Computing geodesics on the Stiefel manifold

Marco Sutti

Postdoctoral fellow at NCTS

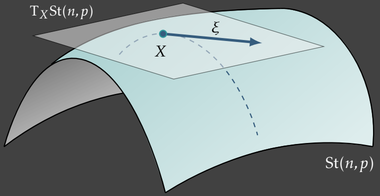
Taipei Postdoc Seminar

October 12, 2022

Overview

- Several applications in optimization, statistics, image and signal processing deal with data belonging to the Stiefel manifold

$$\mathrm{St}(n,p) = \{X \in \mathbb{R}^{n\times p} : \; X^\top X = I_p\}.$$



- Evaluation of the distance between two points on $\mathrm{St}(n,p)$.
- No closed-form solution is known for $\mathrm{St}(n,p)$!

This talk:

I. Motivating example.

II. Geometry of the Stiefel manifold.

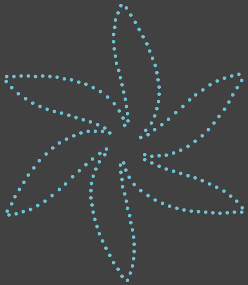
III. Computational framework based on the shooting method.

IV. Some example applications.

I. Motivation

A motivating example: imaging/1

- Need to deal with transformations that are more complicated than similarity transformations (translation/rotation/scaling).
- E.g., distortion, or imaging the same scene from different viewing angles.
- Example: two shapes from the MPEG-7 dataset, with a certain degree of similarity.



$\rightsquigarrow$ How “far” are they from each other?

MPEG-7: [Bober 2001], affine-standardized shapes: [Bryner 2017]

A motivating example: imaging/2

- One usually goes beyond the similarity group to define shape equivalences.
- Geodesics on $\mathrm{St}(n, 2)$, with shapes from the MPEG-7 dataset.

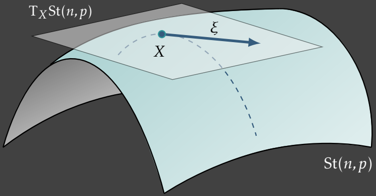
MPEG-7: [Bober 2001], affine-standardized shapes: [Bryner 2017]

II. The Stiefel manifold

The Stiefel manifold and its tangent space

- Set of matrices with orthonormal columns:

  $$\mathrm{St}(n,p) = \{X \in \mathbb{R}^{n\times p} : \; X^\top X = I_p\}.$$



- Tangent space to $\mathcal{M}$ at x: set of all tangent vectors to $\mathcal{M}$ at x, denoted $\mathrm{T}_x\mathcal{M}$. For $\mathrm{St}(n,p)$,

  $$\mathrm{T}_X\mathrm{St}(n,p) = \{\xi \in \mathbb{R}^{n\times p} : \; X^\top \xi + \xi^\top X = 0\}.$$

- Alternative characterization of $\mathrm{T}_X\mathrm{St}(n,p)$:

  $$\mathrm{T}_X\mathrm{St}(n,p) = \{X\Omega + X_\perp K : \; \Omega = -\Omega^\top, \; K \in \mathbb{R}^{(n-p)\times p}\},$$

  where $\mathrm{span}(X_\perp) = \big(\mathrm{span}(X)\big)^\perp$.

Stiefel manifold: [Stiefel, 1935]

Riemannian manifold

A manifold $\mathcal{M}$ endowed with a smoothly-varying inner product (called Riemannian metric g) is called Riemannian manifold.

$\rightsquigarrow$ A couple $(\mathcal{M}, g)$, i.e., a manifold with a Riemannian metric on it.

