# An introduction to optimization on manifolds



## Optimization on smooth manifolds min f(x) subject to $x \in M$

Linear spaces Unconstrained; linear equality constraints Low rank (matrices, tensors) Recommender systems, large-scale Lyapunov equations, ... Orthonormality (Grassmann, Stiefel, rotations) Dictionary learning, SfM, SLAM, PCA, ICA, SBM, Electr. Struct. Comp.... Positivity (positive definiteness, positive orthant) Metric learning, Gaussian mixtures, diffusion tensor imaging, ... Symmetry (quotient manifolds) Invariance under group actions

## A Riemannian structure gives us gradients and Hessians

The essential tools of smooth optimization are defined generally on Riemannian manifolds.

Unified theory, broadly applicable algorithms.

First ideas from the '70s. First practical in the '90s.

















#### Welcome to Manopt!

#### **Toolboxes for optimization on manifolds and matrices**

Optimization on manifolds is a powerful paradigm to address nonlinear optimization problems.

With Manopt, it is easy to deal with various types of constraints and symmetries which arise naturally in applications, such as orthonormality, low rank, positivity and invariance under group actions.

These tools are also perfectly suited for unconstrained optimization with vectors and matrices.





With Bamdev Mishra, P.-A. Absil & R. Sepulchre



Contact

Lead by J. Townsend, N. Koep & S. Weichwald



Lead by Ronny Bergmann

#### manopt.org



t 🔒 Home

A Tutorial L Downloads

OPTIMIZATION ALGORITHMS ON MATRIX MANIFOLDS

P.-A. ABSIL, R. MAHONY & R. SEPULCHRE

🕑 Forum 🔒 About

#### nicolasboumal.net/book

NICOLAS BOUMAL

#### AN INTRODUCTION TO OPTIMIZATION ON SMOOTH MANIFOLDS

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Wi P.-4

#### What is a smooth manifold?



This overview: restricted to embedded submanifolds of linear spaces.

#### What is a smooth manifold? (2)

 $\mathcal{M} \subseteq \mathcal{E}$  is an embedded submanifold of codimension k if: There exists a smooth function  $h: \mathcal{E} \to \mathbf{R}^k$  such that:

 $\mathcal{M} \equiv h(x) = 0$  and  $Dh(x): \mathcal{E} \to \mathbf{R}^k$  has rank  $k \; \forall x \in \mathcal{M}$ 

12

x.

E.g., sphere:  $h(x) = x^T x - 1$  and  $Dh(x)[v] = 2x^T v$ 

These properties allow us to linearize the set:  $h(x + tv) = h(x) + tDh(x)[v] + O(t^2)$ This is  $O(t^2)$  iff  $v \in \ker Dh(x)$ .  $\rightarrow$ Tangent spaces:  $T_x \mathcal{M} = \ker Dh(x)$ , (subspace of  $\mathcal{E}$ )

#### What is a smooth manifold? (3)

 $\mathcal{M} \subseteq \mathcal{E}$  is an embedded submanifold of codimension *k* if:

For each  $x \in \mathcal{M}$  there is a neighborhood U of x in  $\mathcal{E}$  and a smooth function  $h: U \to \mathbf{R}^k$  such that:

a)  $Dh(x): \mathcal{E} \to \mathbf{R}^k$  has rank *k*, and

b) 
$$\mathcal{M} \cap U = h^{-1}(0) = \{y \in U : h(y) = 0\}$$

(Necessary for fixed-rank matrices.)

#### Bootstrapping our set of tools

- 1. Smooth maps
- 2. Differentials of smooth maps
- 3. Vector fields and tangent bundles
- 4. Retractions
- 5. Riemannian metrics
- 6. Riemannian gradients
- 7. Riemannian connections
- 8. Riemannian Hessians
- 9. Riemannian covariant derivatives along curves

#### Smooth maps between manifolds

With  $\mathcal{M}, \mathcal{M}'$  embedded in  $\mathcal{E}, \mathcal{E}'$ ,

**Define:** a map  $F: \mathcal{M} \to \mathcal{M}'$  is smooth if:

There exists a neighborhood U of  $\mathcal{M}$  in  $\mathcal{E}$  and a smooth map  $\overline{F}: U \to \mathcal{E}'$  such that  $F = \overline{F}|_{\mathcal{M}}$ .

