Data processing on manifolds: Some basic ideas of Riemannian computing with applications

Computational Mathematics for Data Science

DTU, Nov. 15-17, 2023

Ralf Zimmermann
University of Southern Denmark (SDU)

Nov. 16, 2023

INSTITUT FOR MATEMATIK
OG DATALOGI

Outline

Outline

Section 1

Matrix manifolds, Lie groups, quotients

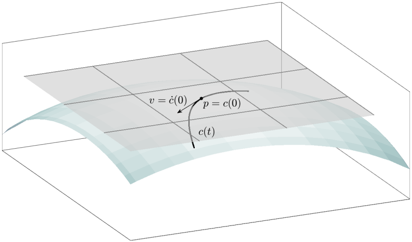
Riemannian Manifolds

Manifolds: Curved 'spaces' that locally look like the flat $\mathbb{R}^n$.

- coordinate charts around every point
- smooth transition between overlapping coordinate charts
  $\rightarrow$ foundation for calculus on manifolds ☺
- Riemannian: tangent spaces with a metric that changes smoothly with the manifold location
- in general: no vector space structure ☹

Riemannian Manifolds

Tangent spaces: local linearization of a manifold



- tangent vectors at $p \in \mathcal{M}$: velocity vectors of curves passing through p (Abstract setting: derivations, i.e., differential operators that induce directional derivatives)
- Option for constructing charts: one-to-one mappings between local manifold domain and tangent space domain

Matrix Manifolds

No generally accepted formal definition (that I am aware of).

Informally: *Sets of matrices (or equivalence classes of matrices), that share certain characterizing properties, which features a Riemannian manifold structure.*

Key idea: "points" = manifold locations represented by matrices

Matrix Manifolds

No generally accepted formal definition (that I am aware of).

Informally: *Sets of matrices (or equivalence classes of matrices), that share certain characterizing properties, which features a Riemannian manifold structure.*

Key idea: "points" = manifold locations represented by matrices

Examples:

- Invertible matrices $GL(n), SPD(n)$
- Matrix Lie groups, i.e., closed subgroups of $GL(n)$: $O(n), SO(n), SL(n), Sp(n), \ldots,$
- Quotients of matrix Lie groups: Stiefel, Grassmann, ...

Textbooks: [Absil et al., 2008], [Sato, 2021], [Boumal, 2023], ...

Numerical challenges

'Adding or subtracting two images of an automobile does not result in a valid image of an automobile.'
[Srivastava and Turaga, 2015, p. 2].

Numerical challenges

'Adding or subtracting two images of an automobile does not result in a valid image of an automobile.'
[Srivastava and Turaga, 2015, p. 2].

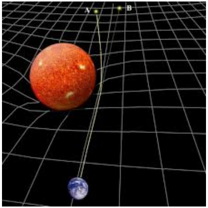
For matrix people:

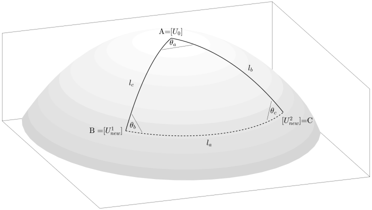
'Adding or subtracting two orthogonal matrices does not result in an orthogonal matrix.'

Similar for: eigenface spaces, computer tomography scans, covariance matrices, rotations in the Euclidean space, reduced-order subspaces, . . .

Numerical challenges

'Adding or subtracting two images of an automobile does not result in a valid image of an automobile.'
[Srivastava and Turaga, 2015, p. 2].
For matrix people:
'Adding or subtracting two orthogonal matrices does not result in an orthogonal matrix.'
Similar for: eigenface spaces, computer tomography scans, covariance matrices, rotations in the Euclidean space, reduced-order subspaces, ...





Shortest paths? Nearest neighbors? Barycenters?

Quotients of Lie groups

Definition (Lie groups)

A **Lie group** G is a differentiable manifold that at the same time forms an algebraic group such that the two group operations

- $G \times G \to G, \quad (g_1, g_2) \mapsto g_1 g_2$ "group multiplication"
- $G \to G, \quad g \mapsto g^{-1}$ "group inversion"

are differentiable.
A **matrix Lie group** matrix Lie group is a subgroup $G \leq GL(n)$ of the general linear group that is closed relative to $GL(n)$.

Definition (Quotients of Lie groups by closed subgroups, [Lee, 2012] §21)

Let G be a Lie group and $H \leq G$ be a Lie subgroup.

(i) For $g \in G$, a subset of G of the form

$$gH = \{gh| \quad h \in H\}$$

is called a **left coset of H**.

Definition (Quotients of Lie groups by closed subgroups, [Lee, 2012] §21)

Let G be a Lie group and $H \leq G$ be a Lie subgroup.

(i) For $g \in G$, a subset of G of the form

$$gH = \{gh| \quad h \in H\}$$

is called a **left coset of H**.

(ii) The set of left cosets is called the **left coset space of G modulo H**, in symbols G/H.

Definition (Quotients of Lie groups by closed subgroups, [Lee, 2012] §21)

Let G be a Lie group and $H \leq G$ be a Lie subgroup.

(i) For $g \in G$, a subset of G of the form

$$gH = \{gh| \quad h \in H\}$$

is called a **left coset of H**.

(ii) The set of left cosets is called the **left coset space of G modulo H**, in symbols G/H.

Theorem (cf. [Lee, 2012], Thm 21.17)

The left coset space G/H inherits a manifold structure such that the quotient map (the canonical projection) $\pi : G \to G/H$ is a smooth submersion.
Dimension: $\dim G/H = \dim G - \dim H$.

Quotient spaces: Why do we care?

- For a smooth submersion $\pi : G \to G/H$, we can split the tangent space at $p \in G$ into

$$T_pG = \ker(d\pi_p) \oplus \ker(d\pi_p)^{\perp} =: \mathcal{V}_p \oplus \mathcal{H}_p.$$

  (Forming the orthogonal complement is with respect to a selected Riemannian metric.)
  Horizontal space "=" tangent space of the quotient:

$$\mathcal{H}_p \cong T_{\pi(p)}G/H.$$

- **Geodesics** that are horizontal in the total space G are mapped to geodesics in the quotient G/H under π.
- In practical calculations, we can work with **horizontal lifts**.

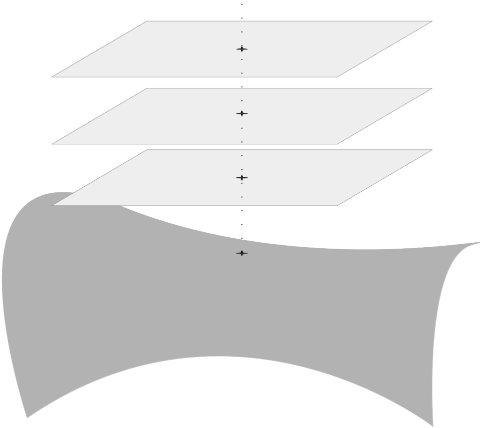


Figure 1: Various horizontal spaces at different points on the fibre. It holds $d\pi_p(\bar{v} + \mathcal{H}_p) = d\pi_p(\mathcal{H}_p)$ for any $\bar{v} \in \mathcal{V}_p$.
Each horizontal space may be used as an explicit representation of the tangent space of the quotient manifold.

Paradigm:

- know your geodesics in the total space
- check that geodesics that start horizontal, stay horizontal
- $\rightarrow$ you have found your geodesics in the quotient space. ☺
  No solving of ODEs required!

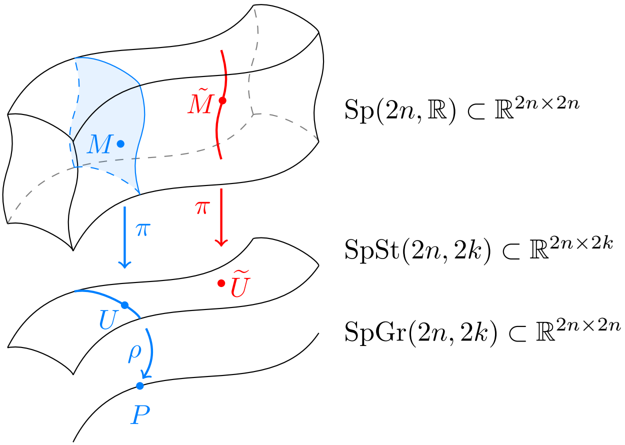
Paradigm:

- know your geodesics in the total space
- check that geodesics that start horizontal, stay horizontal
- $\rightarrow$ you have found your geodesics in the quotient space. ☺
  No solving of ODEs required!

