Optimizing till stationarity on the bounded-rank variety

Nicolas Boumal
EPFL, Institute of Mathematics

Ongoing work with Eitan Levin (CalTech) and Joe Kileel (UT Austin)

UCLouvain INMA-from-home, Feb 9, 2021

My focus today: bounded rank constraints

$$\min_{X \in \mathbf{R}^{m \times n}} f(X) \quad \text{subject to} \quad \operatorname{rank}(X) \leq k$$

Many applications (some with additional structure, e.g., $X \succcurlyeq 0$):

- Model order reduction
- Recommender systems
- Sensor network localization
- Large scale linear matrix equations
- ...

My focus today: bounded rank constraints

$$\min_{X \in \mathbf{R}^{m \times n}} f(X) \quad \text{subject to} \quad \operatorname{rank}(X) \leq k$$

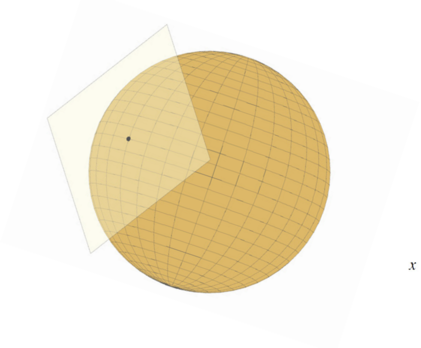
Assume f is smooth (let's say C^∞).

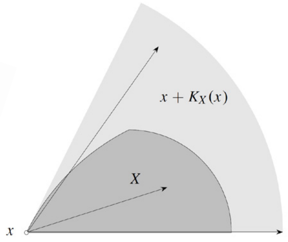
In general, finding a global minimum is NP-hard.*

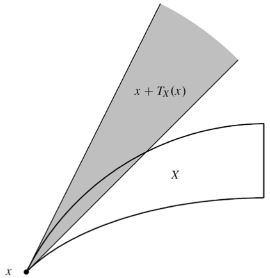
Less ambitious goal of this talk: find a stationary point.

*See for example Gillis & Glineur, *Low-rank matrix approximation with weights or missing data is NP-hard*, SIMAX 2011

Stationarity in general







$$\min_{x \in \mathcal{E}} f(x) \quad \text{subject to} \quad x \in \mathcal{X}$$

The tangent cone $\mathrm{T}_x\mathcal{X}$ collects allowed directions of movement at x.

$$\mathrm{T}_x\mathcal{X} = \left\{ \lim_{i \to \infty} \frac{x_i - x}{\tau_i} : (x_i) \subset \mathcal{X}, \tau_i \subset \mathbf{R}^+, x_i \to x, \tau_i \to 0 \right\}$$

x is stationary if $\mathrm{D}f(x)[v] \geq 0$ for all $v \in \mathrm{T}_x\mathcal{X}$, i.e., $-\nabla f(x) \in (\mathrm{T}_x\mathcal{X})^\circ$.

This is equivalent to the property $\left\| \mathrm{Proj}_{\mathrm{T}_x\mathcal{X}}(-\nabla f(x)) \right\| = 0$.

Gray pictures: Ruszczyński, *Nonlinear Optimization*, 2006

Rank bound or equality: different geometries

The following set is a smooth manifold:

