Schwarz methods for computing geodesics

Marco Sutti

National Center for Theoretical Sciences

Taipei, Taiwan (R.O.C.)

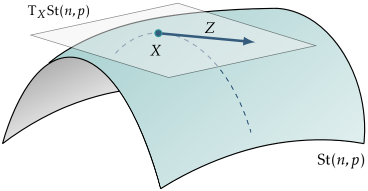
Joint work with Tommaso Vanzan

DD28, KAUST, Jeddah

January 29, 2024

Overview

- Many applications in diverse fields (such as optimization, imaging and signal processing, statistics, …) deal with data belonging to the Stiefel manifold



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

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

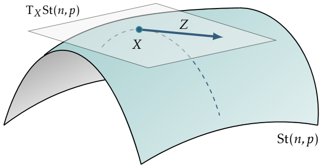
This talk:

I. Introduction to the geometry of the Stiefel manifold.

II. Description of the leapfrog algorithm for computing geodesics.

III. Leapfrog algorithm viewed as a Schwarz method. Present current progress and showcase several ideas for future research directions.

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 $\mathrm{St}(n,p)$,

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

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

- Dimension: since $\dim\big(\mathrm{St}(n,p)\big) = \dim\big(T_X\mathrm{St}(n,p)\big)$, we have

$$\dim(\mathrm{St}(n,p)) = \dim(\mathcal{S}_{\mathrm{skew}}) + \dim(\mathbb{R}^{(n-p)\times p}) = np - \tfrac{1}{2}p(p+1).$$

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 $\mathrm{O}(n)$: $\mathrm{St}(n,p) = \mathrm{O}(n)/\mathrm{O}(n-p)$

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

Optimization on matrix manifolds: [Edelman et al. 1998, Absil et al. 2008, Boumal 2023, …]

Riemannian exponential and logarithm

- Let $x \in \mathcal{M}$, $\xi \in \mathrm{T}_x\mathcal{M}$, and $\gamma(t)$ the geodesic such that $\gamma(0) = x$, $\dot{\gamma}(0) = \xi$. The exponential mapping $\mathrm{Exp}_x : \mathrm{T}_x\mathcal{M} \to \mathcal{M}$ is defined as $\mathrm{Exp}_x(\xi) := \gamma(1)$.
- Corollary: $\mathrm{Exp}_x(t\xi) := \gamma(t)$, for $t \in [0, 1]$.
- $\forall x,\ y \in \mathcal{M}$, the mapping $\mathrm{Exp}_x^{-1}(y) \in \mathrm{T}_x\mathcal{M}$ is called logarithm mapping.

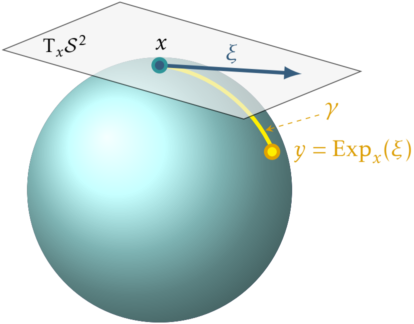
Example. Let $\mathcal{M} = \mathcal{S}^{n-1}$, then the exponential mapping at $x \in \mathcal{S}^{n-1}$ is

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

and the Riemannian logarithm is

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

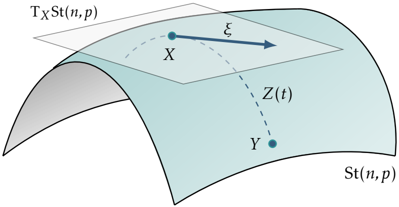
where $y \equiv \gamma(1)$ and P_x is the projector onto $\big(\mathrm{span}(x)\big)^\perp$, i.e., $\mathrm{P}_x = I - xx^\top$.



Riemannian exponential and logarithm on $\mathrm{St}(n,p)$

- Explicit expression (with the canonical metric) of the Riemannian exponential on the Stiefel manifold $\mathrm{St}(n,p)$:

$$Y = \mathrm{Exp}_X(\xi) = Z(1) = [\,X \;\; X_\perp\,] \exp\left(\begin{bmatrix} X^\top \xi & -(X_\perp^\top \xi)^\top \\ X_\perp^\top \xi & O \end{bmatrix}\right) \begin{bmatrix} I_p \\ O_{(n-p)\times p} \end{bmatrix}.$$



- There is no explicit expression for the Riemannian logarithm on the Stiefel manifold.