Successfully applied

- to obtain geodesics on Stiefel- and Grassmann manifolds [Edelman et al., 1998]
- to obtain geodesics on symplectic Stiefel- and Grassmann manifolds [Bendokat and Z., 2021]

Example of a quotient structure: Symp. group, Symp. Stiefel, Symp. Grassmann



Graphic by Thomas Bendokat, taken from [Bendokat and Z., 2021]

Outline

Section 2

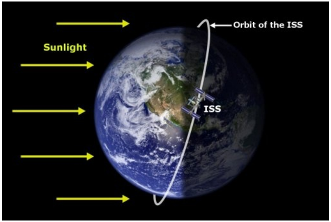
Geodesics matter

Geodesics

- intuitively: shortest connections, Riemannian counterparts to straight lines
- more precisely: stationary points of length functional $\rightarrow$ candidates for extrema

Basis of Riemannian computing: replace $p + tv$ with $c_{p,v}(t)$.

Geodesics



- intuitively: shortest connections, Riemannian counterparts to straight lines
- more precisely: stationary points of length functional $\rightarrow$ candidates for extrema
- characterized by zero covariant acceleration

Basis of Riemannian computing: replace $p + tv$ with $c_{p,v}(t)$.

Geodesic equation(s)

$(\mathcal{M}, g)$ Riemannian manifold with metric $g = (g_p(\cdot,\cdot))_{p\in\mathcal{M}}$.
Geodesic $c : [a, b] \to (\mathcal{M}, g)$ characterized by zero covariant derivative $\to$ ODE

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

Geodesic equation(s)

$(\mathcal{M}, g)$ Riemannian manifold with metric $g = (g_p(\cdot, \cdot))_{p \in \mathcal{M}}$.
Geodesic $c : [a, b] \to (\mathcal{M}, g)$ characterized by zero covariant derivative $\to$ ODE

- $\frac{D\dot{c}}{dt}(t) = 0 \quad \forall t \in [a, b]$.
- in local coordinates (U_φ, φ), $\gamma := \varphi \circ c|_{c^{-1}(U_\varphi)}$:

$$\ddot{\gamma}_k(t) + \sum_{i,j} \dot{\gamma}_i(t)\dot{\gamma}_j(t) \left(\Gamma_{ij}^k \circ \varphi^{-1}\right)(\gamma(t)) = 0 \quad \forall k = 1, \dots, n.$$

Christoffel symbols: $\Gamma_{ij}^k : U_\varphi \to \mathbb{R}$, defined by $\nabla_{\partial_i} \partial_j = \sum_k \Gamma_{ij}^k \partial_k$

Geodesic equation(s)

$(\mathcal{M}, g)$ Riemannian manifold with metric $g = (g_p(\cdot,\cdot))_{p\in\mathcal{M}}$.
Geodesic $c : [a, b] \to (\mathcal{M}, g)$ characterized by zero covariant derivative $\to$ ODE

- $\frac{D\dot{c}}{dt}(t) = 0 \quad \forall t \in [a, b]$.
- in local coordinates (U_φ, φ), $\gamma := \varphi \circ c|_{c^{-1}(U_\varphi)}$:

$$\ddot{\gamma}_k(t) + \sum_{i,j} \dot{\gamma}_i(t)\dot{\gamma}_j(t)\left(\Gamma_{ij}^k \circ \varphi^{-1}\right)(\gamma(t)) = 0 \quad \forall k = 1, \ldots, n.$$

Christoffel symbols: $\Gamma_{ij}^k : U_\varphi \to \mathbb{R}$, defined by $\nabla_{\partial_i}\partial_j = \sum_k \Gamma_{ij}^k \partial_k$

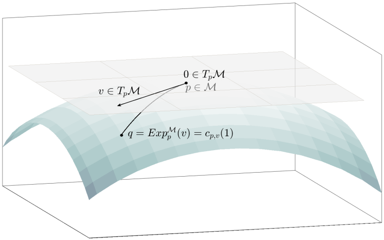
- in vector notation, using Christoffel tensor Γ

$$\ddot{\gamma} + \Gamma_{\gamma(t)}(\dot{\gamma}, \dot{\gamma}) = 0. \qquad \text{[Edelman et al., 1998]}$$

Riemannian normal coordinates

Definition (Riemannian Exponential)

$(\mathcal{M}, g)$ Riemannian manifold, $T_p^e\mathcal{M} := \{v \in T_p\mathcal{M} | \quad 1 \in I_v\}$
Riemannian exponential map at $p \in \mathcal{M}$:
$\mathrm{Exp}_p : T_p^e\mathcal{M} \to \mathcal{M}, \quad v \mapsto \mathrm{Exp}_p(v) := c_v(1)$.



Exp_p is a local diffeo.
$\mathrm{Log}_p = (\mathrm{Exp}_p)^{-1}$ is a coordinate chart.
Riemannian normal coordinates.
The manifold *'plus'* and *'minus'*. (R. Bergmann)

$+_{\mathcal{M}} : \mathrm{Exp}_p(v) = q \approx "p + v = q" \mid -_{\mathcal{M}} : \mathrm{Log}_p(q) = v \approx "q - p = v"$

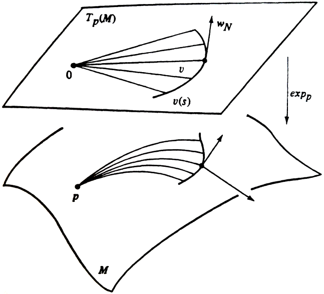
Riemannian normal coordinates

Fact: No isometries between flat and curved spaces possible.
($\rightarrow$ no map of earth that preserves lengths *and* angles.)
But for normal coordinates:

Riemannian normal coordinates

Fact: No isometries between flat and curved spaces possible.
($\rightarrow$ no map of earth that preserves lengths *and* angles.)
But for normal coordinates:

- lengths of geodesic rays are preserved
- geodesic sphere and geodesic rays intersect at right angle,



Gauß Lemma!

Illustration taken from [do Carmo, 1992, p. 69].

Retractions

Retractions: [Absil et al., 2008]

- Maps "tangent space $\rightarrow$ manifold" with derivative Id at 0.

$\Rightarrow$ 1^{st}-order approximations to geodesics/Riemannian exponential, locally invertible

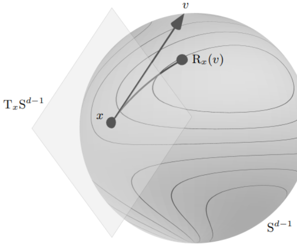
Retractions

Retractions: [Absil et al., 2008]

- Maps "tangent space $\rightarrow$ manifold" with derivative *Id* at 0.

$\Rightarrow$ 1^{st}-order approximations to geodesics/Riemannian exponential, locally invertible

Well-suited for optimization: Cheaper to evaluate. Do not compromise convergence results



taken from [Boumal, 2023, Fig. 3.1]

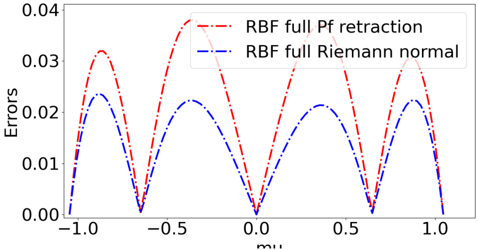
Retractions

Retractions: [Absil et al., 2008]

- Maps "tangent space $\to$ manifold" with derivative *Id* at 0.

$\Rightarrow$ 1^{st}-order approximations to geodesics/Riemannian exponential, locally invertible

Potential additional source of errors/geometry distortion.
Example: Stiefel data interpolation with polar factor retraction.