We call  $\overline{F}$  a smooth extension of F.

#### **Differential** of a smooth map

For 
$$\overline{F}: \mathcal{E} \to \mathcal{E}'$$
, we have  $D\overline{F}(x)[v] = \lim_{t \to 0} \frac{\overline{F}(x+tv) - \overline{F}(x)}{t}$ 

For  $F: \mathcal{M} \to \mathcal{M}'$  smooth, F(x + tv) may not be defined. But picking a smooth extension  $\overline{F}$ , for  $v \in T_x \mathcal{M}$  we say:

**Define:** 
$$DF(x)[v] = \lim_{t \to 0} \frac{\overline{F}(x+tv) - \overline{F}(x)}{t} = D\overline{F}(x)[v].$$

**Claim:** this does not depend on choice of  $\overline{F}$ . **Claim:** we retain linearity, product rule & chain rule.

### Differential of a smooth map (2)

We defined  $DF(x) = D\overline{F}(x)|_{T_x\mathcal{M}}$ . Equivalently:

**Claim:** for each  $v \in T_x \mathcal{M}$ , there exists a smooth curve  $c: \mathbb{R} \to \mathcal{M}$  such that c(0) = x and c'(0) = v.



## **Differential of a smooth map (2)** We defined $DF(x) = D\overline{F}(x)|_{T_x\mathcal{M}}$ . Equivalently:

**Claim:** for each  $v \in T_x \mathcal{M}$ , there exists a smooth curve  $c: \mathbf{R} \to \mathcal{M}$  such that c(0) = x and c'(0) = v.

**Define:** 
$$DF(x)[v] = \lim_{t \to 0} \frac{F(c(t)) - F(x)}{t} = (F \circ c)'(0).$$



#### Vector fields and the tangent bundle

A vector field *V* ties to each *x* a vector  $V(x) \in T_x \mathcal{M}$ .

What does it mean for *V* to be smooth?

**Define:** the tangent bundle of  $\mathcal{M}$  is the disjoint union of all tangent spaces:

$$T\mathcal{M} = \{(x, v) : x \in \mathcal{M} \text{ and } v \in T_x \mathcal{M}\}$$

**Claim:** T $\mathcal{M}$  is a manifold (embedded in  $\mathcal{E} \times \mathcal{E}$ ).

*V* is smooth if it is smooth as a map from  $\mathcal{M}$  to  $T\mathcal{M}$ .

#### Aside: new manifolds from old ones

Given a manifold  $\mathcal{M}$ , we can create a new manifold by considering its tangent bundle.

Here are other ways to recycle:

- **Products** of manifolds:  $\mathcal{M} \times \mathcal{M}'$ ,  $\mathcal{M}^n$
- Open subsets of manifolds\*
- (Quotienting some equivalence relations.)

\*For embedded submanifolds: subset topology.

#### Retractions: moving around

Given  $(x, v) \in T\mathcal{M}$ , we can move away from x along v using any  $c: \mathbf{R} \to \mathcal{M}$  with c(0) = x and c'(0) = v.

Retractions are a smooth choice of curves over  $T\mathcal{M}$ .

**Define:** A retraction is a smooth map  $R: T\mathcal{M} \to \mathcal{M}$ such that if  $c(t) = R(x, tv) = R_x(tv)$ then c(0) = x and c'(0) = v.

x

#### Retractions: moving around (2)

**Define:** a retraction is a smooth map  $R: T\mathcal{M} \to \mathcal{M}$ such that if  $c(t) = R(x, tv) = R_x(tv)$ , then c(0) = x and c'(0) = v.

Equivalently:  $R_x(0) = x$  and  $DR_x(0) = Id$ .