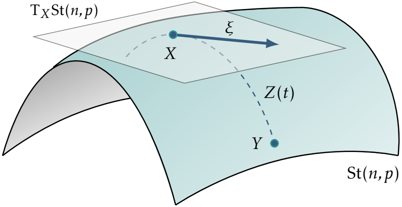
Riemannian distance

- Definition: given x, $y \in \mathcal{M}$, the Riemannian distance $d(x,y)$ is defined as

$$d(x,y) = \min_{\substack{\gamma:[0,1]\to\mathcal{M}\\ \gamma(0)=x,\,\gamma(1)=y}} L[\gamma], \quad \text{where} \quad L[\gamma] = \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t),\dot{\gamma}(t))}\,\mathrm{d}t.$$

- Property: given x, $y \in \mathcal{M}$, and $\xi \in \mathrm{T}_x\mathcal{M}$ such that $\mathrm{Exp}_x(\xi) = y$, the Riemannian distance $d(x,y)$ equals the length of $\xi \equiv \dot{\gamma}(0) \in \mathrm{T}_x\mathcal{M}$, i.e.,

$$d(x,y) = \|\xi\| = \sqrt{\langle \xi, \xi \rangle}.$$



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 $\mathrm{St}(n,p)$!

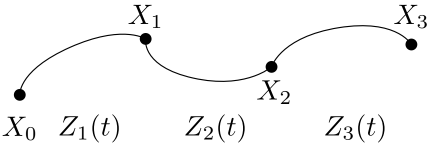
$\rightsquigarrow$ How do we compute $d(X,Y)$ in practice / numerically?

Leapfrog

- Idea: We wish to solve a "global problem" (i.e., for "big" distances between the endpoints). However, we only know how to solve local problems (i.e., on subdomains). E.g., we can compute Log with the single shooting method.

$\rightsquigarrow$ "Think globally, act locally".

- Leapfrog is based on subdivision in $m-1$ subintervals, such that a geodesic can be constructed on each subinterval. Example with $m=4$:



- It considers a piecewise geodesic which is uniquely identified by the m-tuple $\boldsymbol{X} = (X_0, X_1, \dots, X_{m-1}) \in \mathcal{M}^m$.

- By compactness, a convergent subsequence exists and its limit $\boldsymbol{X}^*$ are points that lie on a global geodesic connecting X_0 and X_{m-1}.

Leapfrog: [Noakes 1998]; Shooting methods on the Stiefel manifold: [S. 2023]

Illustration of leapfrog

Illustration of the procedure on $\mathrm{St}(3,1)$, for $m = 4$ points.

Let $\mathfrak{M}\colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ denote the midpoint map defined by
$\mathfrak{M}(U,V) = \mathrm{Exp}_U\left(\frac{1}{2}\mathrm{Log}_U(V)\right)$.

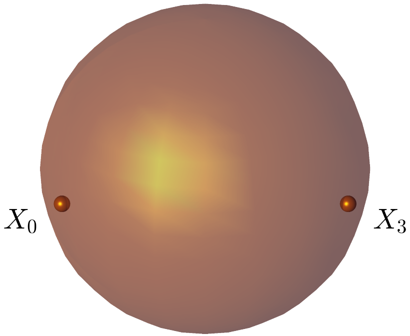


Illustration of leapfrog

Illustration of the procedure on $\mathrm{St}(3,1)$, for $m = 4$ points.

Let $\mathfrak{M}\colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ denote the midpoint map defined by $\mathfrak{M}(U,V) = \mathrm{Exp}_U\left(\frac{1}{2}\mathrm{Log}_U(V)\right)$. Then:

0^{th} iteration:

$$\boldsymbol{X}^{(0)} = (X_0, X_1^{(0)}, X_2^{(0)}, X_3).$$

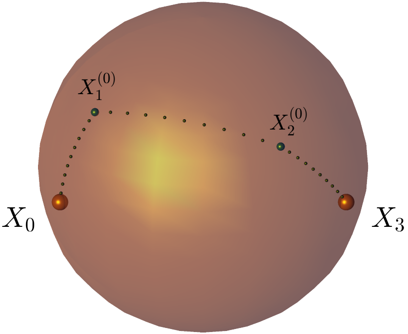


Illustration of leapfrog

Illustration of the procedure on $\mathrm{St}(3,1)$, for $m = 4$ points.

