Connecting geodesics for the Stiefel manifold

UNIVERSITÉ DE GENÈVE

Marco Sutti and Bart Vandereycken

Section of Mathematics, University of Geneva

Overview

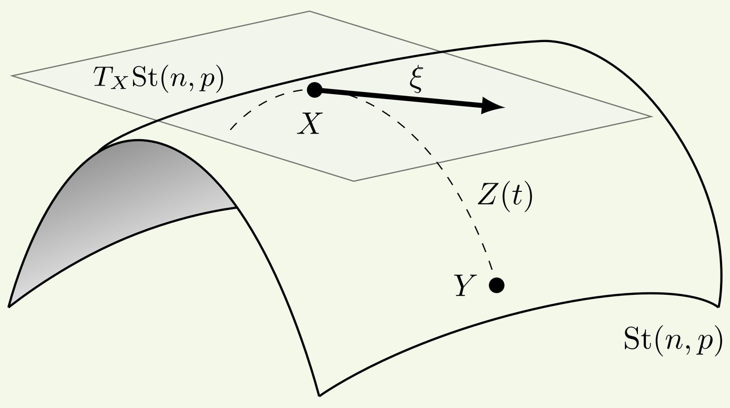
- Several applications in optimization, image and signal processing deal with data belonging to the **Stiefel manifold**
$$\mathrm{St}(n,p) = \{X \in \mathbb{R}^{n\times p}:\ X^\top X = I_p\}.$$
- Some applications require evaluating the geodesic distance between two arbitrary points on $\mathrm{St}(n,p)$. No closed-form solution is known for $\mathrm{St}(n,p)$.
- A new computational framework for computing the geodesic distance is proposed, based on the multiple shooting method and the leapfrog algorithm by L. Noakes.
- **Two example applications**:
  - Karcher mean on the space of probability density functions (PDFs);
  - Interpolation of data belonging to $\mathrm{St}(n,p)$ for parametric model reduction.

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

- **Geodesic**: generalization of straight lines to manifolds.
- When the tangent space $T_X\mathrm{St}(n,p)$ is endowed with the canonical metric
$$g_c(\Delta,\Delta) = \mathrm{tr}(\Delta^\top(I - \tfrac{1}{2}XX^\top)\Delta), \quad \Delta \in T_X\mathrm{St}(n,p),$$
one can get the following ODE for the geodesic $Z \equiv Z(t)$ [1, eq. (2.41)]:
$$\ddot{Z} + \dot{Z}\dot{Z}^\top Z + Z((Z^\top\dot{Z})^2 + \dot{Z}^\top\dot{Z}) = 0.$$
- Closed-form solution for a geodesic $Z(t)$ that realizes a tangent vector Δ with base point X (Ross Lippert [1, eq. (2.42)]):
$$Z(t) = [X\ \ X_\perp]\exp\left(\begin{bmatrix} X^\top\Delta & -(X_\perp^\top\Delta)^\top \\ X_\perp^\top\Delta & O\end{bmatrix} t\right)\begin{bmatrix} I_p \\ O\end{bmatrix}.$$