Red: coordinate charts based on polar factor retraction: RBF on tangent space. Blue: Riemannian normal coordinates: RBF on tangent space

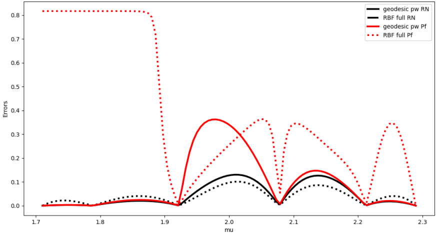
Retractions

Retractions: [Absil et al., 2008]

- Maps "tangent space $\to$ manifold" with derivative *Id* at 0.

$\Rightarrow$ 1^{st}-order approximations to geodesics/Riemannian exponential, locally invertible

Potential additional source of errors/geometry distortion.
Example: Stiefel data interpolation with polar factor retraction.



Red: coordinate charts based on polar factor retraction: piecewise geodesic and RBF on tangent space. Black: Riemannian normal coordinates: piecewise geodesic and RBF on tangent space

Retractions

Retractions: [Absil et al., 2008]

- Maps "tangent space $\rightarrow$ manifold" with derivative *Id* at 0.

$\Rightarrow$ 1^{st}-order approximations to geodesics/Riemannian exponential, locally invertible

Use of retractions can be a bare necessity!

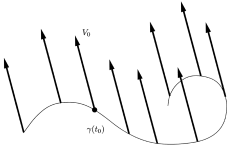
Geodesics on matrix manifolds often feature the matrix exponential.
$\Rightarrow$ Unstable for non-normal matrices.
Severe issue for Symplectic Stiefel geodesics [Bendokat and Z., 2021].
Remedy: Use, e.g., Cayley-trafo for retractions.

Outline

Subsection 2

The Christoffel symbols: Covariant derivatives and Riemannian Hessian

Covariant derivatives

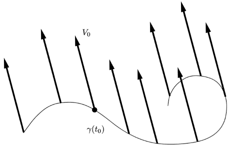


taken from [Lee, 2012, Fig. 4.7]

Let $t \mapsto X(t)$ be a vector field along a curve. Then

$$\frac{DX}{dt}(t) = \dot{X}(t) + \Gamma_{\gamma(t)}(X(t), \dot{\gamma}(t)).$$

Covariant derivatives



taken from [Lee, 2012, Fig. 4.7]

Let $t \mapsto X(t)$ be a vector field along a curve. Then

$$\frac{DX}{dt}(t) = \dot{X}(t) + \Gamma_{\gamma(t)}(X(t), \dot{\gamma}(t)).$$

Covariant derivatives yield

- parallel vector fields $\rightarrow$ parallel vector transport
- Riemannian Hessian $\rightarrow$ second-order optimization schemes

General recipe for computing the Hesse $(1,1)$-form.

- derive the geodesic ODE $\ddot{\gamma} + (\ldots) = 0$. The terms in red depend on $\gamma(t)$ and $\dot{\gamma}(t)$ and constitute the Christoffel tensor $\Gamma_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))) = (\ldots)$.

General recipe for computing the Hesse $(1,1)$-form.

- derive the geodesic ODE $\ddot{\gamma} + (\ldots) = 0$. The terms in red depend on $\gamma(t)$ and $\dot{\gamma}(t)$ and constitute the Christoffel tensor $\Gamma_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t)) = (\ldots)$.
- Find the general form via polarization $\Gamma(v, w) = \frac{1}{4}\left(\Gamma(v+w, v+w) - \Gamma(v-w, v-w)\right).$

General recipe for computing the Hesse $(1,1)$-form.

- derive the geodesic ODE $\ddot{\gamma} + (\ldots) = 0$. The terms in red depend on $\gamma(t)$ and $\dot{\gamma}(t)$ and constitute the Christoffel tensor $\Gamma_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))) = (\ldots)$.
- Find the general form via polarization $\Gamma(v,w) = \frac{1}{4}\left(\Gamma(v+w, v+w) - \Gamma(v-w, v-w)\right)$.
- Compute the Hessian of a scalar function f via the covariant derivative of the gradient field along a geodesic $t \mapsto \gamma(t)$ with starting velocity $\gamma(0) = p, \dot{\gamma}(0) = v$:

$$\begin{aligned}
\mathrm{Hess}f(p)[v] &= (\nabla_v \mathrm{grad}f)(p) = \frac{D}{dt}\Big|_{t=0} \mathrm{grad}f(\gamma(t)) \\
&= \frac{d}{dt}\Big|_{t=0} \mathrm{grad}f(\gamma(t)) + \Gamma_p(\mathrm{grad}f(p), v).
\end{aligned}$$

General recipe for computing the Hesse $(1,1)$-form.

- derive the geodesic ODE $\ddot{\gamma} + (\dots) = 0$. The terms in red depend on $\gamma(t)$ and $\dot{\gamma}(t)$ and constitute the Christoffel tensor $\Gamma_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))) = (\dots)$.
- Find the general form via polarization $\Gamma(v,w) = \frac{1}{4}\left(\Gamma(v+w, v+w) - \Gamma(v-w, v-w)\right)$.
- Compute the Hessian of a scalar function f via the covariant derivative of the gradient field along a geodesic $t \mapsto \gamma(t)$ with starting velocity $\gamma(0) = p, \dot{\gamma}(0) = v$:

$$\begin{aligned} \mathrm{Hess} f(p)[v] &= (\nabla_v \mathrm{grad} f)(p) = \frac{D}{dt}\Big|_{t=0} \mathrm{grad} f(\gamma(t)) \\ &= \frac{d}{dt}\Big|_{t=0} \mathrm{grad} f(\gamma(t)) + \Gamma_p(\mathrm{grad} f(p), v). \end{aligned}$$

Ongoing: applied for constructing a Riemann trust region method on $SpSpt(2n, 2k)$ by Rasmus Jensen.

"Reversed engineering"

What has happened here?

- geodesics from geometric/quotient considerations.
- use the solution to derive the underlying ODE

"Reversed engineering"

What has happened here?

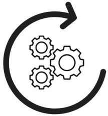
- geodesics from geometric/quotient considerations.
- use the solution to derive the underlying ODE
- use ODE to read off Christoffel tensor
- use Christoffel tensor to compute
  - covariant derivatives
  - parallel vector fields
  - Riemannian Hessian
  - ...

"Reversed engineering"

What has happened here?

- geodesics from geometric/quotient considerations.
- use the solution to derive the underlying ODE
- use ODE to read off Christoffel tensor
- use Christoffel tensor to compute
  - covariant derivatives
  - parallel vector fields
  - Riemannian Hessian
  - ...

Not *"Derive solutions to equations."*, but

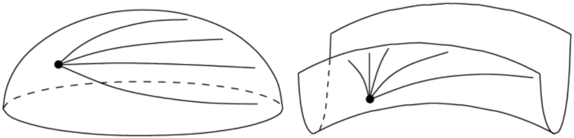


"Derive equations from solutions."

Outline

Subsection 3

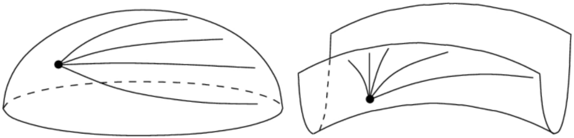
The impact of curvature



[Lee, 2018]

Jacobi fields

- Positive curvature: Geodesics bend towards each other
- Negative curvature: Geodesics spread apart



[Lee, 2018]

Jacobi fields

- Positive curvature: Geodesics bend towards each other
- Negative curvature: Geodesics spread apart

Consequence: Standard approach of data processing by

(1) mapping data onto the tangent space,
(2) processing data in tangent space,
(3) mapping the result back to manifold,

$$\left\{ \begin{array}{l} \text{is benign on positively curved manifolds (Stiefel, Grassmann).} \\ \text{adds extra errors on negatively curved manifolds.} \end{array} \right.$$

Error propagation

Theorem (Errors and curvature [Z., 2020])