Let $\mathfrak{M}\colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ denote the midpoint map defined by $\mathfrak{M}(U,V) = \mathrm{Exp}_U\left(\frac{1}{2}\mathrm{Log}_U(V)\right)$. Then:

0^{th} iteration:

$$\boldsymbol{X}^{(0)} = (X_0, X_1^{(0)}, X_2^{(0)}, X_3).$$

1^{st} iteration:

$$X_1^{(1)} = \mathfrak{M}(X_0, X_2^{(0)}), \quad X_2^{(1)} = \mathfrak{M}(X_1^{(1)}, X_3).$$

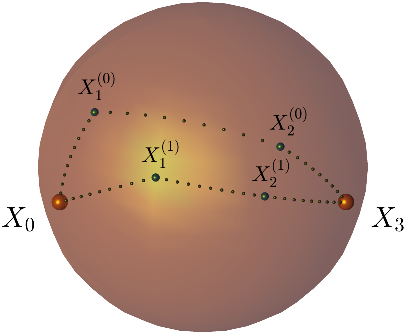


Illustration of leapfrog

Illustration of the procedure on $\mathrm{St}(3,1)$, for $m = 4$ points.

Let $\mathfrak{M}\colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ denote the midpoint map defined by $\mathfrak{M}(U,V) = \mathrm{Exp}_U\left(\frac{1}{2}\mathrm{Log}_U(V)\right)$. Then:

0^{th} iteration:

$$\boldsymbol{X}^{(0)} = (X_0, X_1^{(0)}, X_2^{(0)}, X_3).$$

1^{st} iteration:

$$X_1^{(1)} = \mathfrak{M}(X_0, X_2^{(0)}), \quad X_2^{(1)} = \mathfrak{M}(X_1^{(1)}, X_3).$$

2^{nd} iteration:

$$X_1^{(2)} = \mathfrak{M}(X_0, X_2^{(1)}), \quad X_2^{(2)} = \mathfrak{M}(X_1^{(2)}, X_3).$$

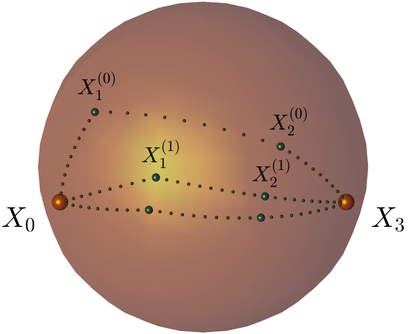


Illustration of leapfrog

Illustration of the procedure on $\mathrm{St}(3,1)$, for $m = 4$ points.

Let $\mathfrak{M}\colon \mathcal{M}\times\mathcal{M}\to\mathcal{M}$ denote the midpoint map defined by $\mathfrak{M}(U,V) = \mathrm{Exp}_U\left(\frac{1}{2}\mathrm{Log}_U(V)\right)$. Then:

0th iteration:

$$\boldsymbol{X}^{(0)} = (X_0, X_1^{(0)}, X_2^{(0)}, X_3).$$

1st iteration:

$$X_1^{(1)} = \mathfrak{M}(X_0, X_2^{(0)}), \quad X_2^{(1)} = \mathfrak{M}(X_1^{(1)}, X_3).$$

2nd iteration:

$$X_1^{(2)} = \mathfrak{M}(X_0, X_2^{(1)}), \quad X_2^{(2)} = \mathfrak{M}(X_1^{(2)}, X_3).$$

……

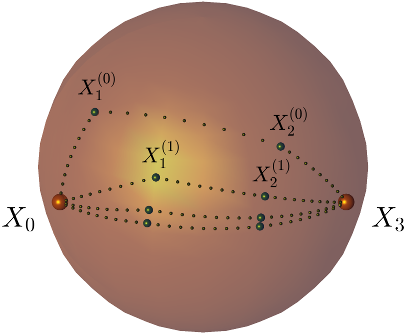


Illustration of leapfrog

Illustration of the procedure on $\mathrm{St}(3,1)$, for $m = 4$ points.