$$\{X \in \mathbf{R}^{m \times n} \colon \operatorname{rank}(X) = k\}$$

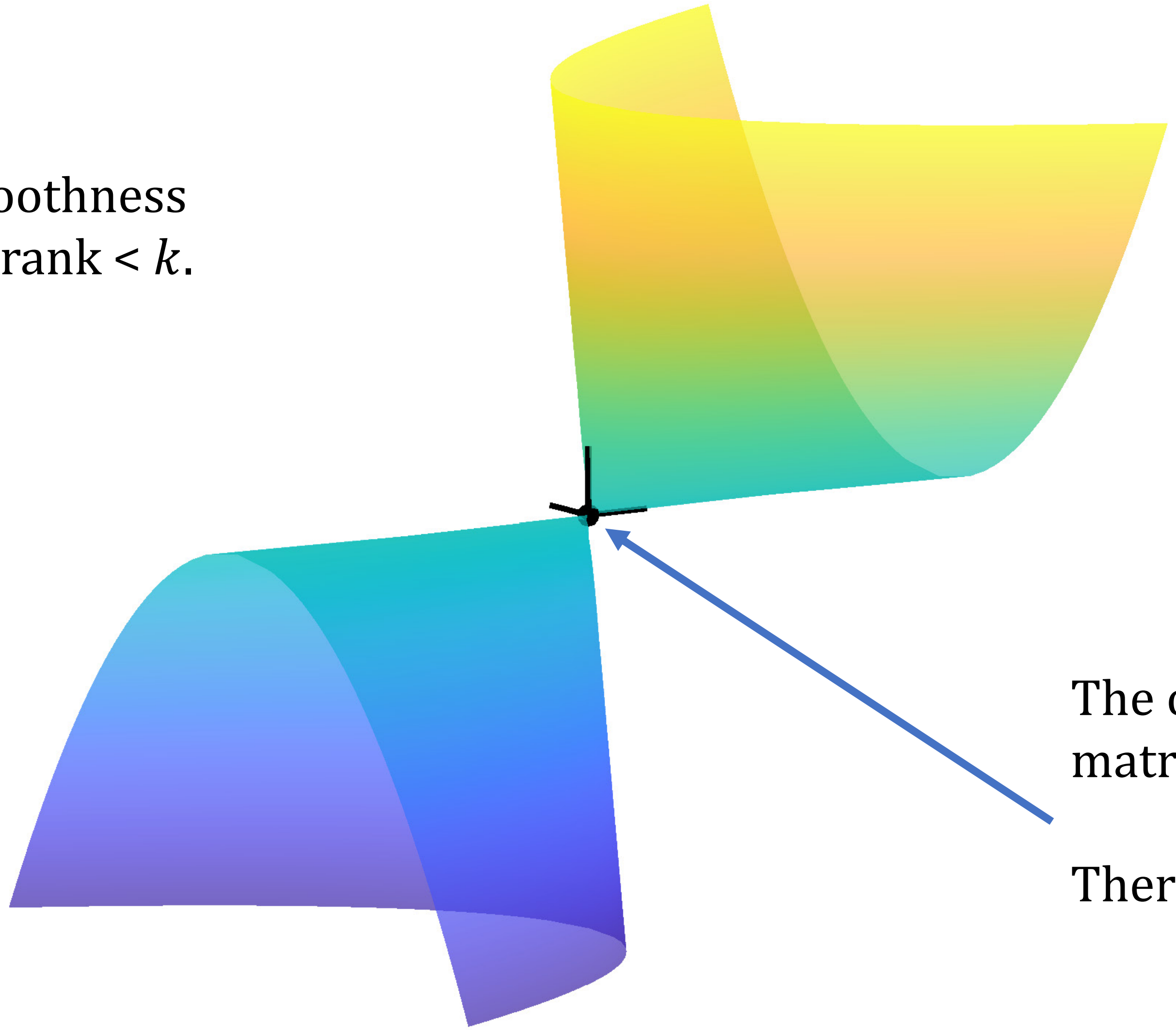
However, the following set is not:

$$\{X \in \mathbf{R}^{m \times n} \colon \operatorname{rank}(X) \leq k\}$$

Let's do a proof by picture for related case of symmetric $X \in \mathbf{R}^{2 \times 2}$.

$$\left\{X \in \mathbf{R}^{2\times 2} : X = X^\top \text{ and } \operatorname{rank}(X) \leq 1\right\} = \left\{\begin{bmatrix} x & y \\ y & z \end{bmatrix} : xz - y^2 = 0\right\}$$

More generally, smoothness fails at all points of rank < k.



The origin is the only matrix of rank zero.

There, the set is not smooth.

Rank constraints: mind the cliff

$$\min_{X \in \mathbf{R}^{m \times n}} f(X) \quad \text{subject to} \quad \operatorname{rank}(X) \leq k$$

If the iterates remain comfortably on the manifold of rank-k, fine.

In practice, this is often the case.

However, if the iterates approach lesser-rank matrices, or worse, converge to one, then smooth optimization theory breaks down.

Optimization algorithms must be able to handle this eventuality.

