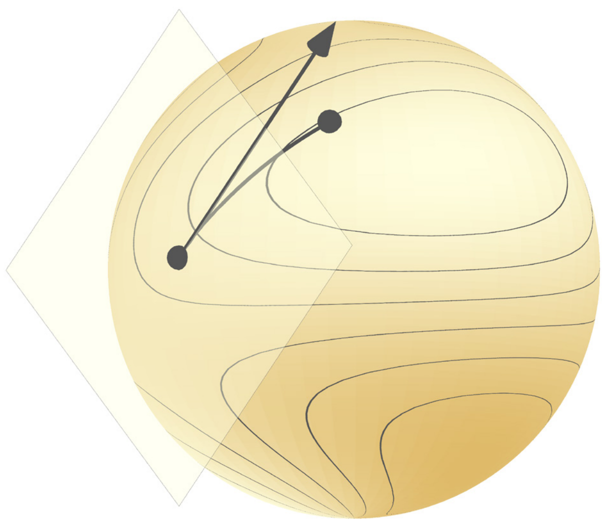
113

Transporters: a proxy for parallel transport



Spring 2023

Optimization on manifolds, MATH 512 @ EPFL

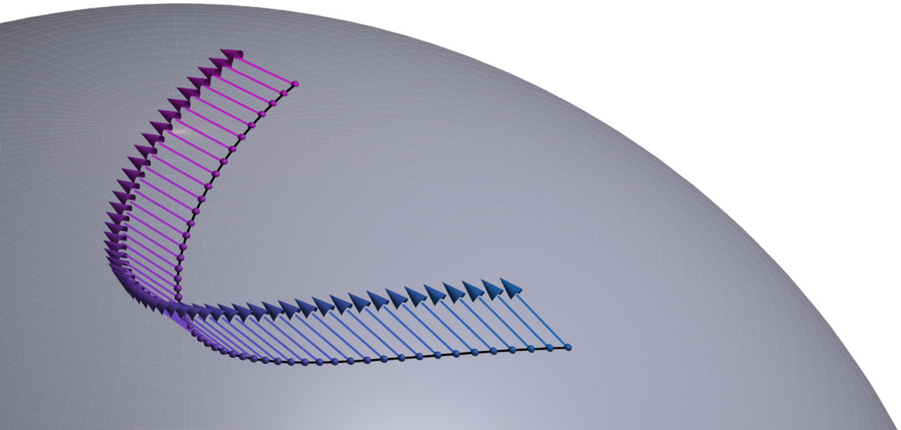
Instructor: Nicolas Boumal

Parallel transport is not always convenient

Computing $\mathrm{PT}^{c}_{t_2 \leftarrow t_1}$ requires solving an ODE. Could be expensive…

Also, you really have to choose a curve; can't just pick x and y.

Theoretically, it is "the" right tool, but it often pays to be pragmatic.



What we need, at least, to move vectors

Given $x, y \in \mathcal{M}$, and $v \in T_x\mathcal{M}$,

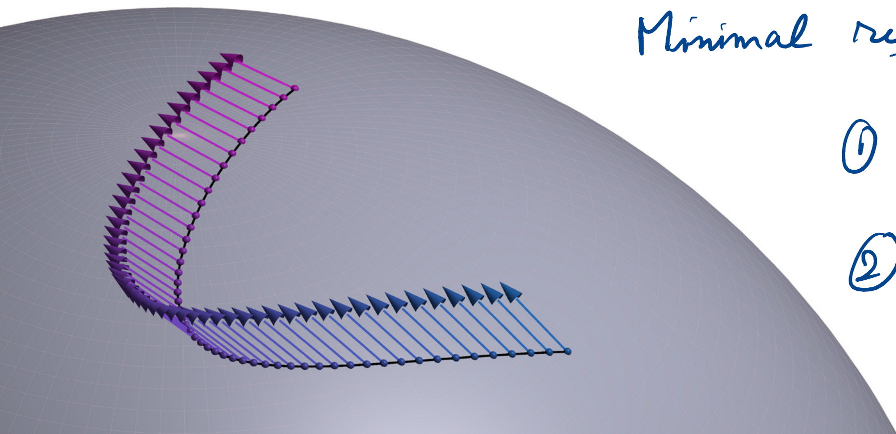
we want a linear map $T_{y \leftarrow x} : T_x\mathcal{M} \to T_y\mathcal{M}$

s.t. $T_{y\leftarrow x}(v) \in T_y\mathcal{M}$ can "play the role of v at y".

Minimal requirements:

① $T_{x \leftarrow x}(v) = v$,

② $T_{y \leftarrow x}$ depends "smoothly" on x and y.



Example 1: If $\mathcal{M}$ is the set of positive definite matrices.

$\mathcal{M}$ is open in its embedding space, so: $T_x\mathcal{M} = \mathcal{E}$.

Pragmatic choice: $T_{y \leftarrow x}(v) = v$, because $T_x\mathcal{M} = T_y\mathcal{M} = \mathcal{E}$.

Example 2: If $\mathcal{M}$ is the set of rotation matrices.

$$\mathcal{M} = \{ X \in \mathbb{R}^{d \times d} : X^T X = I_d, \ \det(X) = 1 \}.$$

$$T_X\mathcal{M} = \{ X\Omega : \Omega \in \mathbb{R}^{d \times d}, \ \Omega + \Omega^T = 0 \}$$

Pragmatic choice: $T_{Y \leftarrow X}(X\Omega) = Y\Omega$. $\Big\}$ Lie groups

Example 3: If $\mathcal{M}$ is embedded in a Euclidean space $\mathcal{E}$.

$T_x\mathcal{M} \subseteq \mathcal{E}$, $\quad T_y\mathcal{M} \subseteq \mathcal{E}$

Pragmatic choice: $T_{y\leftarrow x}(v) = \mathrm{Proj}_y(v)$.

$$\big(\underbrace{(x, v), y}_{T\mathcal{M} \times \mathcal{M}}\big) \longmapsto \underbrace{(y, T_{y\leftarrow x}(v))}_{T\mathcal{M}}$$

Def.: A transporter T provides, for all $x, y \in \mathcal{M}$, a linear map

$$T_{y\leftarrow x}: \mathrm{T}_x\mathcal{M} \to \mathrm{T}_y\mathcal{M}$$

s.t. each $T_{x\leftarrow x}$ is identity and dependence on x, y is smooth.

Use cases

Algorithms (e.g., nonlinear CG and BFGS).

Finite differences: Given any curve with $c(0) = x$ and $c'(0) = u$,

$$\mathrm{Hess}f(x)[u] = \lim_{t\to 0} \frac{T_{x\leftarrow c(t)}\Big(\mathrm{grad}f\big(c(t)\big)\Big) - \mathrm{grad}f(x)}{t} + O(\|\mathrm{grad}f(x)\|_x)$$

Good enough near critical points.

In some cases, the $O(\cdot)$ term is exactly zero: recall slides 111.

(see also §10.6)