Let $\mathfrak{M}\colon \mathcal{M}\times\mathcal{M} \to \mathcal{M}$ denote the midpoint map defined by $\mathfrak{M}(U,V) = \mathrm{Exp}_U\left(\frac{1}{2}\mathrm{Log}_U(V)\right)$. Then:

0^{th} iteration:

$$\boldsymbol{X}^{(0)} = (X_0, X_1^{(0)}, X_2^{(0)}, X_3).$$

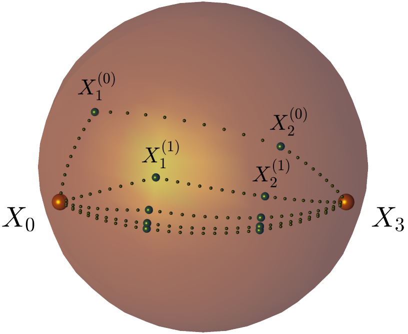
1^{st} iteration:

$$X_1^{(1)} = \mathfrak{M}(X_0, X_2^{(0)}), \quad X_2^{(1)} = \mathfrak{M}(X_1^{(1)}, X_3).$$

2^{nd} iteration:

$$X_1^{(2)} = \mathfrak{M}(X_0, X_2^{(1)}), \quad X_2^{(2)} = \mathfrak{M}(X_1^{(2)}, X_3).$$

……



$\rightsquigarrow$ It is like a Schwarz method! Original observation by Martin J. Gander.

Leapfrog/Schwarz method

Algorithm 1: Overlapping Schwarz method for computing geodesics.

Data: Given two points X_0, X_{m-1}, number of points m.
Result: Geodesic connecting X_0 and X_{m-1}.

```
Compute the initial guess for the intermediate points;
k = 0;
while a stopping criterion is met do
    for i = 1 : m − 2 do
        Compute the midpoint map X_i^(k+1) = 𝔐(X_{i−1}^(k+1), X_{i+1}^(k));
    end for
    Update k ← k + 1;
end while
```

Caveat: It has a sequential nature and converges very slowly.

Preconditioning leapfrog/1

Let's write the idea for the case of the Stiefel manifold $\mathrm{St}(n, p)$ for three subintervals ($m = 4$). Let

$$\boldsymbol{X} = \left(\mathrm{vec}(X_0)^\top,\ \mathrm{vec}(X_1)^\top,\ \mathrm{vec}(X_2)^\top,\ \mathrm{vec}(X_3)^\top\right)^\top,$$

where X_0 and X_3 are the given endpoints, and X_1 and X_2 are our unknowns. For convenience, we define $\boldsymbol{X}_{\mathrm{int}} := (\mathrm{vec}(X_1)^\top,\ \mathrm{vec}(X_2)^\top)^\top$. The **iterative method** is

$$\boldsymbol{X}_{\mathrm{int}}^{k+1} = \Phi(\boldsymbol{X}_{\mathrm{int}}^{k}),$$

where

$$\begin{pmatrix} \mathrm{vec}(X_1) \\ \mathrm{vec}(X_2) \end{pmatrix}^{k+1} = \begin{pmatrix} \mathrm{vec}\Big(\mathfrak{M}(X_0, X_2^k)\Big) \\ \mathrm{vec}\Big(\mathfrak{M}(X_1^k, X_3)\Big) \end{pmatrix}, \qquad \Phi(\boldsymbol{X}_{\mathrm{int}}^{k}) := \begin{pmatrix} \mathrm{vec}\Big(\mathfrak{M}(X_0, X_2^k)\Big) \\ \mathrm{vec}\Big(\mathfrak{M}(X_1^k, X_3)\Big) \end{pmatrix}.$$

In the leapfrog method, the function $\mathfrak{M}$ is the midpoint map defined by

$$\mathfrak{M}(U, V) := \mathrm{Exp}_U\Big(\tfrac{1}{2}\mathrm{Log}_U(V)\Big).$$

Preconditioning leapfrog/2

Hence, we can write the iterative method as

$$\begin{pmatrix} \mathrm{vec}(X_1) \\ \mathrm{vec}(X_2) \end{pmatrix}^{k+1} = \begin{pmatrix} \mathrm{vec}(\mathfrak{M}(X_0, X_2)) \\ \mathrm{vec}(\mathfrak{M}(X_1, X_3)) \end{pmatrix}.$$

Now take the fixed point $\boldsymbol{X}_{\mathrm{int}}^{k+1} = \Phi(\boldsymbol{X}_{\mathrm{int}}^{k})$ for $n \to \infty$, define the function

$$F(\boldsymbol{X}_{\mathrm{int}}) := \boldsymbol{X}_{\mathrm{int}}^{*} - \Phi(\boldsymbol{X}_{\mathrm{int}}^{*}),$$

and apply Newton's method to find the roots $\boldsymbol{X}_{\mathrm{int}}^{*}$ of this nonlinear equation. The Jacobian of F is given by

$$J_{F(\boldsymbol{X}_{\mathrm{int}})} = I - J_{\Phi(\boldsymbol{X}_{\mathrm{int}})},$$

where

$$J_{\Phi(\boldsymbol{X}_{\mathrm{int}})} = \begin{bmatrix} \dfrac{\partial \mathfrak{M}(X_0, X_2)}{\partial X_1} & \dfrac{\partial \mathfrak{M}(X_0, X_2)}{\partial X_2} \\ \dfrac{\partial \mathfrak{M}(X_1, X_3)}{\partial X_1} & \dfrac{\partial \mathfrak{M}(X_1, X_3)}{\partial X_2} \end{bmatrix} = \begin{bmatrix} 0 & \dfrac{\partial \mathfrak{M}(X_0, X_2)}{\partial X_2} \\ \dfrac{\partial \mathfrak{M}(X_1, X_3)}{\partial X_1} & 0 \end{bmatrix}.$$

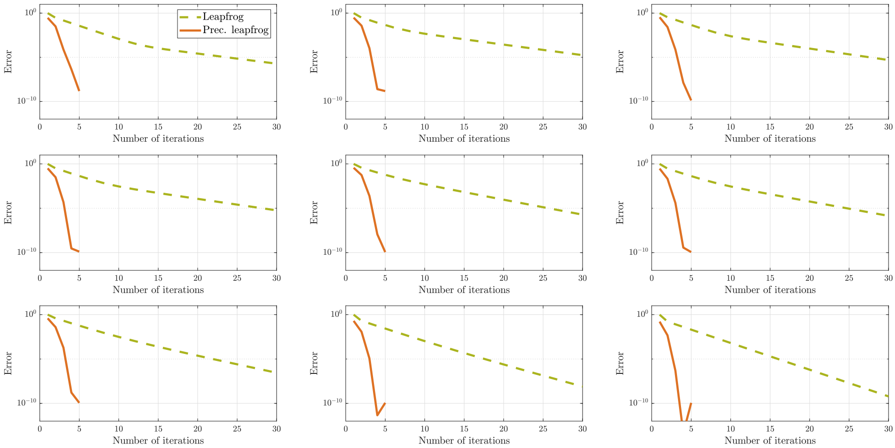
Preconditioning leapfrog/3

Hence $J_{F(\boldsymbol{X}_{\text{int}})}$ has a **block tridiagonal structure**, i.e.,

$$\begin{bmatrix} I_{np} & -\frac{\partial \mathfrak{M}(X_0, X_2)}{\partial X_2} & & & \\ -\frac{\partial \mathfrak{M}(X_1, X_3)}{\partial X_1} & I_{np} & -\frac{\partial \mathfrak{M}(X_1, X_3)}{\partial X_3} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{\partial \mathfrak{M}(X_{m-4}, X_{m-2})}{\partial X_{m-4}} & I_{np} & -\frac{\partial \mathfrak{M}(X_{m-4}, X_{m-2})}{\partial X_{m-2}} \\ & & & -\frac{\partial \mathfrak{M}(X_{m-3}, X_{m-1})}{\partial X_{m-3}} & I_{np} \end{bmatrix}.$$