Let $\mathcal{M}$ be a Riemannian manifold, $q \in \mathcal{M}$ and $\Delta, \tilde{\Delta} \in T_q\mathcal{M}$ $\epsilon := \|\Delta - \tilde{\Delta}\|$ and $\delta = \|\Delta\|$, $\tilde{\delta} = \|\tilde{\Delta}\|$. Assume that $\delta, \tilde{\delta} < 1$. Let $\sigma = span(\Delta, \tilde{\Delta}) \subset T_q\mathcal{M}$ and let $K(q, \sigma)$ be the sectional curvature at q w.r.t. σ.
The Riemannian distance between $p = \mathrm{Exp}_q^{\mathcal{M}}(\Delta)$ and $\tilde{p} = \mathrm{Exp}_q^{\mathcal{M}}(\tilde{\Delta})$ is

$$\mathrm{dist}_{\mathcal{M}}(p, \tilde{p}) \leq |\delta - \tilde{\delta}| + \epsilon(1 - \frac{K_q(\sigma)}{6}\delta + o(\delta^2)) + \mathcal{O}(\epsilon^2).$$

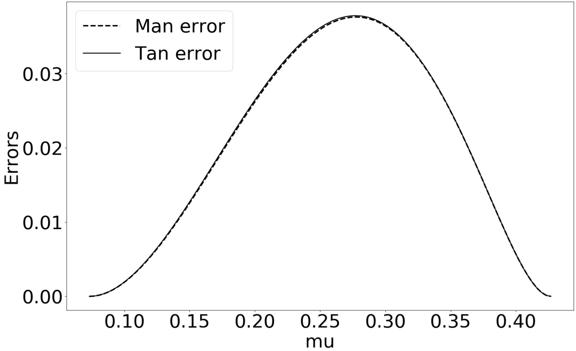


Figure 2: Interpolation of U-factor of parametric SVD data $U(\mu)\Sigma(\mu)V(\mu)^T \in \mathbb{R}^{10{,}000\times 300}$, rank$=10$. [Z., 2020]
Absolute (Hermite) interpolation errors in terms of the Riemannian metric on the tangent space (Tan error) and as measured by the Riemannian distance function on the manifold (Man error).

Curvature has an impact on the injectivity radius $i(\mathcal{M})$ and thus on the size of the domain on which one "can safely perform calculations".

Theorem ([do Carmo, 1992], §13, Prop. 2.13)

If the sectional curvature $K(p, \sigma)$ of a complete, compact Riemannian manifold $\mathcal{M}$ satisfies $K(p, \sigma) \leq C \ \forall p \in \mathcal{M}$ $\sigma \leq T_p\mathcal{M}$, with constant $C > 0$, then:

1. $i(\mathcal{M}) \geq \frac{\pi}{\sqrt{C}}$ or
2. there exists a closed geodesic whose length is less than that of any other closed geodesic, and which is such that $i(\mathcal{M}) = \frac{1}{2}L(\gamma)$.

Curvature has an impact on the injectivity radius $i(\mathcal{M})$ and thus on the size of the domain on which one "can safely perform calculations".

Theorem ([do Carmo, 1992], §13, Prop. 2.13)

If the sectional curvature $K(p, \sigma)$ of a complete, compact Riemannian manifold $\mathcal{M}$ satisfies $K(p, \sigma) \leq C \; \forall p \in \mathcal{M}$ $\sigma \leq T_p\mathcal{M}$, with constant $C > 0$, then:

1. $i(\mathcal{M}) \geq \frac{\pi}{\sqrt{C}}$ or
2. there exists a closed geodesic whose length is less than that of any other closed geodesic, and which is such that $i(\mathcal{M}) = \frac{1}{2}L(\gamma)$.

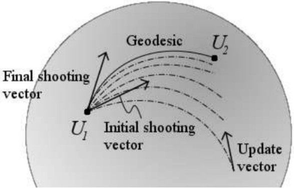
The 'or'-case does not provide a sharper bound for Stiefel. For Stiefel, case (1) is decisive.

Curvature has an impact on the iteration count:

- As a rule: manifold algorithms rely on local linearizations.
  For example: shooting methods to compute Stiefel logarithm [Z. and Hüper, 2022]:
  1 step Euclidean case $\leftrightarrow$ iteration of steps on manifold

Curvature has an impact on the iteration count:

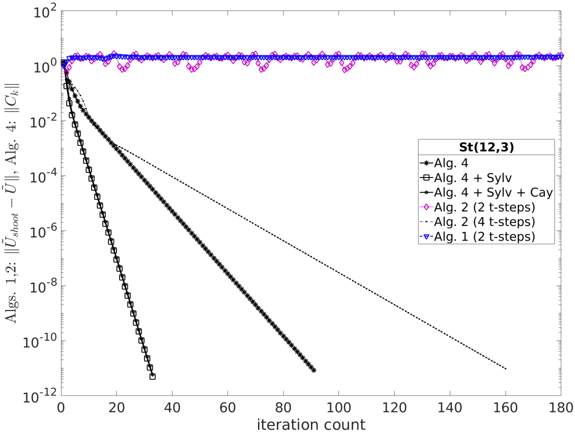
- As a rule: manifold algorithms rely on local linearizations.
  For example: shooting methods to compute Stiefel logarithm [Z. and Hüper, 2022]:
  1 step Euclidean case $\leftrightarrow$ iteration of steps on manifold



(Cartoon taken from [Bryner, 2017])

Curvature has an impact on the iteration count:

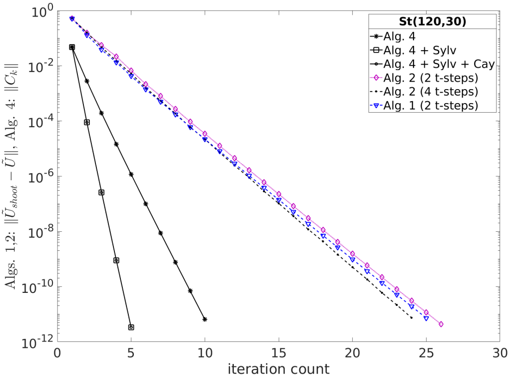
- As a rule: manifold algorithms rely on local linearizations.
  For example: shooting methods to compute Stiefel logarithm [Z. and Hüper, 2022]:
  1 step Euclidean case $\leftrightarrow$ iteration of steps on manifold



dist $= 0.95\pi$

Curvature has an impact on the iteration count:

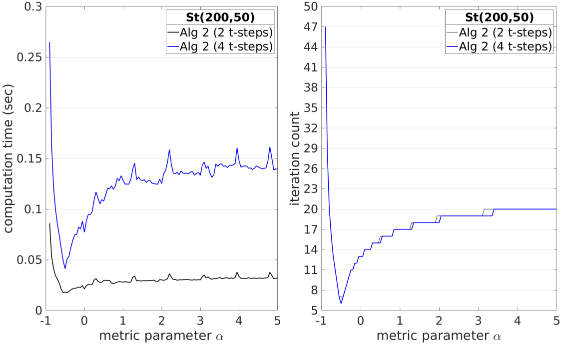
- As a rule: manifold algorithms rely on local linearizations.
  For example: shooting methods to compute Stiefel logarithm [Z. and Hüper, 2022]:
  1 step Euclidean case $\leftrightarrow$ iteration of steps on manifold



$\text{dist} = 1.0\pi$

Curvature has an impact on the iteration count:

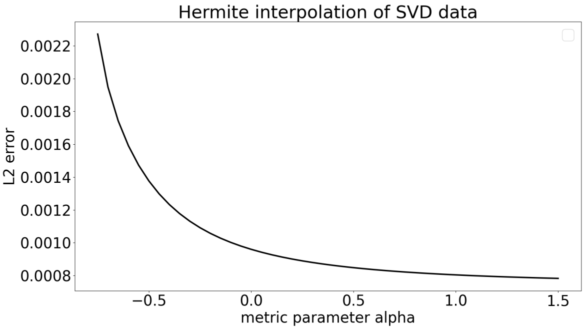
- As a rule: manifold algorithms rely on local linearizations.
  For example: shooting methods to compute Stiefel logarithm [Z. and Hüper, 2022]:
  1 step Euclidean case $\leftrightarrow$ iteration of steps on manifold



Euclid. metric: $\alpha = -\frac{1}{2}$

Curvature has an impact on the iteration count:

- As a rule: manifold algorithms rely on local linearizations.
  For example: shooting methods to compute Stiefel logarithm [Z. and Hüper, 2022]:
  1 step Euclidean case $\leftrightarrow$ iteration of steps on manifold



Euclid. metric: $\alpha = -\frac{1}{2}$

Canonical Stiefel log computations [Z., 2017]

Solving the geodesic endpoint problem for $U, \tilde{U}$ on $St(n,p)$ boils down to a nonlinear matrix equation

$$0 = \begin{pmatrix} 0 & I_p \end{pmatrix} \log_m \left(\begin{pmatrix} M & X_0 \\ N & Y_0 \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & \Phi \end{pmatrix} \right) \begin{pmatrix} 0 \\ I_p \end{pmatrix}, \quad \Phi \in SO(p). \tag{1}$$

The blocks M, N and, in turn X_0, Y_0 are computed from the input data $U, \tilde{U} \in St(n,p)$. The unknown is Φ.