$\rightsquigarrow$ For the Stiefel manifold:

- Embedded metric inherited by $\mathrm{T}_X\mathrm{St}(n,p)$ from the embedding space $\mathbb{R}^{n\times p}$

$$\langle \xi, \eta \rangle = \mathrm{Tr}(\xi^\top \eta), \qquad \xi, \eta \in \mathrm{T}_X\mathrm{St}(n,p).$$

- Canonical metric by seeing $\mathrm{St}(n,p)$ as a quotient of the orthogonal group $\mathrm{O}(n)$: $\mathrm{St}(n,p) = \mathrm{O}(n)/\mathrm{O}(n-p)$

$$\langle \xi, \eta \rangle_\mathrm{c} = \mathrm{Tr}(\xi^\top (I - \tfrac{1}{2}XX^\top)\eta), \qquad \xi, \eta \in \mathrm{T}_X\mathrm{St}(n,p).$$

Metrics and geodesics on $\mathrm{St}(n, p)$

Embedded metric:

$$\langle \xi, \eta \rangle = \mathrm{Tr}(\xi^\top \eta).$$

Canonical metric:

$$\langle \xi, \eta \rangle_{\mathrm{c}} = \mathrm{Tr}(\xi^\top (I - \tfrac{1}{2} X X^\top) \eta).$$

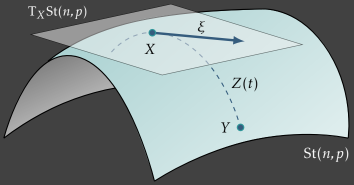
Length of a tangent vector $\xi = X\Omega + X_\perp K$:

$$\|\xi\|_{\mathrm{F}} = \sqrt{\langle \xi, \xi \rangle} = \sqrt{\|\Omega\|_{\mathrm{F}}^2 + \|K\|_{\mathrm{F}}^2}.$$

$$\|\xi\|_{\mathrm{c}} = \sqrt{\langle \xi, \xi \rangle_{\mathrm{c}}} = \sqrt{\tfrac{1}{2}\|\Omega\|_{\mathrm{F}}^2 + \|K\|_{\mathrm{F}}^2}.$$

- Closed-form solution (with the canonical metric) for a geodesic $Z(t)$ that realizes ξ with base point X:

$$Z(t) = [\,X \;\; X_\perp\,] \exp\left(\begin{bmatrix} X^\top \xi & -(X_\perp^\top \xi)^\top \\ X_\perp^\top \xi & O \end{bmatrix} t\right) \begin{bmatrix} I_p \\ O \end{bmatrix}.$$



Riemannian exponential and logarithm

- Given $x \in \mathcal{M}$ and $\xi \in T_x\mathcal{M}$, the exponential mapping $\mathrm{Exp}_x : T_x\mathcal{M} \to \mathcal{M}$ s.t. $\mathrm{Exp}_x(\xi) := \gamma(1)$, with γ being the geodesic with $\gamma(0) = x$, $\dot{\gamma}(0) = \xi$.
- Corollary: $\mathrm{Exp}_x(t\xi) := \gamma(t)$, for $t \in [0, 1]$.
- $\forall x,\ y \in \mathcal{M}$, the mapping $\mathrm{Exp}_x^{-1}(y) \in T_x\mathcal{M}$ is called logarithm mapping.

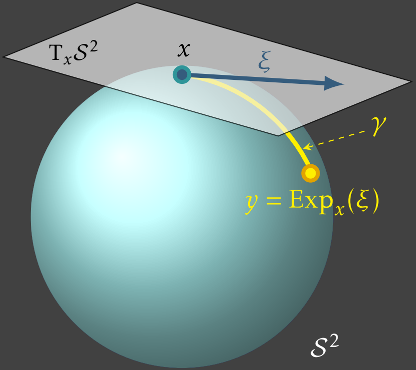
Example. Let $\mathcal{M} = \mathcal{S}^{n-1}$, then the exponential mapping at $x \in \mathcal{S}^{n-1}$ is

$$y = \mathrm{Exp}_x(\xi) = x\cos(\|\xi\|) + \frac{\xi}{\|\xi\|}\sin(\|\xi\|),$$

and the Riemannian logarithm is

$$\mathrm{Log}_x(y) = \xi = \arccos(x^\top y)\frac{\mathrm{P}_x y}{\|\mathrm{P}_x y\|},$$

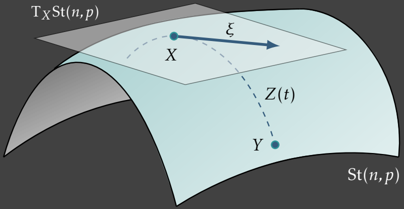
where $y \equiv \gamma(1)$ and P_x is the projector onto $\big(\mathrm{span}(x)\big)^\perp$, i.e., $\mathrm{P}_x = I - xx^\top$.