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

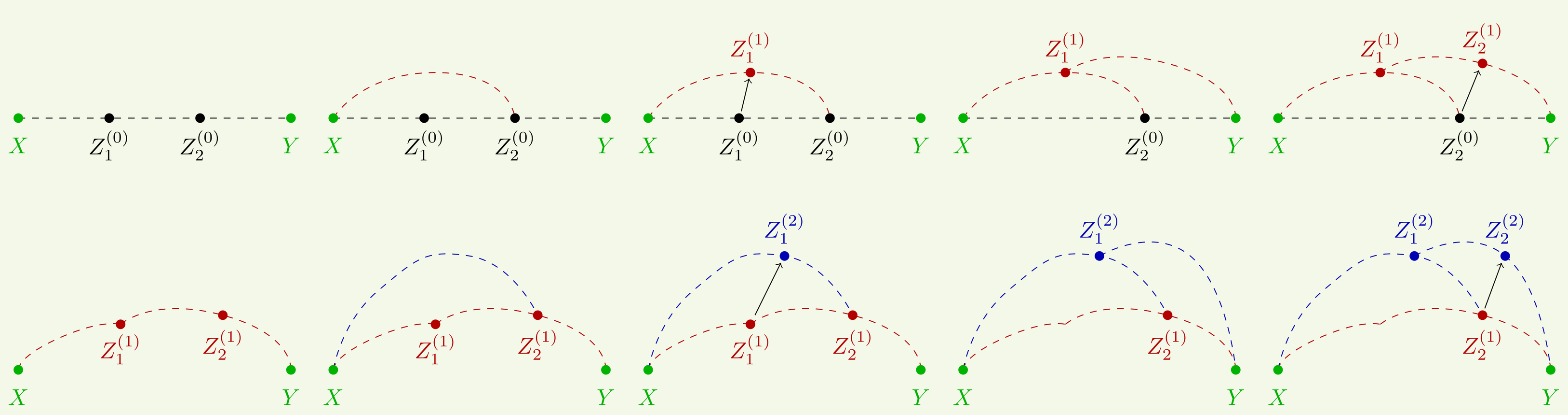
- Given $X,\ Y \in \mathrm{St}(n,p)$, the **geodesic distance** $d(X,Y)$ is the length of $\Delta_* \equiv \dot{Z}(0) \in T_X\mathrm{St}(n,p)$ s.t. the Riemannian exponential mapping $\mathrm{Exp}_X(\Delta_*) = Y$.
- Equivalent to: Find the **Riemannian logarithm** of Y with base point X, i.e., $\mathrm{Log}_X(Y) = \Delta_*$.



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

Geodesics via leapfrog (by L. Noakes [2])

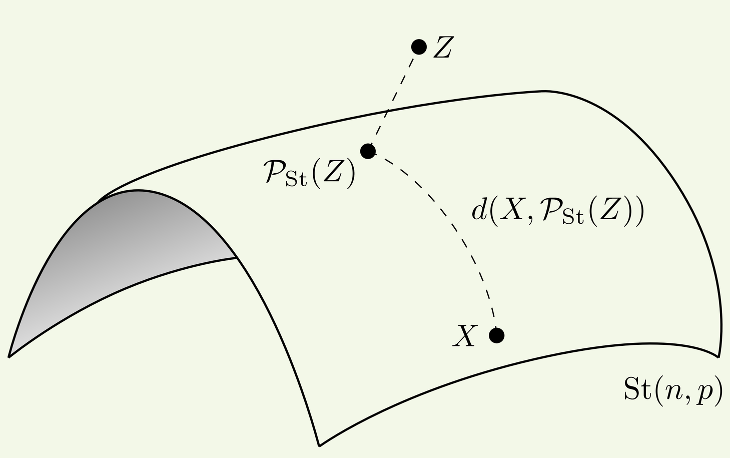
- Based on subdivision, s.t. single shooting works well on each subinterval.
- Illustration of two iterations of the procedure, for m points:



- Global convergence to Δ_*, but very slow. Deteriorates when $m \to \infty$.

Geodesics via nonlinear block Gauss–Seidel [3]