Writing $\log_m \left(\begin{pmatrix} M & X_0 \\ N & Y_0 \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & \Phi \end{pmatrix} \right) = \begin{pmatrix} A & -B^T \\ B & C \end{pmatrix} \in \text{skew}(2p)$,
this means finding an orthogonal Φ such that $C = 0$.

Intuition: Need to find a rotation Φ such that the tangent vector becomes horizontal!

Canonical Stiefel log computations [Z., 2017]
Algorithm based on Baker-Campell-Hausdorff formula (BCH, Dynkin)

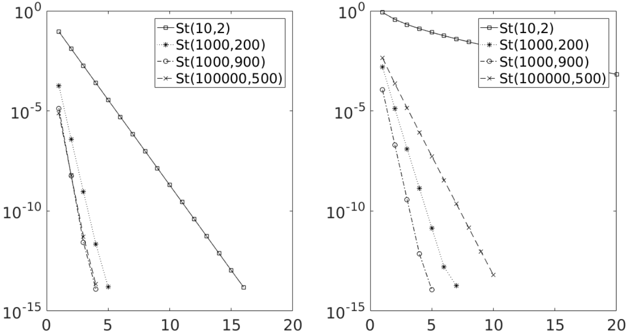
$$V_0 := \begin{pmatrix} M & X_0 \\ N & Y_0 \end{pmatrix}, \quad \log_m(V_0) := \begin{pmatrix} A_0 & -B_0^T \\ B_0 & C_0 \end{pmatrix},$$
$$W_0 := \begin{pmatrix} I_p & 0 \\ 0 & \Phi_0 \end{pmatrix}, \quad \log_m(W_0) = \begin{pmatrix} 0 & 0 \\ 0 & \log_m(\Phi_0) \end{pmatrix}.$$

BCH: $\log_m(V_0 W_0) \approx \log_m(V_0) + \log_m(W_0)$.
Geometric interpretation:

$$\begin{aligned} \log_m(V_0 W_0) = \log_m(V_0) + \log_m(W_0) \quad &\Leftrightarrow \\ V_0 W_0 = W_0 V_0 \quad &\Leftrightarrow \quad [V_0, W_0] = 0 \end{aligned}$$

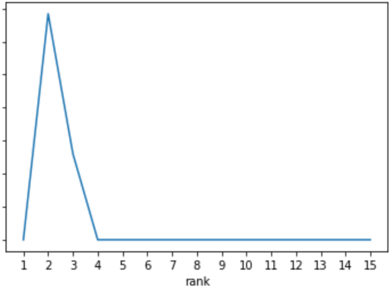
$\Leftrightarrow$ zero sectional curvature of plane spanned by V_0, W_0

Canonical Stiefel log computations [Z., 2017]



Smallest dimension → largest iteration count **and** largest error!

Explanation: For Stiefel (and Grassmann) the maximal sectional curvature is attained for tangent planes spanned by rank-2 matrices.



Experiments with (pseudo-) random data on $St(n.p)$. Number of cut points found in the range $[0.891\pi, 0.987\pi]$ sorted rank of the velocity tangent matrix.

(taken from Master thesis project of Jakob Stoye, [Stoye, 2023])

How to get curvature information?

Enter again into play: our good old quotient construction.

How to get curvature information?

Enter again into play: our good old quotient construction.

Theorem ([Gallier and Quintance, 2020], Prop. 23.29)

Let $\mathcal{M} = G/H$ be a homogeneous space with G a connected Lie group, assume that $\mathfrak{g}$ admits an $Ad(G)$-invariant inner product $\langle\cdot,\cdot\rangle$ and let $\mathfrak{m} = \mathfrak{h}^\perp$ be the orthogonal complement of $\mathfrak{h}$ with respect to $\langle\cdot,\cdot\rangle$. ($\mathfrak{h} = T_{id}H$ is vertical space at id, $\mathfrak{m}$ is horizontal). *Then*

1. *The space G/H is reductive with respect to the decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$.*
2. *Under the G-invariant metric induced by the inner product, the homogeneous space G/H is naturally reductive.*
3. *The sectional curvature at span$\{X, Y\} \subset \mathfrak{m}$ is determined by*

$$\langle R(X, Y)X, Y\rangle = \frac{1}{4}\|[X, Y]_{\mathfrak{m}}\|^2 + \|[X, Y]_{\mathfrak{h}}\|^2. \tag{3}$$

for $X \perp Y$, $\|X\| = \|Y\| = 1$. *(The subscripts $_{\mathfrak{h},\mathfrak{m}}$ indicate projections.)*

Useful matrix inequalities for curvature estimates

For any two matrices $A, B \in \mathbb{R}^{m \times n}$, with $m, n \geq 2$,

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

[Wu and Chen, 1988]

Related: the (settled) Böttcher-Wenzel conjecture for real, square matrices

$$\|AB - BA\|_F \leq \sqrt{2}\|A\|_F\|B\|_F$$

[Böttcher and Wenzel, 2008, Vong and Jin, 2008].

Useful matrix inequalities for curvature estimates

For any two matrices $A, B \in \mathbb{R}^{m\times n}$, with $m, n \geq 2$,

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

[Wu and Chen, 1988]

Related: the (settled) Böttcher-Wenzel conjecture for real, square matrices

$$\|AB - BA\|_F \leq \sqrt{2}\|A\|_F\|B\|_F$$

[Böttcher and Wenzel, 2008, Vong and Jin, 2008].

Something along these lines must have been known to Wong [Wong, 1967, Wong, 1968], who provides sharp bounds for the sectional curvature on the Grassmann manifold.

Outline

Section 3

Optimization, interpolation, MOR

Outline

Subsection 1

Symplectic Model Order Reduction

Symplectic Model Order Reduction

[Peng and Mohseni, 2016, Afkham and Hesthaven, 2017, Buchfink et al., 2020] ...

Full order model (FOM)

Hamilton's equations

$$\begin{cases} \dot{x}(t,\mu) = J_{2n}\nabla H_\mu(x), \\ x(0,\mu) = x_0(\mu) \in \mathbb{R}^{2n}, \end{cases}$$

with states $x(t,\mu) \in \mathbb{R}^{2n}$, parameters $\mu \in \Gamma \subset \mathbb{R}^d$, and Hamiltonian $H_\mu \in C^\infty(\mathbb{R}^{2n})$.

Snapshot matrix S with column vectors $x(t_i,\mu_j)$ being samples of the full system.

Symplectic Model Order Reduction

[Peng and Mohseni, 2016, Afkham and Hesthaven, 2017, Buchfink et al., 2020] ...

Full order model (FOM)

Hamilton's equations

$$\begin{cases} \dot{x}(t,\mu) = J_{2n} \nabla H_\mu(x), \\ x(0,\mu) = x_0(\mu) \in \mathbb{R}^{2n}, \end{cases}$$

with states $x(t,\mu) \in \mathbb{R}^{2n}$, parameters $\mu \in \Gamma \subset \mathbb{R}^d$, and Hamiltonian $H_\mu \in C^\infty(\mathbb{R}^{2n})$.

Snapshot matrix S with column vectors $x(t_i, \mu_j)$ being samples of the full system.