Riemannian distance on $\mathrm{St}(n,p)$

- Property: Given X, $Y \in \mathrm{St}(n,p)$, s.t. $\mathrm{Exp}_X(\xi) = Y$, the Riemannian distance $d(X,Y)$ equals the length of $\xi \equiv \dot{Z}(0) \in \mathrm{T}_X\mathrm{St}(n,p)$:

$$d(X,Y) = \|\xi\|_{\mathrm{c}} = \sqrt{\langle \xi, \xi \rangle_{\mathrm{c}}}.$$



Equivalent to: Compute the length of the Riemannian logarithm of Y with base point X, i.e.,

$$\mathrm{Log}_X(Y) = \xi.$$

- No closed-form solution is known for $\mathrm{St}(n,p)$!

$\rightsquigarrow$ How do we compute $d(X,Y)$ in practice / numerically?

III. The shooting method

Single shooting for BVPs

- Boundary value problem (BVP): Find $w(x)\colon [a,b] \to \mathbb{R}$ that satisfies

$$w'' = f(x, w, w'), \quad \text{with BCs} \quad \begin{cases} w(a) = \alpha, \\ w(b) = \beta. \end{cases}$$

- Recast it as an initial value problem (IVP): Find $w(x)$ that satisfies

$$w'' = f(x, w, w'), \quad \text{with ICs} \quad \begin{cases} w(a) = \alpha, \\ w'(a) = s. \end{cases}$$

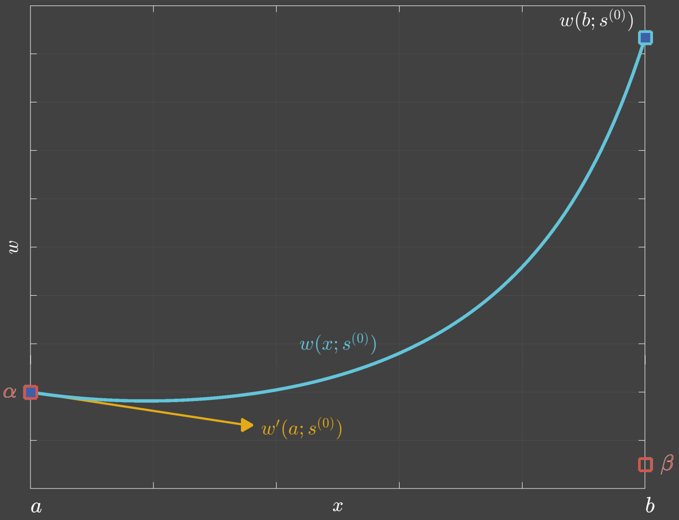
In general, this has a unique solution $w(x) \equiv w(x; s)$ which depends on s.

$\rightsquigarrow$ Single shooting method for BVPs:

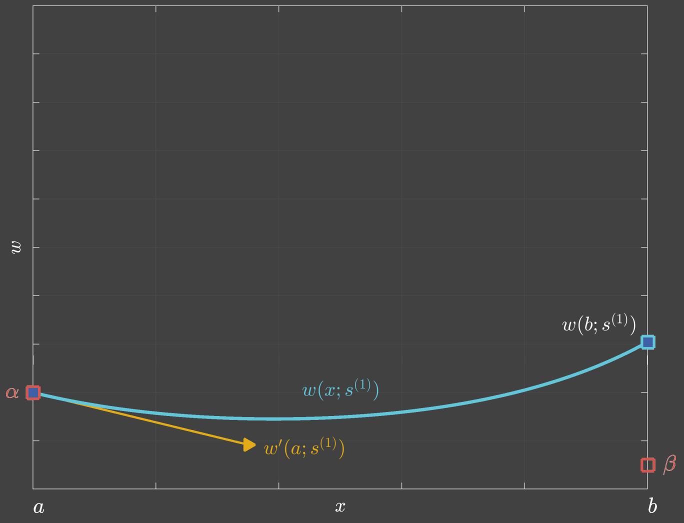
- Define $F(s) = w(b; s) - \beta$.
- Find $\bar{s}$ s.t. $F(\bar{s}) = 0$. Usually, with Newton's method.

BVPs and shooting methods: see, e.g., [Stoer/Bulirsch 1991]