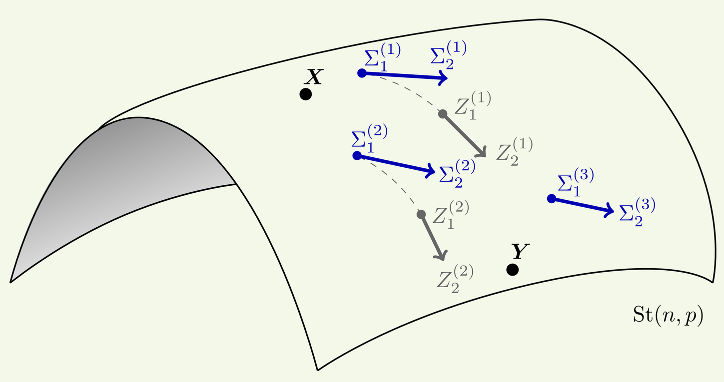
- **Alternating minimization**: cyclically minimize F over each block variable X_i
$$\min_{X\in\mathcal{X}} F(X_1,\ldots,X_s)$$
while fixing the other blocks at their last updated values.
  - Let X_i^k denote the value of X_i after its kth update, and let
$$F_i^k(X_i) = F(X_1^k,\ldots,X_{i-1}^k,X_i,X_{i+1}^{k-1},\ldots,X_s^{k-1}), \quad \forall i,\ \forall k.$$
  - At each step, the update is [3, Eq. (1.3a)]
$$X_i^k = \arg\min_{X_i\in\mathcal{X}_i^k} F_i^k(X_i).$$
- **Nonlinear block Gauss–Seidel** or **block coordinate descent method** [3].
  - Theory in [3] applies only in Euclidean space $\mathbb{R}^n$, not on Riemannian manifolds.
  - Smooth extension of Riemannian distance function $d\colon\ \mathrm{St}(n,p)\times\mathrm{St}(n,p) \to \mathbb{R}_{\geq 0}$ as $d^2_{\mathrm{ext}}\colon\ \mathrm{St}(n,p)\times\mathbb{R}^{n\times p} \to \mathbb{R}_{\geq 0}$:
$$d^2_{\mathrm{ext}}(X,Z) = d^2(X,\mathcal{P}(Z)) + \|\mathcal{P}(Z) - Z\|_F^2,$$
where $\mathcal{P}\colon \mathbb{R}^{n\times p} \to \mathrm{St}(n,p)$ is the projector on $\mathrm{St}(n,p)$.



  - Proof of convergence based on showing local strong convexity of $d^2_{\mathrm{ext}}(X_{i-1}^k, X_i)$.
  - Connection to the **Karcher mean** of two points X_{i-1}^k and X_{i-1}^{k-1}.

Geodesics via multiple shooting

- Enforce continuity conditions of Z and $\dot{Z}$ at the interfaces between subintervals.
- Fast convergence to Δ_*.
- $\Sigma_1^{(k)}$: point on $\mathrm{St}(n,p)$ relative to the kth subinterval.
- $\Sigma_2^{(k)}$: tangent vector to $\mathrm{St}(n,p)$ at $\Sigma_1^{(k)}$.



System of nonlinear equations:

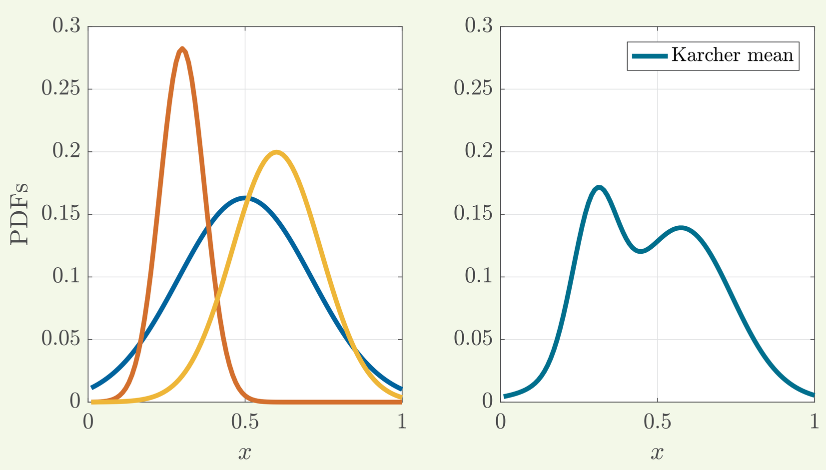
For each subinterval k, we have an **explicit expression** for the Jacobian $G^{(k)}$.

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

- Complexity of multiple shooting with *condensing* is $O(mn^3p^3)$.