Reduced model (ROM)

Approximation

$$\begin{cases} \dot{y}(t,\mu) = J_{2k} \nabla (H_\mu \circ U)(y), \\ y(0,\mu) = U^+ x_0(\mu) \in \mathbb{R}^{2k}, \end{cases}$$

subject to

$$\min_{U \in \mathbb{R}^{2n \times 2k}} \|S - UU^+ S\|_F$$

$$\text{where } U^T J_{2n} U = J_{2k}.$$

Assumption: $x(t,\mu) \approx Uy(t,\mu)$.

Holy grail? Proper symplectic decomposition? POD/SVD with symplectic structure?
With the help of Riemannian optimization?

Geometry of symplectic Stiefel and Grassmann:
[Bendokat and Z., 2021]

- Quotient space structure
- tangent spaces
- metrics, Riemannian/pseudo
- Riemannian exponential + retractions (Cayley)
- Riemannian gradients

Related: [Gao et al., 2021b, Gao et al., 2021a]
Can PSD be used to find a “symplectic SVD” or can a “true symplectic SVD” be used to solve PSD?

Numerical experiment: 1D parametric Schrödinger

FOM simulations: Störmer-Verlet time-stepping scheme, $h = \Delta t = 0.01$, $[t_0, t_e] = [0, 20]$.

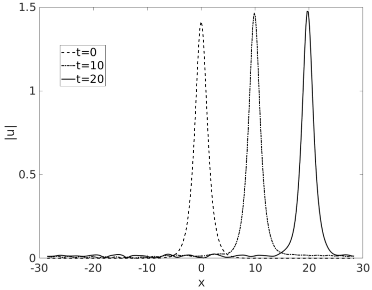


Figure 3: Probability density $|u(t,x,\epsilon)| = \sqrt{q^2(t,x,\epsilon) + p^2(t,x,\epsilon)}$ for time instants $t = 0, 10, 20$.

Take snapshots at every 10th time step. Snapshot matrix:

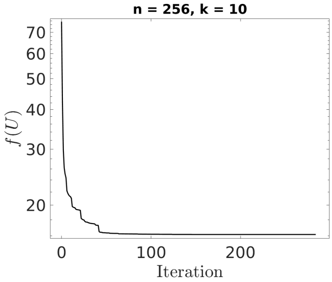
$$S = \left(\begin{pmatrix} q(t_1) \\ p(t_1) \end{pmatrix}, \dots, \begin{pmatrix} q(t_m) \\ p(t_m) \end{pmatrix} \right) \in \mathbb{R}^{512 \times 201}$$

Numerical experiment: 1D parametric Schrödinger

Reduced dimension $k = 10$.
Solve

$$\min_{UU^+ \in \mathrm{SpGr}(2n,2k)} \|S - UU^+ S\|_F,$$

relative error
$\|S - UU^+ S\|_F / \|S\|_F$.



Start	init. error	opt. error	iters.
(0) $U_0 = E$	1.0	0.067	646
(a) cotangent lift	0.261	0.067	284
(b) complex SVD	0.174	0.067	385
(c) SVD-like decomp.	0.0853	0.067	297

(a),(b):[Peng and Mohseni, 2016], (c): [Buchfink et al., 2020] relying on [Xu, 2003]

Outline

Subsection 2

Multivariate Hermite interpolation

Interpolation via optimization

The Riemannian barycenter / Fréchet mean of a sample data set $\{p_1, \dots, p_k\} \subset \mathcal{M}$ on a manifold:
Minimizer of

$$\mathcal{M} \ni q \mapsto L(q) = \frac{1}{2} \sum_{j=1}^{k} w_j \, \mathrm{dist}(q, p_j)^2,$$

where

- $\mathrm{dist}(q, p_j)$: Riemannian distance between $q, p_j \in \mathcal{M}$
- $w_j \geq 0$: scalar weights, $\sum_{j=1}^{k} w_j = 1$. (pos. measure of unit weight).

Existence and uniqueness criteria, further details: [Karcher, 1977], [Afsari et al., 2013].

Interpolation via optimization

Let $\{\varphi_j : \omega \mapsto \varphi_j(\omega) \in \mathbb{R} \mid j = 1, \dots, k\}$ be multivariate scalar-valued interpolation weight functions with $\varphi_l(\omega_j) = \delta_{lj}$ and $\sum_{j=1}^{k} \varphi_j(\omega) \equiv 1$: $\leftarrow$ signed measure of unit weight.

(constructed, e.g., from Lagrange polynomials, [Sander, 2016], radial basis functions, [Buhmann, 2003], Kriging)

Interpolation via optimization

Let $\{\varphi_j : \omega \mapsto \varphi_j(\omega) \in \mathbb{R} \mid j = 1, \dots, k\}$ be multivariate scalar-valued interpolation weight functions with $\varphi_l(\omega_j) = \delta_{lj}$ and $\sum_{j=1}^k \varphi_j(\omega) \equiv 1$: $\leftarrow$ signed measure of unit weight.

(constructed, e.g., from Lagrange polynomials, [Sander, 2016], radial basis functions, [Buhmann, 2003], Kriging)

Interpolant at ω^*: $q^* := \arg\min_{q \in \mathcal{M}} L(q, \omega^*)$, where

$$L(q, \omega) := \frac{1}{2} \sum_{j=1}^{k} \varphi_j(\omega) \operatorname{dist}(q, p_j)^2. \tag{4}$$

Precise conditions for the local existence and uniqueness under signed unit measures: [Sander, 2016, Theorems 3.1 & 3.19]. Under these conditions, the local minima are smooth in (q, ω), if the φ_j are smooth, [Sander, 2016, Theorems 3.19 & 4.1].

Interpolation via optimization

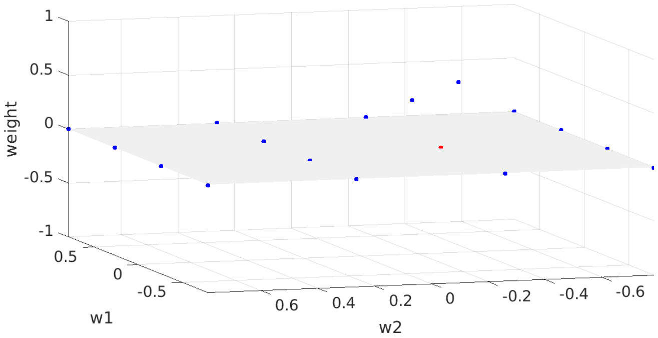


Figure 4: Barycentric interpolation: attached to each sample location (blue dots) is a weight function φ_j. The weight functions get excited depending on their distance to the trial location (red dot), the total weight always sums up to 1. Once the weights are determined, the corresponding Riemannian barycenter (aka Fréchet mean) is computed.

Barycentric Hermite Interpolation

Idea: [Z. and Bergmann, 2023], similar idea for Riem. continuation in [Séguin and Kressner, 2022]

- Local minima (= interpolants) characterized by zeros of the parametric gradient field

$$(q,\omega) \quad \mapsto \quad \operatorname{grad}_q L(q,\omega) = -\sum_{j=1}^{k} \varphi_j(\omega) \operatorname{Log}_q(p_j) \tag{5}$$

Barycentric Hermite Interpolation

Idea: [Z. and Bergmann, 2023], similar idea for Riem. continuation in [Séguin and Kressner, 2022]

- Local minima (= interpolants) characterized by zeros of the parametric gradient field

$$(q,\omega) \quad \mapsto \quad \operatorname{grad}_q L(q,\omega) = -\sum_{j=1}^{k} \varphi_j(\omega)\, \operatorname{Log}_q(p_j) \tag{5}$$

- parameterize the zero sets via the implicit function theorem

Barycentric Hermite Interpolation

Idea: [Z. and Bergmann, 2023], similar idea for Riem. continuation in [Séguin and Kressner, 2022]

- Local minima (= interpolants) characterized by zeros of the parametric gradient field

$$(q,\omega) \quad \mapsto \quad \operatorname{grad}_q L(q,\omega) = -\sum_{j=1}^{k} \varphi_j(\omega) \operatorname{Log}_q(p_j) \qquad (5)$$

- parameterize the zero sets via the implicit function theorem
- differentiate the implicit function, applied to (5) this yields

$$v_l^i = \operatorname{Hess}_q L(p_l,\omega_l)[v_l^i] = \sum_{j=1, j\neq l}^{k} \partial_i \varphi_j(\omega_l) \operatorname{Log}_{p_l}(p_j) \qquad (6)$$

- **Theorem:** For p fixed, the Hesse form of $q \mapsto \frac{1}{2}\operatorname{dist}(q,p)^2$ at p is the identity, $\operatorname{Hess}_q L(p) = \operatorname{id}_{T_p\mathcal{M}} \colon T_p\mathcal{M} \to T_p\mathcal{M}$.