Typical choice:  $R_x(v) =$ projection of x + v to  $\mathcal{M}$ :

• 
$$\mathcal{M} = \mathbf{R}^n, R_{\chi}(v) = x + v$$
  
•  $\mathcal{M} = \mathbf{S}^{n-1}, R_{\chi}(v) = \frac{x+v}{\|x+v\|}$   
•  $\mathcal{M} = \mathbf{R}^{m \times n}, R_{\chi}(V) = \mathrm{SVD}_{r}(X+V)$ 

#### Towards gradients: reminders from $\mathbf{R}^n$

The gradient of a smooth  $\overline{f} : \mathbf{R}^n \to \mathbf{R}$  at x is defined by:

$$D\overline{f}(x)[v] = \langle \operatorname{grad} \overline{f}(x), v \rangle$$
 for all  $v \in \mathbf{R}^n$ .

In particular, with the inner product  $\langle u, v \rangle = u^{\top} v$ ,

grad
$$\bar{f}(x)_i = \langle \operatorname{grad}\bar{f}(x), e_i \rangle = \mathrm{D}\bar{f}(x)[e_i] = \frac{\partial \bar{f}}{\partial x_i}(x).$$

Note: grad  $\overline{f}$  is a smooth vector field on  $\mathbb{R}^n$ .

#### **Riemannian metrics**

 $T_x \mathcal{M}$  is a linear space (subspace of  $\mathcal{E}$ ).

Pick an inner product  $\langle \cdot, \cdot \rangle_x$  for each  $T_x \mathcal{M}$ .

**Define:**  $\langle \cdot, \cdot \rangle_{\chi}$  defines a Riemannian metric on  $\mathcal{M}$  if for any two smooth vector fields V, W the function  $x \mapsto \langle V(x), W(x) \rangle_{\chi}$  is smooth.

A Riemannian manifold is a manifold with a Riemannian metric.

#### Riemannian submanifolds

Let  $\langle \cdot, \cdot \rangle$  be the inner product on  $\mathcal{E}$ .

Since  $T_x \mathcal{M}$  is a linear subspace of  $\mathcal{E}$ , one choice is:

$$\langle u, v \rangle_{\chi} = \langle u, v \rangle$$

**Claim:** this defines a Riemannian metric on  $\mathcal{M}$ .

With this metric,  $\mathcal{M}$  is a Riemannian submanifold of  $\mathcal{E}$ .

#### Riemannian gradient

Let  $f: \mathcal{M} \to \mathbf{R}$  be smooth on a Riemannian manifold.

**Define:** the Riemannian gradient of *f* at *x* is the unique tangent vector at *x* such that:

 $Df(x)[v] = \langle \operatorname{grad} f(x), v \rangle_{\chi} \text{ for all } v \in T_{\chi}\mathcal{M}$ 

**Claim:** grad *f* is a smooth vector field. **Claim:** if *x* is a local optimum of *f*, grad f(x) = 0. Gradients on Riemannian submanifolds Let  $\overline{f}$  be a smooth extension of f. For all  $v \in T_x \mathcal{M}$ :

$$\langle \operatorname{grad} f(x), v \rangle_{x} = \mathrm{D} f(x)[v]$$
  
=  $\mathrm{D} \overline{f}(x)[v] = \langle \operatorname{grad} \overline{f}(x), v \rangle$ 

Assume  $\mathcal{M}$  is a Riemannian submanifold of  $\mathcal{E}$ . Since  $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_x$ , by uniqueness we conclude:

$$\operatorname{grad} f(x) = \operatorname{Proj}_x \left( \operatorname{grad} \overline{f}(x) \right)$$

 $\operatorname{Proj}_{x}$  is the orthogonal projector from  $\mathcal{E}$  to  $T_{x}\mathcal{M}$ .

## A first algorithm: gradient descent For $\overline{f}$ : $\mathbf{R}^n \to \mathbf{R}$ , $x_{k+1} = x_k - \alpha_k \operatorname{grad} \overline{f}(x_k)$ For $f: \mathcal{M} \to \mathbf{R}$ , $x_{k+1} = R_{x_k} (-\alpha_k \operatorname{grad} f(x_k))$

For the analysis, need to understand  $f(x_{k+1})$ :

The composition  $f \circ R_x$ :  $T_x \mathcal{M} \to \mathbf{R}$  is on a linear space, hence we may Taylor expand it.