Karcher mean of univariate probability density functions

- **Karcher mean**: one possible notion of mean on a Riemannian manifold $\mathcal{M}$, defined by the optimization problem
$$\mu = \arg\min_{p\in\mathcal{M}} \frac{1}{2N}\sum_{i=1}^N d(p,q_i)^2,$$
where $d(p,q_i)$ is the distance between two points on $\mathcal{M}$.



- $\mathcal{S}^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$ can be used to approximate $\mathcal{S}^\infty$, which represents the **space of univariate PDFs** on the unit interval $[0,1]$, i.e., $\mathcal{P} = \{g : [0,1] \to \mathbb{R}_{\geq 0} : \int_0^1 g(x)\,\mathrm{d}x = 1\}$.
- **Example**: Karcher mean of 3 PDFs, sampled at 100 points, which makes them belonging to $\mathrm{St}(100,1)$.

Model reduction with POD and interpolation on $\mathrm{St}(n,r)$

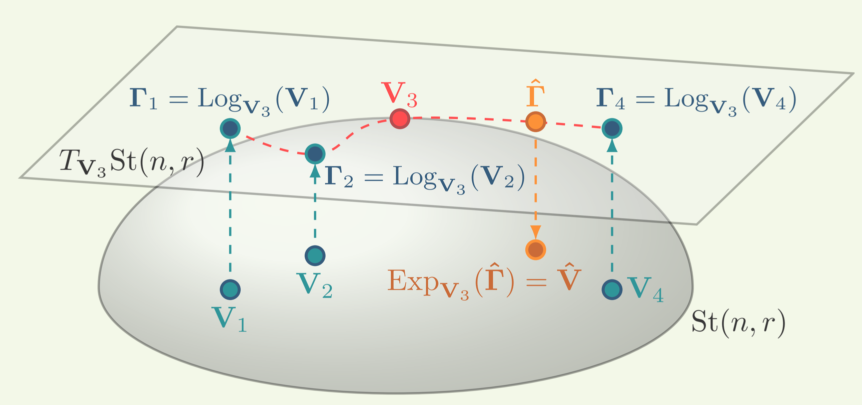
- Model reduction for dynamical systems parametrized with $\boldsymbol{p} = [p_1,\ldots,p_d]^\top$:

$$\begin{cases} \dot{x}(t;\boldsymbol{p}) = A(\boldsymbol{p})\,x(t;\boldsymbol{p}) + B(\boldsymbol{p})\,u(t), \\ y(t;\boldsymbol{p}) = C(\boldsymbol{p})\,x(t;\boldsymbol{p}), \end{cases} \xrightarrow{\text{reduction}} \begin{cases} \dot{x}_r(t;\boldsymbol{p}) = A_r(\boldsymbol{p})\,x_r(t;\boldsymbol{p}) + B_r(\boldsymbol{p})\,u(t), \\ y_r(t;\boldsymbol{p}) = C_r(\boldsymbol{p})\,x_r(t;\boldsymbol{p}), \end{cases}$$

