stankiewicz, B. R. Noack, R. King, F. Thiele, and G. Tadmor. Continuous mode interpolation for control-oriented models of fluid flow. In R. King, editor, *Active Flow Control*, pages 260–278. Springer, Berlin – Heidelberg, 2007.

[75] E. Nava-Yazdani and K. Polthier. De Casteljau's algorithm on manifolds. *Comput. Aided Geom. Des.*, 30(7):722–732, 2013.

[76] T. S. Nguyen. A real time procedure for affinely dependent parametric model order reduction using interpolation on Grassmann manifolds. *Int. J. Numer. Methods Eng.*, 93(8):818–833, 2013.

[77] L. Noakes, G. Heinzinger, and B. Paden. Cubic splines on curved spaces. *IMA J. Math. Control Inf.*, 6(4):465–473, 1989.

[78] H. Panzer, J. Mohring, R. Eid, and B. Lohmann. Parametric model order reduction by matrix interpolation. *Automatisierungstechnik*, 58(8):475–484, 2010.
[79] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.*, 25(1):127, 2006.
[80] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *Int. J. Comput. Vis.*, 66(1):41–66, 2006.
[81] T. Popiel and L. Noakes. Bézier curves and C2 interpolation in Riemannian manifolds. *J. Approx. Theory*, 148(2):111–127, 2007.
[82] I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schröder. Multiscale representations for manifold-valued data. *SIAM J. Multiscale Model. Simul.*, 4(4):1201–1232, 2005.
[83] Q. Rentmeesters. Algorithms for data fitting on some common homogeneous spaces. PhD thesis, Université Catholique de Louvain, Louvain, Belgium 2013.
[84] Q. Rentmeesters and P.-A. Absil. Algorithms comparison for Karcher mean computation of rotation matrices and diffusion tensors. In *Proceedinngs of the 19th European Signal Processing Conference (EUSIPCO 2011)*, Barcelona, Spain, Aug. 29–Sept. 2, 2011.
[85] M. Rewienski and J. White. Model order reduction for nonlinear dynamical systems based on trajectory piecewise-linear approximations. *Linear Algebra Appl.*, 415(2-–3):426–454, 2006. Special Issue on Order Reduction of Large-Scale Systems.
[86] W. Rossmann. *Lie Groups: An Introduction Through Linear Groups*. *Oxford Graduate Texts in Mathematics*. Oxford University Press, 2006.
[87] T. S. Nguyen, P.-Y. Gousenbourger, E. Massart, and P.-A. Absil. Online balanced truncation for linear time-varying systems using continuously differentiable interpolation on grassmann manifold. Technical Report UCL-INMA-2019.01, University of Louvain, 2019.
[88] C. Samir, P.-A. Absil, A. Srivastava, and E. Klassen. A gradient-descent method for curve fitting on Riemannian manifolds. *Found. Comput. Math.*, 12(1):49–73, 2012.
[89] C. Samir and I. Adouani. C1 interpolating Bézier path on Riemannian manifolds, with applications to 3D shape space. *Appl. Math. Comput.*, 348:371–384, 2019.
[90] O. Sander. Interpolation und Simulation mit nichtlinearen Daten. *GAMM Rundbr.*, 1:6–12, 2015.
[91] O. Sander. Geodesic finite elements of higher order. *IMA J. Numer. Anal.*, 36(1):238–266, 2016.
[92] A. Srivastava and P. K. Turaga. *Riemannian computing in computer vision*. Springer, 2015.
[93] F. Steinke, M. Hein, J. Peters, and B. Schoelkopf. *Manifold-valued Thin-Plate Splines with Applications in Computer Graphics*. *Computer Graphics Forum*, 2008.
[94] G. Tadmor, O. Lehmann, B. R. Noack, and M. Morzyński. Galerkin models enhancements for flow control. In B. R. Noack, M. Morzyński and G. Tadmor, editors, *Reduced-Order Modelling for Flow Control*, pages 151–252. Springer, Vienna, 2011.
[95] P. K. Turaga, A. Veeraraghavan, and R. Chellappa. Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
[96] B. Vandereycken, P.-A. Absil, and S. Vandewalle. A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank. *IMA J. Numer. Anal.*, 33(2):481–514, 2012.
[97] G. Weickum, M. S. Eldred, and K. Maute. Multi-point extended reduced order modeling for design optimization and uncertainty analysis. In *Proceedings of the 2nd AIAA Multidisciplinary Design Optimization Specialist Conference, number AIAA 2006-2145*, Newport, RI, May 1–4, 2006.
[98] Y.-C. Wong. Differential geometry of Grassmann manifolds. *Proc. Natl. Acad. Sci. USA*, 57:589–594, 1967.

[99] R. Zimmermann. Gradient-enhanced surrogate modeling based on proper orthogonal decomposition. *J. Comput. Appl. Math.*, 237(1):403–418, 2013.
[100] R. Zimmermann. A locally parametrized reduced order model for the linear frequency domain approach to time-accurate computational fluid dynamics. *SIAM J. Sci. Comput.*, 36(3):B508–B537, 2014.
[101] R. Zimmermann. Local parametrization of subspaces on matrix manifolds via derivative information. In B. Karasözen, M. Manguoğlu, M. Tezer-Sezgin, S. Göktepe and Ö. Uğur, editors, *Numerical Mathematics and Advanced Applications ENUMATH 2015*, pages 379–387. Springer, Cham, 2016.
[102] R. Zimmermann. A matrix-algebraic algorithm for the Riemannian logarithm on the Stiefel manifold under the canonical metric. *SIAM J. Matrix Anal. Appl.*, 38(2):322–342, 2017.
[103] R. Zimmermann. Hermite interpolation and data processing errors on Riemannian matrix manifolds. *SIAM J. Sci. Comput.*, 42(5):A2593–A2619, 2020. https://doi.org/10.1137/19M1282878.
[104] R. Zimmermann and K. Debrabant. Parametric model reduction via interpolating orthonormal bases. In F. A. Radu, K. Kumar, I. Berre, D. N. Nordbotten and I. S. Pop, editors, *Numerical Mathematics and Advanced Applications ENUMATH 2017*. Springer, Cham, 2018.
[105] R. Zimmermann and K. Willcox. An accelerated greedy missing point estimation procedure. *SIAM J. Sci. Comput.*, 38(5):A2827–A2850, 2016.