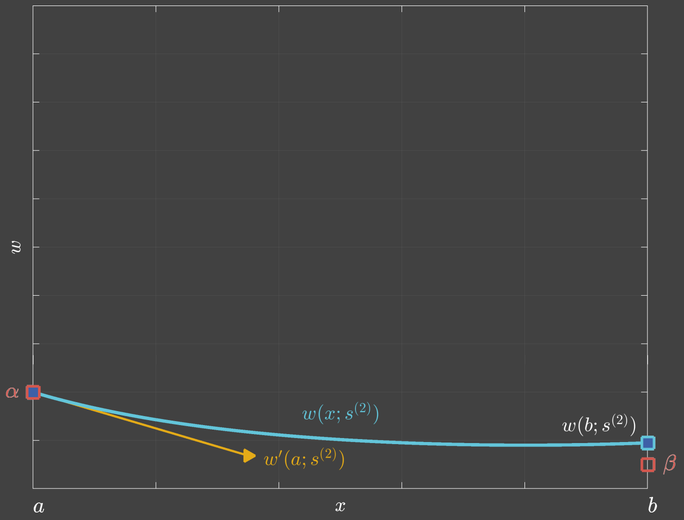
Single shooting for BVPs: example



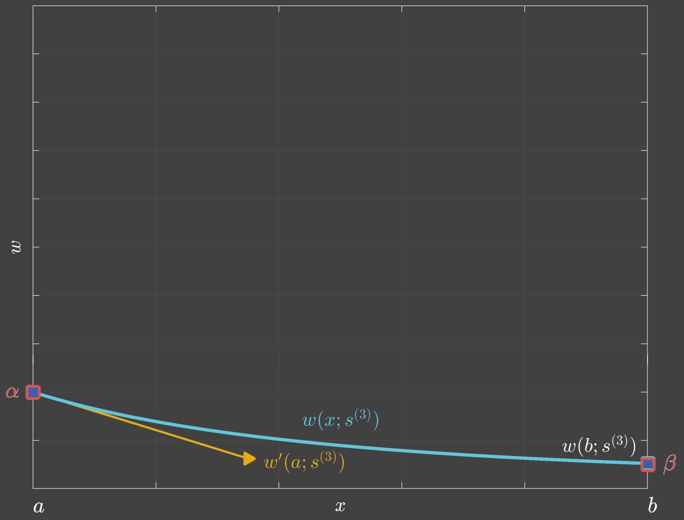
Single shooting for BVPs: example



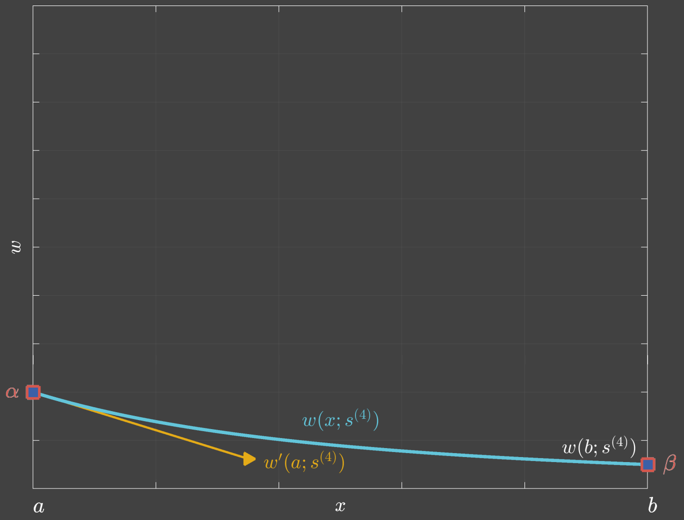
Single shooting for BVPs: example



Single shooting for BVPs: example



Single shooting for BVPs: example

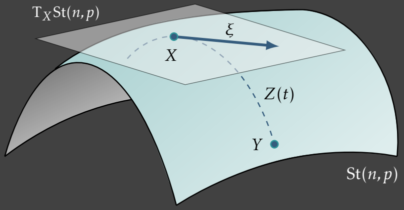


Stiefel geodesics via single shooting/1

- Problem statement:
  Find $\xi \equiv \dot{Z}(0) \in \mathrm{T}_X \mathrm{St}(n,p)$
  that satisfies the BVP

$$\ddot{Z} = -\dot{Z}\dot{Z}^\top Z - Z((Z^\top \dot{Z})^2 + \dot{Z}^\top \dot{Z}),$$

$$\text{with BCs} \quad \begin{cases} Z(0) = X, \\ Z(1) = Y. \end{cases}$$



- Recall: we have the explicit solution: $Z(t) = [\,X\ X_\perp\,] \exp\left(\begin{bmatrix} X^\top \xi & -(X_\perp^\top \xi)^\top \\ X_\perp^\top \xi & O \end{bmatrix} t\right) \begin{bmatrix} I_p \\ O \end{bmatrix}.$

$\rightsquigarrow$ Single shooting for Stiefel geodesics:

- Define $F(\xi) = Z_{(t=1,\xi)} - Y$.
- Find ξ s.t. $F(\xi) = 0$ with Newton's method.

Shooting method on the Stiefel manifold: [Bryner 2017, S. 2020, S./Vandereycken 2020]

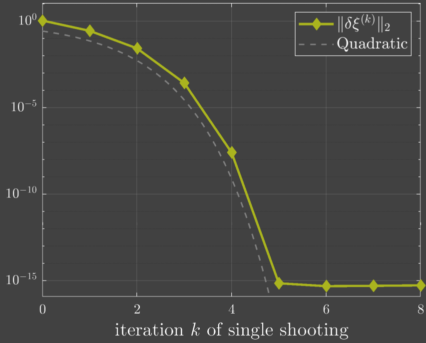
Stiefel geodesics via single shooting/2

- Numerical experiment on $\mathrm{St}(15,4)$.
- Monitored quantity: norm of the residual $\delta\xi^{(k)}$ of $F(\xi^{(k)}) = Z_{(t=1,\xi^{(k)})} - Y$.

⊕ Quadratic convergence.

⊖ A good initial guess $\xi^{(0)}$ is needed.

- Local problem (X and Y "close") can be solved very well by single shooting.
- A variable step size might be used to make the shooting more robust.



MATLAB code available: github.com/MarcoSutti/LFMS_Stiefel

IV. Applications

Model order reduction/1

- Model order reduction (MOR) for dynamical systems parametrized according to $p = [p_1, \dots, p_d]^\top$.
- For each parameter p_i in a set $\{p_1, p_2, \dots, p_K\}$, use proper orthogonal decomposition (POD) to derive a reduced-order basis $V_i \in \mathrm{St}(n, r)$, $r \ll n$.