Even computing stationary points is tricky

$$\min_{X \in \mathbf{R}^{m \times n}} f(X) \quad \text{subject to} \quad \text{rank}(X) \leq k$$

projected gradient descent method*

$$X_{i+1} = \text{Proj}_{\mathbf{R}^{m\times n}_{\leq k}}\left(X_i + \alpha_i \text{Proj}_{\mathrm{T}_{X_i}\mathbf{R}^{m\times n}_{\leq k}}\big(-\nabla f(X_i)\big)\right)$$

*Schneider & Uschmajew, SIOPT 2015,
Convergence Results for Projected Line-Search Methods on Varieties of Low-Rank Matrices Via Łojasiewicz Inequality

Even computing stationary points is tricky

$$\min_{X \in \mathbf{R}^{m \times n}} f(X) \quad \text{subject to} \quad \text{rank}(X) \leq k$$

There exist f and X_0 for which a projected gradient descent method* with Armijo backtracking produces iterates $X_1, X_2, X_3, \ldots$ such that:

1. $\text{rank}(X_i) = k$ for all i,
2. The stationarity measure goes to zero as $i \to \infty$,
3. The sequence converges to a feasible matrix X,
4. Yet the limit X is not stationary. – We might be far from any!

*Schneider & Uschmajew, SIOPT 2015,
Convergence Results for Projected Line-Search Methods on Varieties of Low-Rank Matrices Via Łojasiewicz Inequality

Apocalypses in general (algorithm agnostic)

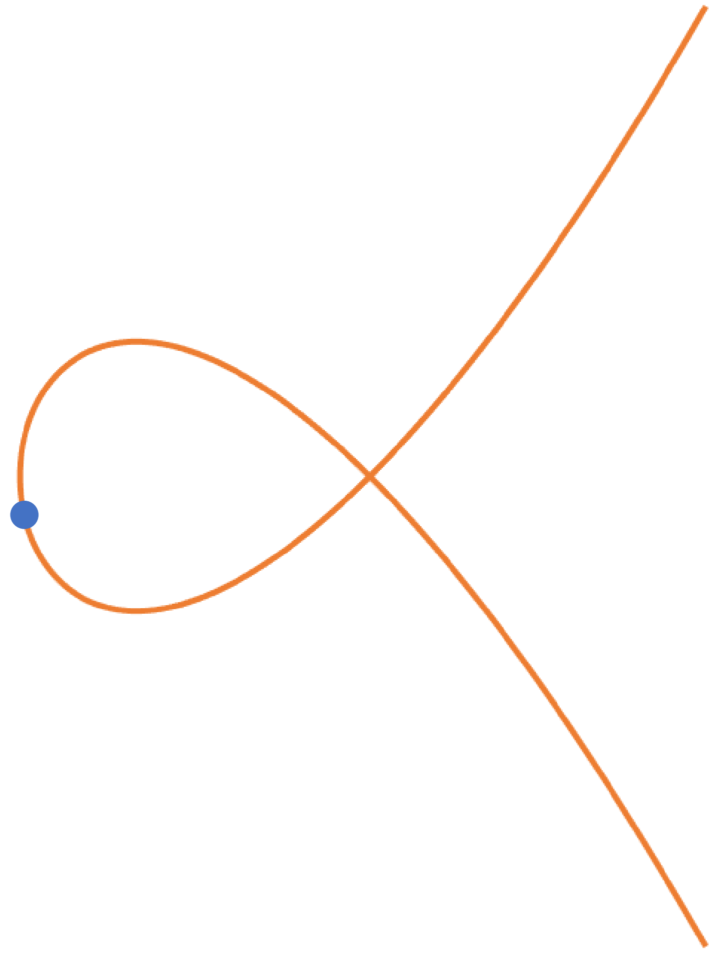
$$\min_{x \in \mathcal{E}} f(x) \quad \text{subject to} \quad x \in \mathcal{X}$$

The tangent cone $\mathrm{T}_x\mathcal{X}$ collects allowed directions of movement at x.

x is stationary if $\left\|\mathrm{Proj}_{\mathrm{T}_x\mathcal{X}}\left(-\nabla f(x)\right)\right\| = 0$.