Equation (6) yields a set of linear equation systems.
Write $\mathrm{Log}_{p_l}(p_j) \in T_{p_l}\mathcal{M}$ in a local frame.
Here: $\dim(\mathcal{M}) = \dim(T_{p_l}\mathcal{M}) = m$.

$$\mathrm{Log}_{p_l}(p_j) = x^j_{l,1} E^l_1 + \cdots + x^j_{l,m} E^l_m.$$

Likewise:

$$v^i_l = \alpha^i_{l,1} E^l_1 + \cdots + \alpha^i_{l,m} E^l_m.$$

Equation system for derivatives of coefficient functions:

$$\begin{pmatrix} x^1_{l,1} & \dots & x^{l-1}_{l,1} & x^{l+1}_{l,1} & \dots & x^k_{l,1} \\ \vdots & & \vdots & \vdots & & \vdots \\ x^1_{l,m} & \dots & x^{l-1}_{l,m} & x^{l+1}_{l,m} & \dots & x^k_{l,m} \\ 1 & \dots & 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \partial_i \varphi_1(\omega_l) \\ \vdots \\ \partial_i \varphi_{l-1}(\omega_l) \\ \partial_i \varphi_{l+1}(\omega_l) \\ \vdots \\ \partial_i \varphi_k(\omega_l) \end{pmatrix} = \begin{pmatrix} \alpha^i_{l,1} \\ \vdots \\ \vdots \\ \alpha^i_{l,m} \\ 0 \end{pmatrix} := \alpha^{\mathbf{i}}_{\mathbf{l}}. \quad (7)$$

Hermite data on $SO(3)$

Academic test function:

$$f:[a,b]^2 \to SO(3), \quad (\omega_1,\omega_2) \mapsto \exp_m X(\omega_1,\omega_2), \text{ where}$$

$$X(\omega_1,\omega_2) = \begin{pmatrix} 0 & \omega_1^2 + \frac{1}{2}\omega_2 & \sin\left(4\pi(\omega_1^2+\omega_2^2)\right) \\ -\omega_1^2 - \frac{1}{2}\omega_2 & 0 & \omega_1 + \omega_2^2 \\ -\sin(4\pi(\omega_1^2+\omega_2^2)) & -\omega_1 - \omega_2^2 & 0 \end{pmatrix}.$$

The sample values $P_j = \exp_m X(\omega^j)$ at $\omega^j = (\omega_1^j, \omega_2^j)$ and the corresponding partial derivatives $V_j^i = \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} \exp_m(X(\omega^j + te_i)) = \mathrm{d}(\exp_m)(X(\omega^j))[\partial_i X(\omega^j)]$, $i=1,2$ of the test function can be obtained by Mathias' theorem, see [Higham, 2008, Thm. 3.6]:

$$\exp_m \begin{pmatrix} X(\omega^j) & \partial_i X(\omega^j) \\ 0 & X(\omega^j) \end{pmatrix} = \begin{pmatrix} \exp_m(X(\omega^j)) & \mathrm{d}(\exp_m)(X(\omega^j))[\partial_i X(\omega^j)] \\ 0 & \exp_m(X(\omega^j)) \end{pmatrix}.$$

Sampling plan: 7×7 Chebychev grid

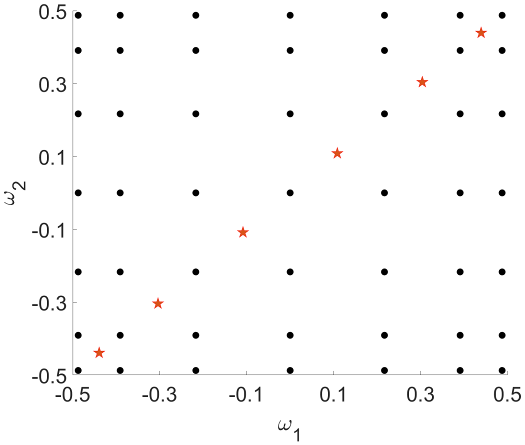


Figure 5: Black dots: Chebychev 7×7 grid on the domain $[-0.5, 0.5]^2$. Red stars: trial locations that are used for visualization purposes in the upcoming Figure 9.

Interpolation errors

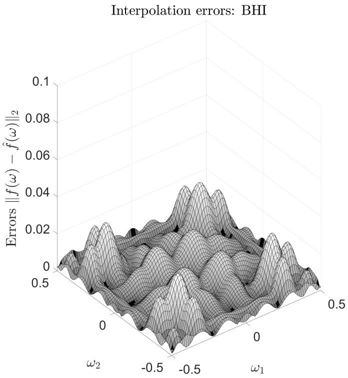


Figure 6: Error surfaces for $SO(3)$-interpolation on a Chebychev 7×7 grid. Left: Barycentric Hermite Interpolation (BHI). Right: Tangent Space Hermite Interpolation (THI).

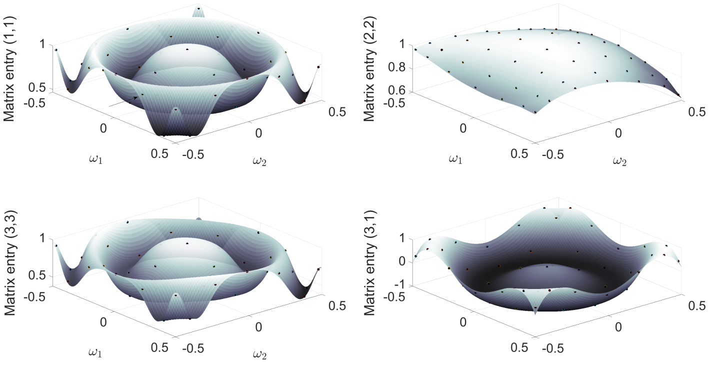


Figure 7: Plots of some selected interpolated matrix component functions $(\omega_1, \omega_2) \to \left(\hat{f}(\omega_1, \omega_2)\right)_{i,j} \in \mathbb{R}$. The black dots indicate the Chebychev 7×7 sample grid.

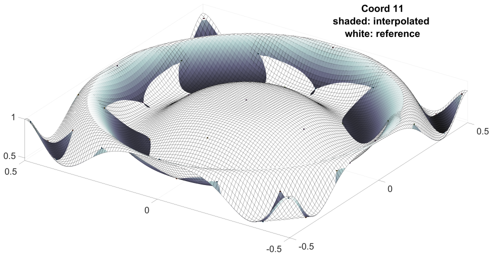


Figure 8: Interpolated matrix component function $\hat{P}_{11} = (\hat{f}(\omega))_{11}$ (shaded surface) and the reference matrix component $P_{11} = f(\omega)$ (white surface) together with the sample locations on a Chebychev 7×7 grid.

Multivariate Hermite interpolation

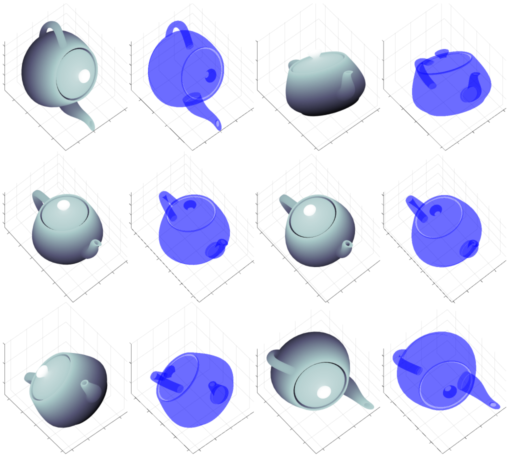


Figure 9: (from upper left to lower right): reference rotations (gray) and interpolated $SO(3)$-matrices (blue) at the 6 trial points displayed in Fig. 5. The rotation matrices are visualized via their action on the tea pot object.