### A first algorithm: gradient descent

$$x_{k+1} = R_{x_k} \left( -\alpha_k \operatorname{grad} f(x_k) \right)$$

The composition  $f \circ R_x$ :  $T_x \mathcal{M} \to \mathbf{R}$  is on a linear space, hence we may Taylor expand it:

$$f(R_x(v)) = f(R_x(0)) + \langle \operatorname{grad}(f \circ R_x)(0), v \rangle_x + O(||v||_x^2)$$
$$= f(x) + \langle \operatorname{grad} f(x), v \rangle_x + O(||v||_x^2)$$

Indeed:  $D(f \circ R_x)(0)[v] = Df(R_x(0))[DR_x(0)[v]] = Df(x)[v].$ 

#### Gradient descent on $\mathcal M$

- **A1**  $f(x) \ge f_{\text{low}}$  for all  $x \in \mathcal{M}$
- A2  $f(R_x(v)) \le f(x) + \langle v, \operatorname{grad} f(x) \rangle + \frac{L}{2} ||v||^2$

Algorithm: 
$$x_{k+1} = R_{x_k} \left( -\frac{1}{L} \operatorname{grad} f(x_k) \right)$$



Complexity:  $\|\operatorname{grad} f(x_k)\| \leq \varepsilon$  with some  $k \leq 2L(f(x_0) - f_{\text{low}})\frac{1}{\varepsilon^2}$ 

$$\mathbf{A2} \Rightarrow f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{L} \| \operatorname{grad} f(\mathbf{x}_k) \|^2 + \frac{1}{2L} \| \operatorname{grad} f(\mathbf{x}_k) \|^2$$
$$\Rightarrow f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{1}{2L} \| \operatorname{grad} f(\mathbf{x}_k) \|^2$$
(for contradiction)  
$$\mathbf{A1} \Rightarrow f(\mathbf{x}_0) - f_{\text{low}} \geq \sum_{k=0}^{K-1} f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) > \frac{\varepsilon^2}{2L} K$$

#### Tips and tricks to get the gradient

Use definition as starting point:  $Df(x)[v] = \cdots$ 

Cheap gradient principle

Numerical check of gradient: Taylor  $t \mapsto f(R_x(tv))$ Manopt: checkgradient(problem)

Automatic differentiation: Python, Julia Not Matlab :/—this being said, for theory, often need to manipulate gradient "on paper" anyway.

#### On to second-order methods

Consider  $\overline{f} : \mathbf{R}^n \to \mathbf{R}$  smooth. Taylor says:

$$\bar{f}(x+v) \approx \bar{f}(x) + \langle \operatorname{grad}\bar{f}(x), v \rangle + \frac{1}{2} \langle v, \operatorname{Hess}\bar{f}(x)[v] \rangle$$

If  $\text{Hess}\bar{f}(x) > 0$ , quadratic model minimized for v s.t.:

$$\operatorname{Hess} \overline{f}(x)[v] = -\operatorname{grad} \overline{f}(x)$$

From there, we can construct Newton's method etc.

Towards Hessians: reminders from  $\mathbb{R}^n$ Consider  $\overline{f}: \mathbb{R}^n \to \mathbb{R}$  smooth.

The Hessian of  $\overline{f}$  at x is a linear operator which tells us how the gradient vector field of  $\overline{f}$  varies:

$$\operatorname{Hess}\overline{f}(x)[v] = \operatorname{Dgrad}\overline{f}(x)[v]$$

With  $\langle u, v \rangle = u^{\mathsf{T}} v$ , yields:  $\operatorname{Hess} \overline{f}(x)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ . Notice that  $\operatorname{Hess} \overline{f}(x)[v]$  is a vector in  $\mathbf{R}^n$ .

#### A difficulty on manifolds

A smooth vector field V on  $\mathcal{M}$  is a smooth map: we already have a notion of how to differentiate it.