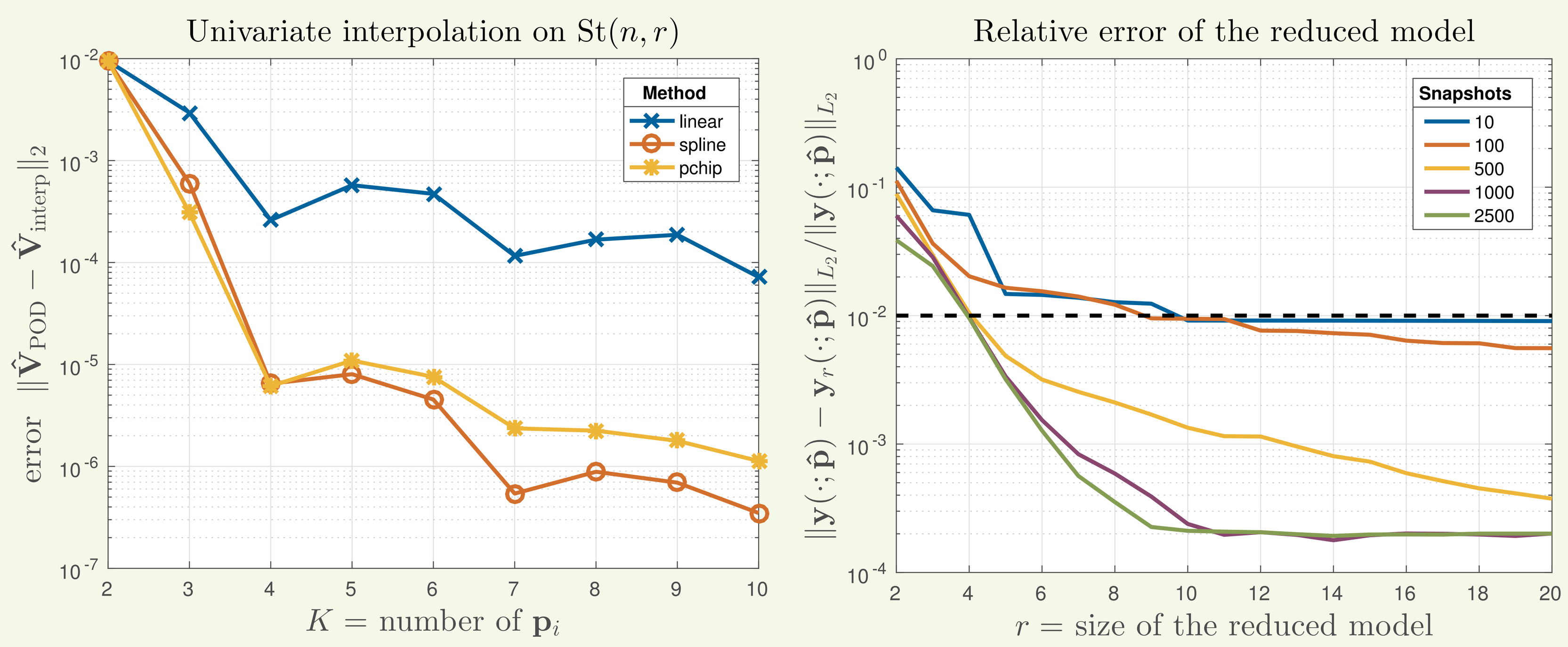
$x(t;\boldsymbol{p}) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^q$, $A(\boldsymbol{p}) \in \mathbb{R}^{n\times n}$, $B(\boldsymbol{p}) \in \mathbb{R}^{n\times m}$, $C(\boldsymbol{p}) \in \mathbb{R}^{q\times n}$.

$x_r = V^\top x$, $A_r = V^\top A V$, $B_r = V^\top B$, $C_r = CV$, $V \equiv V(\boldsymbol{p}) \in \mathrm{St}(n,r)$.

- For each parameter in a set of parameter values $\{\boldsymbol{p}_1,\boldsymbol{p}_2,\ldots,\boldsymbol{p}_K\}$, use proper orthogonal decomposition (POD) to derive a reduced-order basis $\boldsymbol{V}_i \in \mathrm{St}(n,r)$.



- This yields a set of local basis matrices $\{\boldsymbol{V}_1,\boldsymbol{V}_2,\ldots,\boldsymbol{V}_K\}$.
- Given a new parameter value $\hat{p}$, a basis $\hat{\boldsymbol{V}}$ can be obtained by **interpolating the local basis matrices on a tangent space to $\mathrm{St}(n,r)$**.
- **Application**: 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 $\boldsymbol{y}(\cdot;\hat{\boldsymbol{p}})$ and $\boldsymbol{y}_r(\cdot;\hat{\boldsymbol{p}})$ is about 1%.



Essential references

[1] A. Edelman, T. A. Arias, and S. T. Smith. The Geometry of Algorithms with Orthogonality Constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[2] J. L. Noakes. A global algorithm for geodesics. *Journal of the Australian Mathematical Society. Series A. Pure Mathematics and Statistics*, 65(1):37–50, 1998.

[3] Y. Xu and W. Yin. A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, 2013.