$$\begin{cases} \dot{x}(t;p) = A(p)\,x(t;p) + B(p)\,u(t), \\ y(t;p) = C(p)\,x(t;p), \end{cases}$$

$$x(t;p) \in \mathbb{R}^n,\ \ u(t) \in \mathbb{R}^m,\ y(t) \in \mathbb{R}^q,$$

$$A(p) \in \mathbb{R}^{n\times n},\ B(p) \in \mathbb{R}^{n\times m},\ C(p) \in \mathbb{R}^{q\times n}.$$

$$\xrightarrow{\text{reduction}}$$

$$\begin{cases} \dot{x}_r(t;p) = A_r(p)\,x_r(t;p) + B_r(p)\,u(t), \\ y_r(t;p) = C_r(p)\,x_r(t;p), \end{cases}$$

$$x_r = V^\top x,\ A_r = V^\top A V,\ B_r = V^\top B,$$

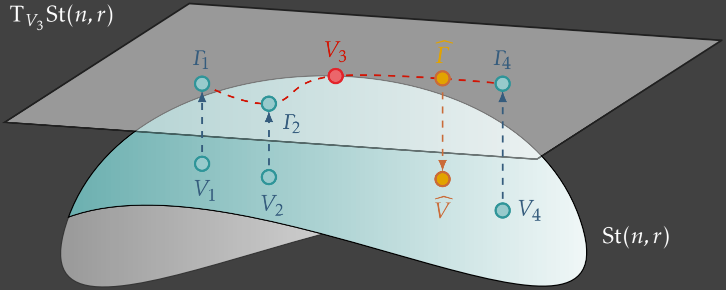
$$C_r = CV,\ \ V \equiv V(p) \in \mathrm{St}(n, r),\ r \ll n.$$

$\rightsquigarrow$ This gives a set of local basis matrices $\{V_1, V_2, \dots, V_K\}$.

MOR, POD: [Benner/Gugercin/Willcox 2015]

Model order reduction/2

- Given a new parameter value $\hat{p}$, a basis $\widehat{V}$ can be obtained by interpolating the local basis matrices on a tangent space to $\mathrm{St}(n, r)$.
- For interpolation on $\mathrm{T}_{V_3}\mathrm{St}(n, r)$, the distance is needed.

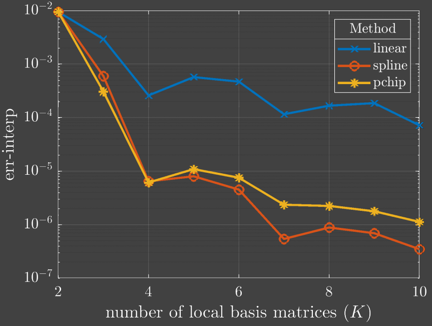


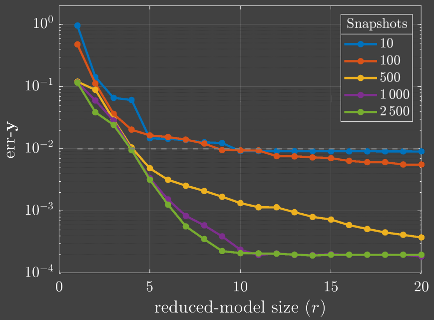
Interpolation in the tangent space to a manifold: [Hüper/Silva Leite 2007, Amsallem 2010, Amsallem/Farhat 2011]

Model order reduction/3

Transient heat equation on a square domain, with 4 disjoint discs.

- FEM discretization with $n = 1169$. Simulation for $t \in [0, 500]$, with $\Delta t = 0.1$.
- 500 snapshot POD over 5000 timeframes, with a reduced model of size $r = 4$.
- Relative error between $y(\cdot;\hat{p})$ and $y_r(\cdot;\hat{p})$ is about 1%.





Details for these experiments: [S. 2020]

Riemannian center of mass

- Notion of mean on a Riemannian manifold $\mathcal{M}$, defined by the optimization problem

$$\mu = \operatorname*{argmin}_{p \in \mathcal{M}} \frac{1}{2N} \sum_{i=1}^{N} d^2(p, q_i),$$

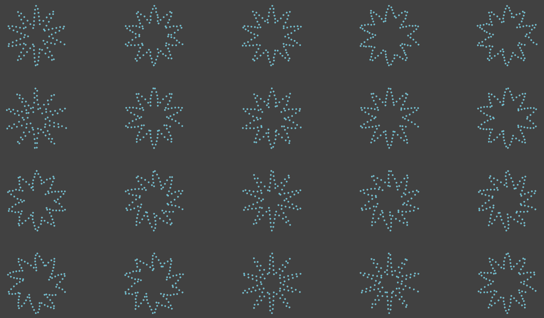
where $d(p, q_i)$ is the Riemannian distance on $\mathcal{M}$, and $q_i \in \mathcal{M}$, for $i = 1, \ldots, N$.

- For $\mathrm{St}(n, p)$, the distances $d(p, q_i)$ are computed with our algorithm.

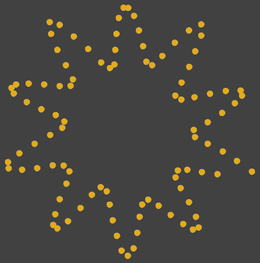
Riemannian center of mass: [Cartan 1920s, Calabi 1958, Grove/Karcher 1973]

Riemannian center of mass of a shape set

- "device7" shape set from the MPEG-7 dataset.



- Riemannian center of mass:



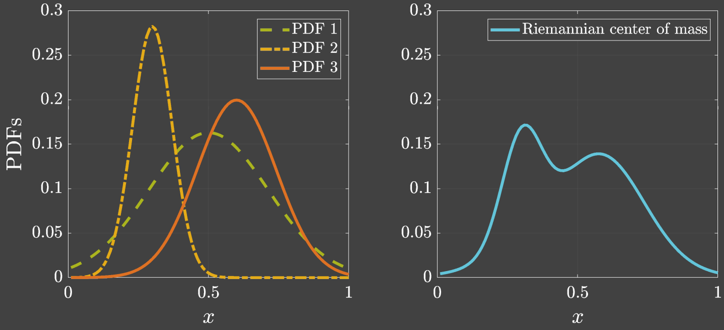
MPEG-7: [Bober 2001], affine-standardized shapes: [Bryner 2017]

Riemannian center of mass for summary statistics

Summary Statistics: $\mathcal{S}^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$ can be used to approximate $\mathcal{S}^\infty$, the space of univariate probability density functions (PDFs) on the unit interval $[0,1]$, i.e.,

$$\mathcal{P} = \left\{ g\colon [0,1] \to \mathbb{R}_{\geqslant 0} :\ \int_0^1 g(x)\,\mathrm{d}x = 1 \right\}.$$

Example: Riemannian center of mass of 3 PDFs, sampled at 100 points, thus making them belong to $\mathrm{St}(100,1) \equiv \mathcal{S}^{99}$.

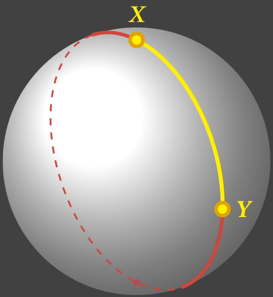


Functional and shape data analysis: [Srivastava/Klassen 2016]

Conclusions

This talk:

- Computing the Riemannian distance can be a hard problem.
- Computational framework: shooting method.
- Applications in imaging, model order reduction, and summary statistics.



Outlook:

- Recent advances in numerical algorithms: [Zimmermann 2017, Zimmermann/Hüper 2022].
- Other novel applications on $\mathrm{St}(n,p)$ for: EEG data [Yamamoto et al. 2021], brain network harmonics [Chen et al. 2021], clustering problems [Huang et al. 2022], ...

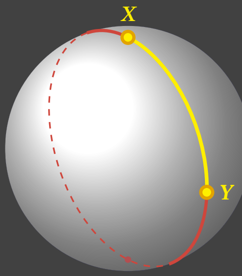
$\rightsquigarrow$ MATLAB code: github.com/MarcoSutti/LFMS_Stiefel

$\rightsquigarrow$ Download slides: marcosutti.net/research.html#talks

V. Bonus material

Geodesics

- Generalization of straight lines to manifolds.
- Locally curves of shortest length, but globally they may not be.



- Hopf–Rinow theorem guarantees the existence of a length-minimizing geodesic connecting any two given points.

Hopf–Rinow Theorem

Theorem ([Hopf/Rinow]) Let $(\mathcal{M}, g)$ be a (connected) Riemannian manifold. Then the following conditions are equivalent:

1. Closed and bounded subsets of $\mathcal{M}$ are compact;
2. $(\mathcal{M}, g)$ is a complete metric space;
3. $(\mathcal{M}, g)$ is geodesically complete, i.e., for any $x \in \mathcal{M}$, the exponential map Exp_x is defined on the entire tangent space $\mathrm{T}_x\mathcal{M}$.

Any of the above implies that given any two points $x, y \in \mathcal{M}$, there exists a length-minimizing geodesic connecting these two points.

The Stiefel manifold is compact/complete/geodesically complete.

$\rightsquigarrow$ Length-minimizing geodesics exist.

Riemannian Geometry, Sakai 1992

The orthogonal group as a special case of $\mathrm{St}(n, p)$

- If $p = n$, then the Stiefel manifold reduces to the orthogonal group

$$\mathrm{O}(n) = \{X \in \mathbb{R}^{n \times n} \colon X^\top X = I_n\},$$

and the tangent space at X is given by

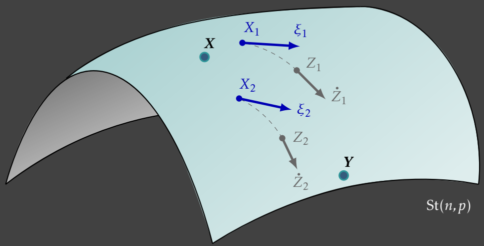
$$\mathrm{T}_X \mathrm{O}(n) = \{X\Omega \colon \Omega^\top = -\Omega\} = X\mathcal{S}_{\mathrm{skew}}(n).$$

- Furthermore, if $X = I_n$, we have $\mathrm{T}_{I_n}\mathrm{O}(n) = \mathcal{S}_{\mathrm{skew}}(n)$. This means that the tangent space to $\mathrm{O}(n)$ at the identity matrix I_n is the set of skew-symmetric n-by-n matrices $\mathcal{S}_{\mathrm{skew}}(n)$.
- In the language of Lie groups, we say that $\mathcal{S}_{\mathrm{skew}}(n)$ is the Lie algebra of the Lie group $\mathrm{O}(n)$.

Geodesics via multiple shooting

Global problem (X and Y “far”)

- Based on subdivision.
- Enforce continuity conditions of Z and $\dot{Z}$ at the interfaces between subintervals.



X_k: point on $\mathrm{St}(n, p)$ relative to the k-th subinterval.

ξ_k: tangent vector to $\mathrm{St}(n, p)$ at X_k.

Geodesics via multiple shooting

System of nonlinear equations:

$$F(\Sigma) = \begin{bmatrix} Z_1^{(1)} - \Sigma_1^{(2)} \\ Z_2^{(1)} - \Sigma_2^{(2)} \\ Z_1^{(2)} - \Sigma_1^{(3)} \\ Z_2^{(2)} - \Sigma_2^{(3)} \\ \vdots \\ r_1 = \Sigma_1^{(1)} - Y_0 \\ r_2 = \Sigma_1^{(m)} - Y_1 \end{bmatrix} = 0, \quad \xrightarrow{\text{linearize}} \quad \underbrace{\begin{bmatrix} G^{(1)} & -I & O & & O \\ O & G^{(2)} & -I & \ddots & \\ & \ddots & \ddots & \ddots & O \\ O & & \ddots & G^{(m-1)} & -I \\ C & O & & O & D \end{bmatrix}}_{=:DF(\Sigma)} \delta\Sigma = -F(\Sigma).$$

- ⊕ Fast convergence to ξ.
- ⊖ A very good initial guess $\xi^{(0)}$ is still needed.