**Example:** with 
$$f(x) = \frac{1}{2}x^{T}Ax$$
 on the sphere,  
 $V(x) = \operatorname{grad} f(x) = \operatorname{Proj}_{x}(Ax) = Ax - (x^{T}Ax)x$   
 $DV(x)[u] = \cdots = \operatorname{Proj}_{x}(Au) - (x^{T}Ax)u - (x^{T}Au)x$ 

**Issue:** DV(x)[u] may not be a tangent vector at x!

#### Connections: A tool to differentiate vector fields

Let  $\mathcal{X}(\mathcal{M})$  be the set of smooth vector fields on  $\mathcal{M}$ . Given  $U \in \mathcal{X}(\mathcal{M})$  and smooth f, (Uf)(x) = Df(x)[U(x)].

A map  $\nabla : \mathcal{X}(\mathcal{M}) \times \mathcal{X}(\mathcal{M}) \to \mathcal{X}(\mathcal{M})$  is a connection if: 1.  $\nabla_{fU+gW}(V) = f\nabla_U V + g\nabla_W V$ 2.  $\nabla_U(aV + bW) = a\nabla_U V + b\nabla_U W$ 3.  $\nabla_U(fV) = (Uf)V + f\nabla_U V$ 

**Example:** for  $\mathcal{M} = \mathcal{E}$ ,  $(\nabla_U V)(x) = DV(x)[U(x)]$ **Example:** for  $\mathcal{M} \subseteq \mathcal{E}$ ,  $(\nabla_U V)(x) = \operatorname{Proj}_x(DV(x)[U(x)])$ 

#### Riemannian connections: A unique and favorable choice

Let  $\mathcal{M}$  be a Riemannian manifold.

**Claim:** there exists a unique connection  $\nabla$  on  $\mathcal{M}$  s.t.:

4. 
$$(\nabla_U V - \nabla_V U)f = U(Vf) - V(Uf)$$

5. 
$$U\langle V, W \rangle = \langle \nabla_U V, W \rangle + \langle V, \nabla_U W \rangle$$

It is called the Riemannian connection (Levi-Civita).

**Claim:** if  $\mathcal{M}$  is a Riemannian submanifold of  $\mathcal{E}$ , then

 $(\nabla_U V)(x) = \operatorname{Proj}_x(\operatorname{D} V(x)[U(x)])$ 

is the Riemannian connection on  $\mathcal{M}$ .

#### Riemannian Hessians

**Claim:**  $(\nabla_U V)(x)$  depends on U only through U(x). This justifies the notation  $\nabla_u V$ ; e.g.:  $\nabla_u V = \operatorname{Proj}_x(DV(x)[u])$ 

**Define:** the Riemannian Hessian of  $f: \mathcal{M} \to \mathbf{R}$  at x is a linear operator from  $T_x \mathcal{M}$  to  $T_x \mathcal{M}$  defined by:

$$\operatorname{Hess} f(x)[u] = \nabla_{\!\!u} \operatorname{grad} f$$

where  $\nabla$  is the Riemannian connection.

**Claim:** Hess f(x) is self-adjoint. **Claim:** if x is a local minimum, then Hess  $f(x) \ge 0$ .

## Hessians on Riemannian submanifolds Hess $f(x)[u] = \nabla_u \operatorname{grad} f(x)$

On a Riemannian submanifold of a linear space,

$$\nabla_u V = \operatorname{Proj}_x(\mathrm{D}V(x)[u])$$

**Combining:** 

 $\text{Hess}f(x)[u] = \text{Proj}_x(\text{Dgrad}f(x)[u])$ 

#### Hessians on Riemannian submanifolds (2)

 $\text{Hess}f(x)[u] = \text{Proj}_x(\text{Dgrad}f(x)[u])$ 

**Example:**  $f(x) = \frac{1}{2}x^{T}Ax$  on the sphere in  $\mathbb{R}^{n}$ .  $V(x) = \operatorname{grad} f(x) = \operatorname{Proj}_{x}(Ax) = Ax - (x^{T}Ax)x$   $DV(x)[u] = \operatorname{Proj}_{x}(Au) - (x^{T}Ax)u - (x^{T}Au)x$  $\operatorname{Hess} f(x)[u] = \operatorname{Proj}_{x}(Au) - (x^{T}Ax)u$ 

Remarkably, grad f(x) = 0 and  $Hess f(x) \ge 0$  iff x optimal.