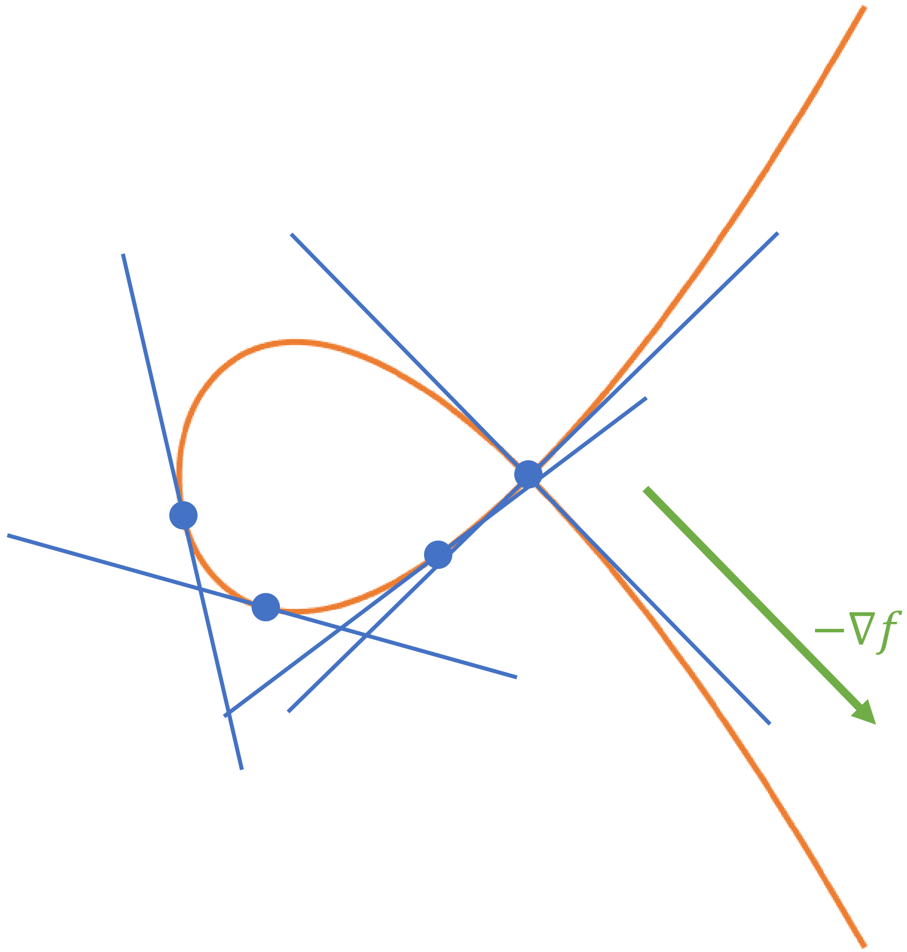
x is apocalyptic if there exists a sequence $x_i \to x$ and a function f such that $\left\|\mathrm{Proj}_{\mathrm{T}_{x_i}\mathcal{X}}\left(-\nabla f(x_i)\right)\right\| \to 0$ yet x is not stationary.

x is apocalyptic if there exists a sequence $x_i \to x$ and a function f such that $\left\|\mathrm{Proj}_{\mathrm{T}_{x_i}\mathcal{X}}\left(-\nabla f(x_i)\right)\right\| \to 0$ yet x is not stationary.



When can apocalypses occur?

x is apocalyptic if there exists a sequence $x_i \to x$ and a function f such that $\left\| \mathrm{Proj}_{\mathrm{T}_{x_i}\mathcal{X}}\left(-\nabla f(x_i)\right) \right\| \to 0$ yet x is not stationary.



When can apocalypses occur?

When tangent cones change suddenly, adding new directions:

$$\left(\limsup_{i\to\infty} \mathrm{T}_{x_i}\mathcal{X}\right)^\circ \not\subseteq (\mathrm{T}_x\mathcal{X})^\circ$$

x is apocalyptic if there exists a sequence $x_i \to x$ and a function f such that $\left\|\mathrm{Proj}_{\mathrm{T}_{x_i}\mathcal{X}}\left(-\nabla f(x_i)\right)\right\| \to 0$ yet x is not stationary.

When can apocalypses occur?

When tangent cones change suddenly, adding new directions:

$$\left(\limsup_{i\to\infty} \mathrm{T}_{x_i}\mathcal{X}\right)^\circ \not\subseteq (\mathrm{T}_x\mathcal{X})^\circ$$

x is apocalyptic if there exists a sequence $x_i \to x$ and a function f such that $\left\|\mathrm{Proj}_{\mathrm{T}_{x_i}\mathcal{X}}\left(-\nabla f(x_i)\right)\right\| \to 0$ yet x is not stationary.

A first take-away, and two positive notes

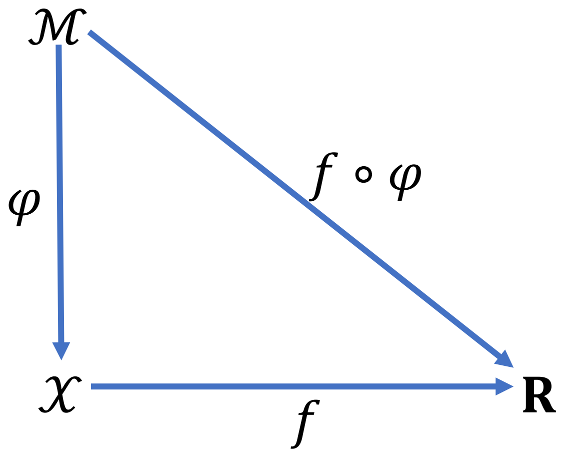
Apocalypses are a geometric feature of a search space $\mathcal{X}$.
They cause blind spots for gradient-based methods.

Two positive notes:

Convex sets and manifolds with boundaries have no apocalypses.

Using second-order information of f, we can find stationary points.

To find stationary points, use lifts



Let $\mathcal{M} = \mathbf{R}^{m\times k} \times \mathbf{R}^{n\times k}$ and $\mathcal{E} = \mathbf{R}^{m\times n}$.
Consider the smooth map $\varphi(L, R) = LR^\top$ from $\mathcal{M}$ to $\mathcal{E}$.

Notice: $\varphi(\mathcal{M}) = \{X \in \mathbf{R}^{m\times n}: \text{rank}(X) \leq k\}$: it is a smooth lift.

Thm*: If (L, R) is 2-critical for $f \circ \varphi$, then $LR^\top$ is stationary for f.

Thm: If f has compact sublevel sets, then a modified version of the trust-region method on $f \circ \varphi$ finds 2-critical points, always.

* Ha, Liu & Barber, SIOPT 2021,
An equivalence between critical points for rank constraints versus low-rank factorizations

Summary

Optimization on nonsmooth sets: watch out for apocalypses.

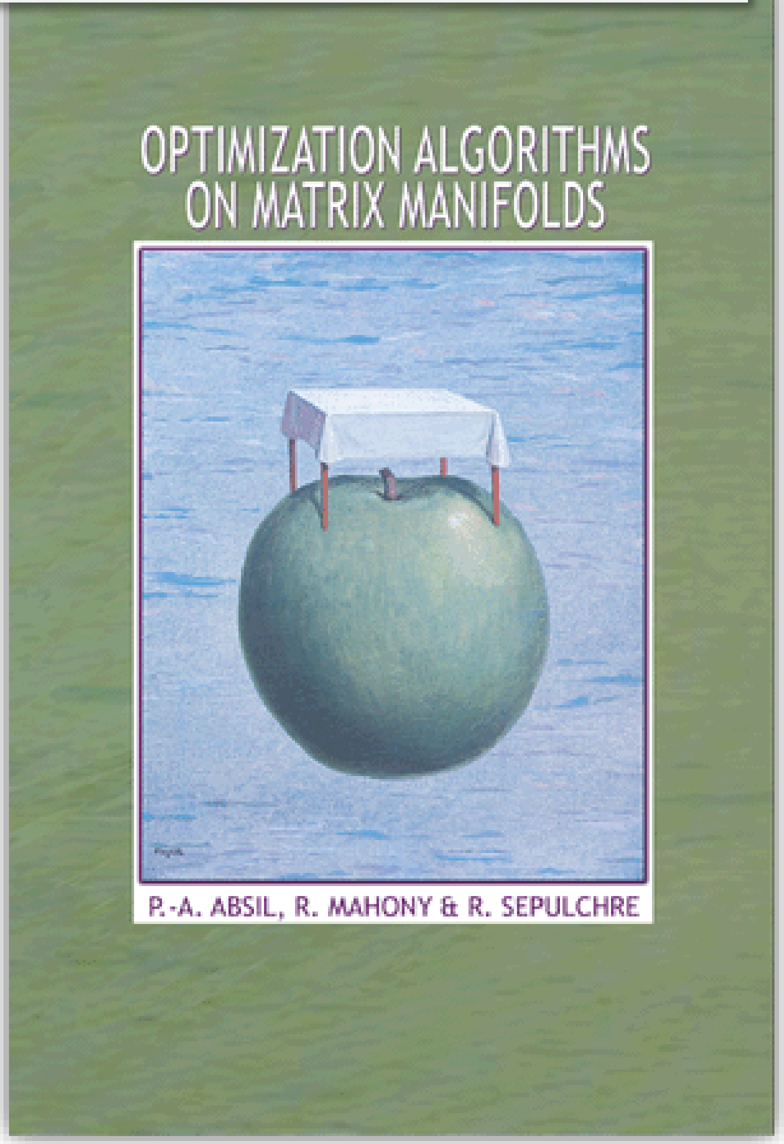
Exist on bounded-rank variety; not on convex sets / manifolds with boundary.

We can use lifts to move the problem to a smooth manifold.

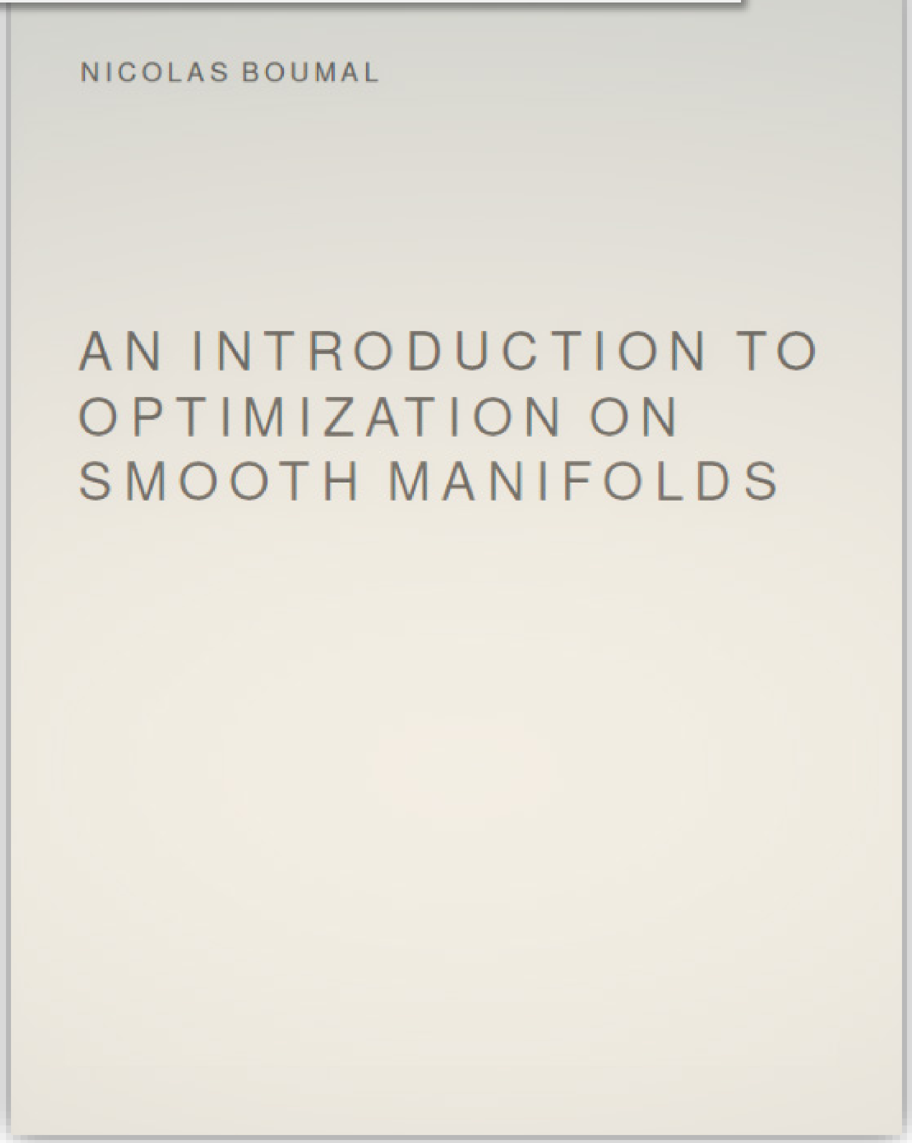
Can be done for other nonsmooth sets: desingularization, symmetry, shadows...

If the lift has nice properties (e.g., 2-critical $\mapsto$ stationary),
it tells us how to use the Hessian of f to avoid apocalypses.

press.princeton.edu/absil



nicolasboumal.net/book



press.princeton.edu/absil

nicolasboumal.net/book

manopt.org

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

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

Lead by Ronny Bergmann