Parameter settings: Interpolation on $SO(3)$			
Manifold	domain D	#samples	threshold
$SO(3)$	$[-0.5, 0.5]^2$	$k = 49$ (Cheby.)	$\tau = 1.0 \cdot 10^{-6}$

Results: barycentric Hermite interpolation (BHI)			
Wall clock time		Interpolation error	
offline	online	max	avg
0.41s	0.077s	0.029	0.0069
Results: tangent space Hermite interpolation (THI)			
Wall clock time		Interpolation error	
offline	online	max	avg
0.73s	0.0023s	0.027	0.0065

Table 1: Associated with Figure 6.
'offline': construction of the interpolant
'online': time for querying the interpolant at a trial location.
Details: [Z. and Bergmann, 2023].

Summary & Conclusion

- Riemann Exp and Log are fundamental to data processing.
  Even when you use retractions in practice, it is valuable to know the true geodesics.
- Lie groups and Lie group quotients are very well-studied objects.
  $\rightarrow$ Geodesics by geometric arguments (rather than by solving ODEs)
- Obtain geometric info from geodesic equation.
  $\rightarrow$ Covariant derivative, parallel transport, Riemannian Hessian,...
- Large (sectional) curvature spoils the performance/iteration count of geometric methods.
  For Stiefel & Grassmann: Curvature max at "rank-2 tangent planes". $\rightarrow$ Algorithms (generically) more benign in larger dims.

Summary & Conclusion

At proof-of-concept stage:

- Computing a PSD via Riemannian optimization on symplectic Stiefel for Hamiltonian MOR
- Mutlivariate Hermite interpolation
- What about really high dimensions?
- *"More sophisticated, nicer theoretical properties"* does not necessarily mean *"better results in practice"*

Summary & Conclusion

At proof-of-concept stage:

- Computing a PSD via Riemannian optimization on symplectic Stiefel for Hamiltonian MOR
- Mutlivariate Hermite interpolation
- What about really high dimensions?
- *"More sophisticated, nicer theoretical properties"* does not necessarily mean *"better results in practice"*

Open matrix issues:

- matrix exponential/general matrix functions for symplectic matrices?
- true symplectic counterpart to SVD?

The end

Thank you for your attention!

Questions?

Linear algebra
Matrix analysis
Coordinate systems
Subspaces
Norm bounds
Conditioning
Stability
...

Differential geometry
Manifolds
Geodesic paths
Normal coordinates
Christoffel symbols
Curvature
Jacobi fields
...

References I

Absil, P.-A., Mahony, R., and Sepulchre, R. (2008).
Optimization Algorithms on Matrix Manifolds.
Princeton University Press, Princeton, New Jersey.

Afkham, B. M. and Hesthaven, J. S. (2017).
Structure preserving model reduction of parametric Hamiltonian systems.
SIAM Journal on Scientific Computing, 39(6):A2616–A2644.

Afsari, B., Tron, R., and Vidal, R. (2013).
On the convergence of gradient descent for finding the Riemannian center of mass.
SIAM Journal on Control and Optimization, 51(3):2230–2260.

References II

Bendokat, T. and Z., R. (2021).
The real symplectic Stiefel and Grassmann manifolds: metrics, geodesics and applications.

Böttcher, A. and Wenzel, D. (2008).
The Frobenius norm and the commutator.
Linear Algebra Appl., 429:1864–1885.

Boumal, N. (2023).
An Introduction to Optimization on Smooth Manifolds.
Cambridge University Press, Cambridge.

References III

Bryner, D. (2017).
Endpoint geodesics on the Stiefel manifold embedded in Euclidean space.
SIAM Journal on Matrix Analysis and Applications, 38(4):1139–1159.

Buchfink, P., Haasdonk, B., and Rave, S. (2020).
PSD-greedy basis generation for structure-preserving model order reduction of Hamiltonian systems.
Proceedings of the Conference Algoritmy, pages 151–160.

Buhmann, M. D. (2003).
Radial Basis Functions, volume 12 of *Cambridge Monographs on Applied and Computational Mathematics*.
Cambridge University Press, Cambridge, UK.

References IV

do Carmo, M. P. (1992).
Riemannian Geometry.
Mathematics: Theory & Applications. Birkhäuser Boston.

Edelman, A., Arias, T. A., and Smith, S. T. (1998).
The geometry of algorithms with orthogonality constraints.
SIAM Journal on Matrix Analysis and Applications, 20(2):303–353.

Gallier, J. and Quaintance, J. (2020).
Differential Geometry and Lie Groups: A Computational Perspective.
Geometry and Computing. Springer International Publishing.

References V

Gao, B., S., N. T., Absil, P.-A., and Stykel, T. (2021a).
Geometry of the symplectic Stiefel manifold endowed with the Euclidean metric.
In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, pages 789–796, Cham. Springer International Publishing.

Gao, B., S., N. T., Absil, P.-A., and Stykel, T. (2021b).
Riemannian optimization on the symplectic Stiefel manifold.
SIAM Journal on Optimization, 31(2):1546–1575.

Higham, N. J. (2008).
Functions of Matrices: Theory and Computation.
Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

References VI

Karcher, H. (1977).
Riemannian center of mass and mollifier smoothing.
Communications on Pure and Applied Mathematics, 30(5):509–541.

Lee, J. M. (2012).
Introduction to Smooth Manifolds.
Graduate Texts in Mathematics. Springer New York.

Lee, J. M. (2018).
Introduction to Riemannian Manifolds.
Graduate Texts in Mathematics. Springer International Publishing, Cham, 2nd edition.

References VII

Peng, L. and Mohseni, K. (2016).
Symplectic model reduction of Hamiltonian systems.
SIAM Journal on Scientific Computing, 38(1):A1–A27.

Sander, O. (2016).
Geodesic finite elements of higher order.
IMA Journal of Numerical Analysis, 36(1):238–266.

Sato, H. (2021).
Riemannian Optimization and Its Applications.
SpringerBriefs in Electrical and Computer Engineering.
Springer International Publishing.

Séguin, A. and Kressner, D. (2022).
Continuation methods for Riemannian optimization.
SIAM Journal on Optimization, 32(2):1069–1093.

References VIII

Srivastava, A. and Turaga, P. K. (2015).
Riemannian computing in computer vision.
Springer International Publishing.

Stoye, J. (2023).
On the injectivity radius of the Stiefel manifold and the algorithmic domain of convergence of the canonical Riemannian logarithm.
Master's thesis, Technical University Braunschweig.

Vong, S.-W. and Jin, X.-Q. (2008).
Proof of Böttcher and Wenzel's conjecture.
Oper. Matrices, 2:435–442.

References IX

Wong, Y.-C. (1967).
Differential geometry of Grassmann manifolds.
Proceedings of the National Academy of Sciences of the United States of America, 57:589–594.

Wong, Y.-C. (1968).
Sectional curvatures of Grassmann manifolds.
Proceedings of the National Academy of Sciences of the United States of America, 60(1):75–79.

Wu, G. L. and Chen, W. H. (1988).
A matrix inequality and its geometric applications.
Acta Math. Sinica, 31(3):348–355.

References X

Xu, H. (2003).
An SVD-like matrix decomposition and its applications.
Linear Algebra and its Applications, 368:1–24.

Z., R. (2017).
A matrix-algebraic algorithm for the Riemannian logarithm on the Stiefel manifold under the canonical metric.
SIAM Journal on Matrix Analysis and Applications, 38(2):322–342.

Z., R. (2020).
Hermite interpolation and data processing errors on Riemannian matrix manifolds.
SIAM Journal on Scientific Computing, 42(5):A2593–A2619.

References XI

Z., R. and Bergmann, R. (2023).
Multivariate Hermite interpolation of manifold-valued data.
to appear in: SIAM Journal on Scientific Computing.

Z., R. and Hüper, K. (2022).
Computing the Riemannian logarithm on the Stiefel manifold: Metrics, methods, and performance.
SIAM Journal on Matrix Analysis and Applications, 43(2):953–980.