#### Newton, Taylor and Riemann

Now that we have a Hessian, we might guess:

$$f(R_x(v)) \approx m_x(v) = f(x) + \langle \operatorname{grad} f(x), v \rangle_x + \frac{1}{2} \langle v, \operatorname{Hess} f(x)[v] \rangle_x$$

If Hess f(x) is invertible,  $m_x: T_x \mathcal{M} \to \mathbf{R}$  has one critical point, solution of this linear system:

$$\text{Hess}f(x)[v] = -\text{grad}f(x) \text{ for } v \in T_x \mathcal{M}$$

Newton's method on  $\mathcal{M}$ :  $x_{next} = R_x(v)$ .

**Claim:** if  $\text{Hess}f(x^*) > 0$ , quadratic local convergence.

#### We need one more tool...

The truncated expansion

$$f(R_x(v)) \approx f(x) + \langle \operatorname{grad} f(x), v \rangle_x + \frac{1}{2} \langle v, \operatorname{Hess} f(x)[v] \rangle_x$$

is not quite correct (well, not always...)

To see why, let's expand *f* along some smooth curve.

We need one more tool... (2) Let's expand *f* along some smooth curve.

With  $c: \mathbf{R} \to \mathcal{M}$  s.t. c(0) = x and c'(0) = v, the composition  $g = f \circ c$  maps  $\mathbf{R} \to \mathbf{R}$ , so Taylor holds:  $g(t) \approx g(0) + tg'(0) + \frac{t^2}{2}g''(0)$ g(0) = f(x)

 $g'(t) = Df(c(t))[c'(t)] = \langle \operatorname{grad} f(c(t)), c'(t) \rangle_{c(t)}$  $g'(0) = \langle \operatorname{grad} f(x), v \rangle_{x}$ 

g''(0) = ???

#### Differentiating vector fields along curves

A map  $Z: \mathbb{R} \to T\mathcal{M}$  is a vector field along  $c: \mathbb{R} \to \mathcal{M}$  if Z(t) is a tangent vector at c(t).

Let  $\mathcal{X}(c)$  denote the set of smooth such fields.

**Claim:** there exists a unique operator  $\frac{D}{dt}: \mathcal{X}(c) \to \mathcal{X}(c)$  s.t. 1.  $\frac{D}{dt}(aY + bZ) = a \frac{D}{dt}Y + b \frac{D}{dt}Z$ 2.  $\frac{D}{dt}(gZ) = g'Z + g \frac{D}{dt}Z$ 3.  $\frac{D}{dt}(U(c(t))) = \nabla_{c'(t)}U$ 4.  $\frac{d}{dt}\langle Y(t), Z(t)\rangle_{c(t)} = \left\langle \frac{D}{dt}Y(t), Z \right\rangle_{c(t)} + \left\langle Y(t), \frac{D}{dt}Z(t) \right\rangle_{c(t)}$ 

where  $\nabla$  is the Riemannian connection.

#### Differentiating fields along curves (2)

**Claim:** there exists a unique operator  $\frac{D}{dt}$ :  $\mathcal{X}(c) \to \mathcal{X}(c)$  s.t.

1. 
$$\frac{D}{dt}(aY + bZ) = a\frac{D}{dt}Y + b\frac{D}{dt}Z$$
  
2. 
$$\frac{D}{dt}(gZ) = g'Z + g\frac{D}{dt}Z$$
  
3. 
$$\frac{D}{dt}(U(c(t))) = \nabla_{c'(t)}U$$
  
4. 
$$\frac{d}{dt}\langle Y(t), Z(t)\rangle_{c(t)} = \left\langle \frac{D}{dt}Y(t), Z\right\rangle_{c(t)} + \left\langle Y(t), \frac{D}{dt}Z(t)\right\rangle_{c(t)}$$

where  $\nabla$  is the Riemannian connection.