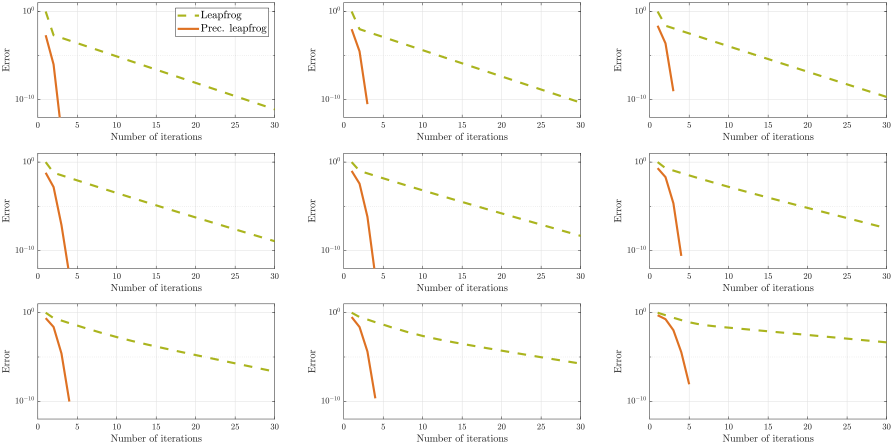
Numerical experiments/1

St$(10, p)$, varying $p = 2 : 1 : 10$, $d(X, Y) = 0.8\pi$, $m = 4$.



Numerical experiments/2

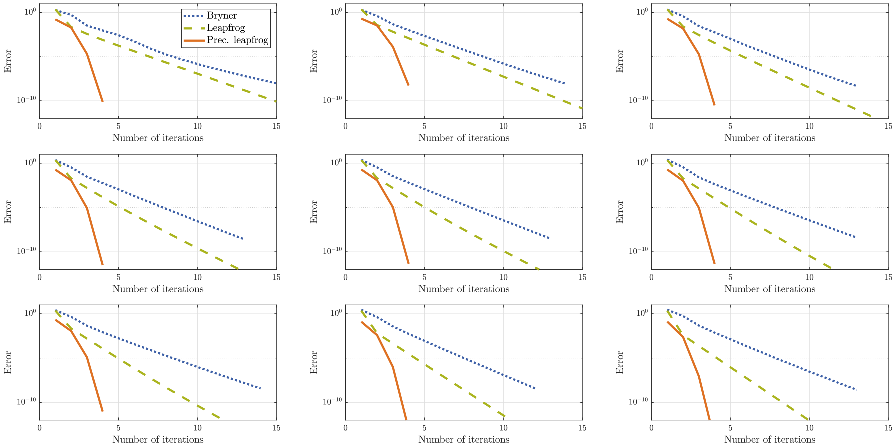
St$(10, 4)$, varying distance $d(X, Y) = 0.1\pi : 0.1\pi : 0.9\pi$, $m = 4$.



Numerical experiments/3

Comparison with “shooting” method of Bryner, 2017.

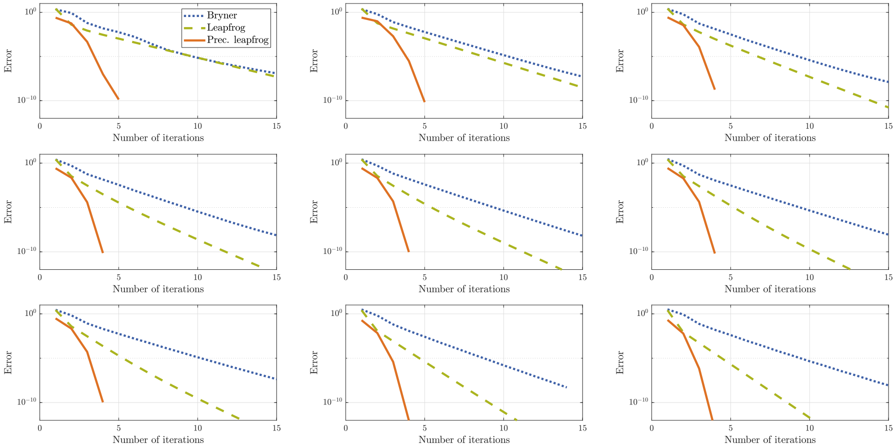
$\mathrm{St}(10, p)$, varying $p = 2:1:10$, $d(X, Y) = 0.7\pi$, $m = 4$.



Numerical experiments/4

Comparison with “shooting” method of Bryner, 2017.

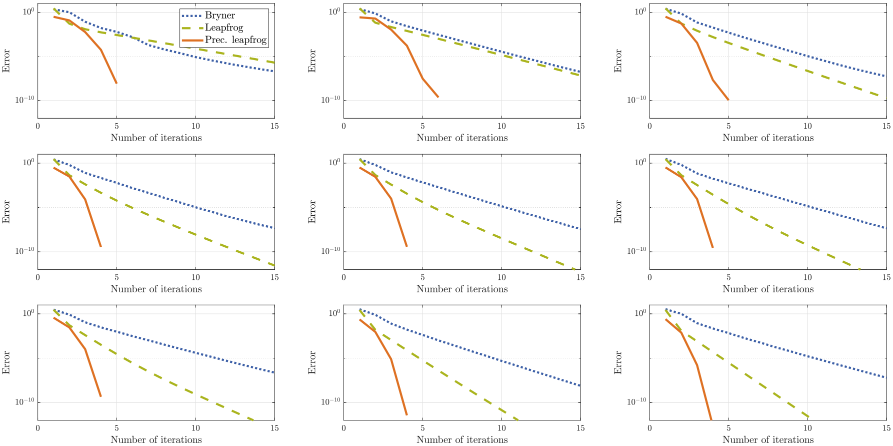
$\mathrm{St}(10, p)$, varying $p = 2:1:10$, $d(X, Y) = 0.8\pi$, $m = 4$.



Numerical experiments/5

Comparison with “shooting” method of Bryner, 2017.

$\mathrm{St}(10, p)$, varying $p = 2 : 1 : 10$, $d(X, Y) = 0.85\pi$, $m = 4$.



Some observations

- A major disadvantage of the leapfrog (Schwarz) method is its sequential nature.
- Convergence deteriorates as the number of subdomains (m) increases.
- The preconditioned leapfrog addresses this problem, but it is very expensive to form the Jacobian matrix.
- For larger $d(X, Y)$, leapfrog needs many more iterations to converge, but the number of iterations in the preconditioned version is independent of $d(X, Y)$.
- For $p \to n$, leapfrog needs fewer iterations to converge; but the convergence behavior of its preconditioned version seems to be independent of p.
- While Bryner's "shooting" method is computationally cheaper, it takes many more iterations to reach the same accuracy of both leapfrog and preconditioned leapfrog, especially for large $d(X, Y)$.

Summary and research outlook

This talk:

- Introduction to the geometry of the Stiefel manifold.
- Leapfrog, an existing method for computing geodesic, is a Schwarz method.
- Use ideas from DDM field to improve on that (work in progress).

Open questions and outlook:

- Convergence deterioration of leapfrog $\rightsquigarrow$ Introduce a coarse-grid correction like in the multigrid method. $\rightsquigarrow$ "Two-level leapfrog method"?
- Computational cost of preconditioned leapfrog: Is it possible to reduce the cost of forming the Jacobian matrix in preconditioned leapfrog?
- Use the method of Bryner within leapfrog to "conquer" the subproblems.
- Explore parallelization: simultaneously process subdomains with no overlap?

$\rightsquigarrow$ Download slides:

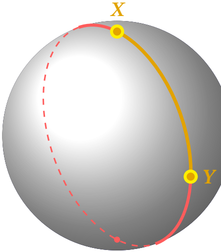
marcosutti.net/research.html#talks

Thank you for your attention!

Bonus material

Geodesics

- Generalization of straight lines to manifolds.
- Locally they are curves of shortest length, but globally they may not be.
- In general, they are defined as critical points of the length functional $L[\gamma]$, and may or may not be minima.



- The fundamental 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 $\mathrm{T}_x\mathcal{M}$.

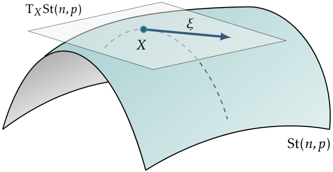
Any of the above implies that given any two points $x, y \in \mathcal{M}$, there exists a length-minimizing geodesic connecting these two points.

The Stiefel manifold is compact/complete/geodesically complete.

$\rightsquigarrow$ Length-minimizing geodesics exist.

Riemannian Geometry, Sakai 1992

Metrics 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 $\xi = X\Omega + X_\perp K$:

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

Example for $p = 3$: $\Omega = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$, then $\|\Omega\|_{\mathrm{F}}^2 = 2a^2 + 2b^2 + 2c^2$.

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

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

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