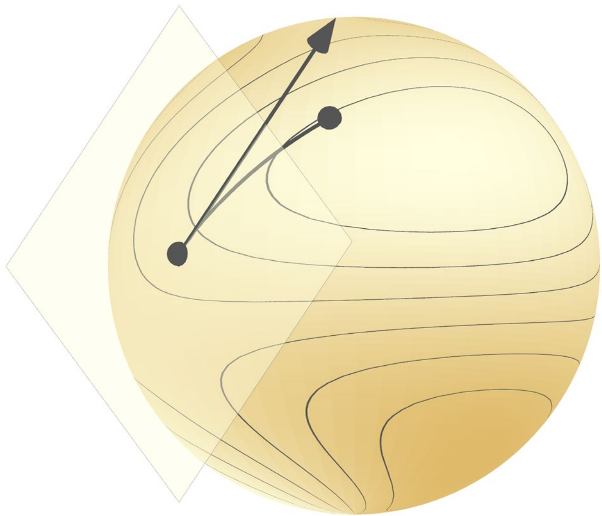
002

Bird's-eye view, and aims

Spring 2023

Optimization on manifolds, MATH 512 @ EPFL

Instructor: Nicolas Boumal

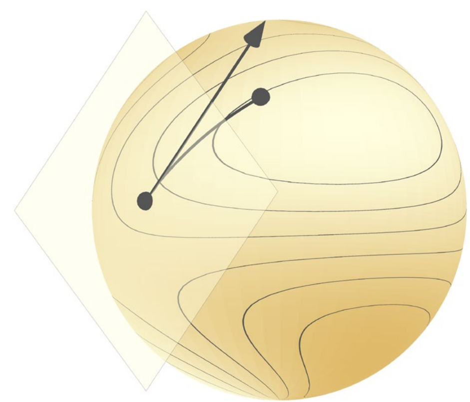


What tools do we need solve $\min_{x \in \mathcal{M}} f(x)$?

To be concrete, say $\mathcal{M}$ is the unit sphere:

$$S^{n-1} = \{x \in \mathbf{R}^n : x^\top x = 1\}$$

1. What does it mean that S^{n-1} is smooth?
2. What does it mean for $f : S^{n-1} \to \mathbf{R}$ to be smooth?



① $h(x) = x^\top x - 1$; $S^{n-1} = \{ x \in \mathbb{R}^n : h(x) = 0 \}$.

Pick $x \in S^{n-1}$; Consider $v \in \mathbb{R}^n$.

$$h(x+v) \simeq \underbrace{h(x)}_{=0} + \underbrace{Dh(x)[v]}_{=0 \text{ if } v \in \ker Dh(x)}$$

$$Dh(x)[v] = \lim_{t \to 0} \frac{h(x+tv) - h(x)}{t} = \lim_{t \to 0} \frac{t x^T v + t v^T x + t^2 v^T v}{t}$$

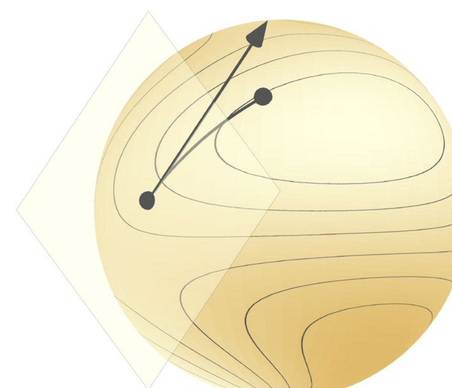
$$= 2x^T v.$$

$$\ker Dh(x) = \{ v \in \mathbb{R}^n : x^T v = 0 \} \triangleq T_x S^{n-1}.$$

② $\bar{f} : \mathbb{R}^n \to \mathbb{R}$, smooth (C^∞); ? $f = \bar{f}|_{S^{n-1}} : S^{n-1} \to \mathbb{R}$ smooth.

Simplest algorithm to minimize $g : \mathbf{R}^n \to \mathbf{R}$ is gradient descent:

$$x_{k+1} = x_k - \alpha \operatorname{grad} g(x_k)$$



How to extend this for $f : \mathrm{S}^{n-1} \to \mathbf{R}$?

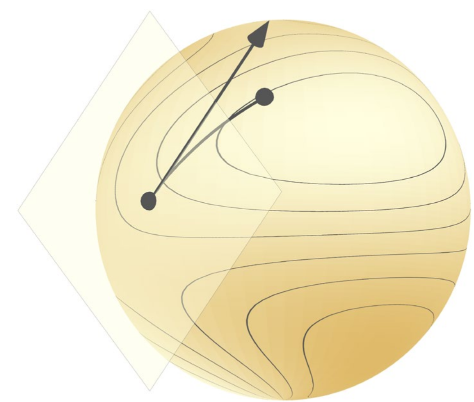
3. How can we move around on the sphere?

Given $x \in S^{n-1}$, $v \in T_x S^{n-1}$,

pick a new point on the sphere:

$$R_x(v) = \frac{x+v}{\|x+v\|} \quad \text{(one possible choice).}$$

retraction



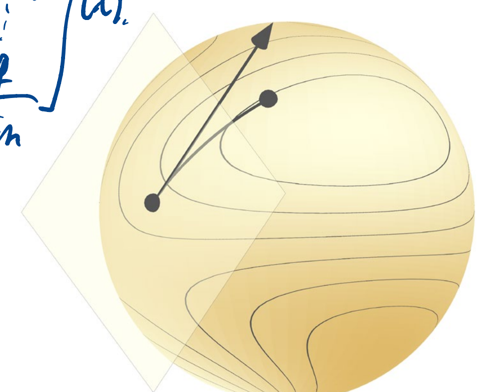
4. What could be a reasonable notion of gradient for $f : S^{n-1} \to \mathbf{R}$?

$g : \mathbb{R}^n \to \mathbb{R}$; $\qquad Dg(x)[v] = \lim_{t \to 0} \dfrac{g(x+tv) - g(x)}{t}$.

$\langle \underbrace{\mathrm{grad}\, g(x)}_{\mathbb{R}^n}, v \rangle = Dg(x)[v]$; need inner product.

If $\langle u, v \rangle = u^T v$, then $\mathrm{grad}\, g(x) = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{bmatrix}(x)$.

$\langle \cdot , \cdot \rangle_x$



Beyond gradient descent

More advanced optimization algorithms in $\mathbf{R}^n$ may use Hessians.

The Hessian of f is the derivative of its gradient vector field.

Differentiating vector fields on manifolds requires care.

We'll study Riemannian connections,

and build from there to construct, e.g., trust-region methods.

If you work diligently, you will be:

Fluent in basic differential and Riemannian geometry,

with an emphasis on computation, beginning with embedded geometry;

Familiar with both first- and second-order algorithms on manifolds,

in terms of both mathematics and implementation;

Acquainted with some advanced topics to be determined,

such as quotient manifolds, geodesic convexity, vector transports++, ...

Ready to tackle research and applied challenges in+with $\min_{\mathcal{M}} f$.