**Claim:** if 
$$\mathcal{M}$$
 is a Riemannian submanifold of  $\mathcal{E}$ ,  
$$\frac{\mathrm{D}}{\mathrm{d}t}Z(t) = \operatorname{Proj}_{c(t)}\left(\frac{\mathrm{d}}{\mathrm{d}t}Z(t)\right)$$

1. 
$$\frac{D}{dt}(aY + bZ) = a\frac{D}{dt}Y + b\frac{D}{dt}Z$$
  
2. 
$$\frac{D}{dt}(gZ) = g'Z + g\frac{D}{dt}Z$$
  
3. 
$$\frac{D}{dt}(U(c(t))) = \nabla_{c'(t)}U$$
  
4. 
$$\frac{d}{dt}\langle Y(t), Z(t)\rangle_{c(t)} = \left\langle \frac{D}{dt}Y(t), Z\right\rangle_{c(t)} + \left\langle Y(t), \frac{D}{dt}Z(t)\right\rangle_{c(t)}$$

With g(t) = f(c(t)) and  $g'(t) = \langle \operatorname{grad} f(c(t)), c'(t) \rangle_{c(t)}$ :

$$g''(t) = \left\langle \frac{\mathrm{D}}{\mathrm{d}t} \operatorname{grad} f(c(t)), c'(t) \right\rangle_{c(t)} + \left\langle \operatorname{grad} f(c(t)), \frac{\mathrm{D}}{\mathrm{d}t} c'(t) \right\rangle_{c(t)}$$
$$= \left\langle \nabla_{c'(t)} \operatorname{grad} f, c'(t) \right\rangle_{c(t)} + \left\langle \operatorname{grad} f(c(t)), c''(t) \right\rangle_{c(t)}$$
$$g''(0) = \left\langle \operatorname{Hess} f(x)[v], v \right\rangle_{x} + \left\langle \operatorname{grad} f(x), c''(0) \right\rangle_{x}$$

With g(t) = f(c(t)) and  $g'(t) = \langle \operatorname{grad} f(c(t)), c'(t) \rangle_{c(t)}$ :

$$g''(t) = \left\langle \frac{\mathrm{D}}{\mathrm{d}t} \operatorname{grad} f(c(t)), c'(t) \right\rangle_{c(t)} + \left\langle \operatorname{grad} f(c(t)), \frac{\mathrm{D}}{\mathrm{d}t} c'(t) \right\rangle_{c(t)}$$

$$= \left\langle \nabla_{c'(t)} \operatorname{grad} f, c'(t) \right\rangle_{c(t)} + \left\langle \operatorname{grad} f(c(t)), c''(t) \right\rangle_{c(t)}$$
$$g''(0) = \left\langle \operatorname{Hess} f(x)[v], v \right\rangle_{x} + \left\langle \operatorname{grad} f(x), c''(0) \right\rangle_{x}$$

$$f(c(t)) = f(x) + t \cdot \langle \operatorname{grad} f(x), v \rangle_{x} + \frac{t^{2}}{2} \cdot \langle \operatorname{Hess} f(x)[v], v \rangle_{x}$$
$$+ \frac{t^{2}}{2} \cdot \langle \operatorname{grad} f(x), c''(0) \rangle_{x} + O(t^{3})$$

The annoying term vanishes at critical points and for special curves (special retractions). Mostly fine for optimization.

#### Trust-region method: Newton's with a safeguard

With the same tools, we can design a Riemannian trust-region method: RTR (Absil, Baker & Gallivan '07).

Approximately minimizes quadratic model of  $f \circ R_{x_k}$ restricted to a ball in  $T_{x_k}\mathcal{M}$ , with dynamic radius.

Complexity known. Excellent performance, also with approximate Hessian (e.g., finite differences of grad f).

In Manopt, call trustregions(problem).

#### nicolasboumal.net/book

## In the lecture notes:

- Proofs for all claims in these slides
- References to the (growing) literature
- Short descriptions of applications
- Fully worked out manifolds E.g.: Stiefel, fixed-rank matrices, general  $\{x \in \mathcal{E}: h(x) = 0\}$
- Details about computation, pitfalls and tricks E.g.: how to compute gradients, checkgradient, checkhessian
- Theory for general manifolds
- Theory for quotient manifolds
- Discussion of more advanced geometric tools E.g.: distance, exp, log, transports, Lipschitz, finite differences
- Basics of geodesic convexity

AN INTRODUCTION TO OPTIMIZATION ON SMOOTH MANIFOLDS

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSIT