Apocalypse on singular sets: optimization with bounded rank

Nicolas Boumal

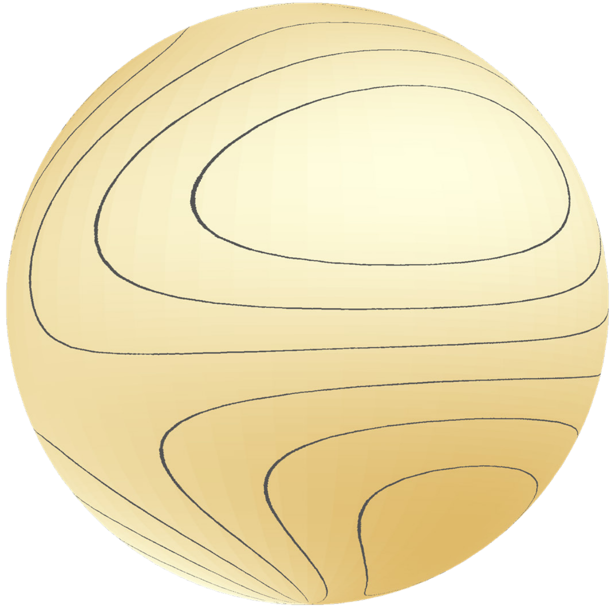
EPFL, Institute of Mathematics

Ongoing work with Eitan Levin and Joe Kileel

ManLearn workshop, ICPR-from-home, Jan 11, 2021

Optimization on smooth manifolds

$$\min_x f(x) \text{ subject to } x \in \mathcal{M}$$



Linear spaces

Unconstrained; linear equality constraints

Fixed-rank matrices, tensors

Recommender systems, large-scale Lyapunov equations, …

Orthonormal matrices (sphere, Stiefel, rotations, ...)

Dictionary learning, SfM, SLAM, PCA, ICA, SBM, Electr. Struct. Comp.…

Positive definite matrices, positive vectors

Metric learning, Gaussian mixtures, diffusion tensor imaging, …

Quotients through symmetries

Invariance under group actions

press.princeton.edu/absil

nicolasboumal.net/book

manopt.org

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

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

Lead by Ronny Bergmann

manopt.org
nicolasboumal.net/book

My talk today however is about **nonsmooth** sets.

... though we'll see smoothness is never far.

A classical geometric fact first

The following set is a smooth manifold:

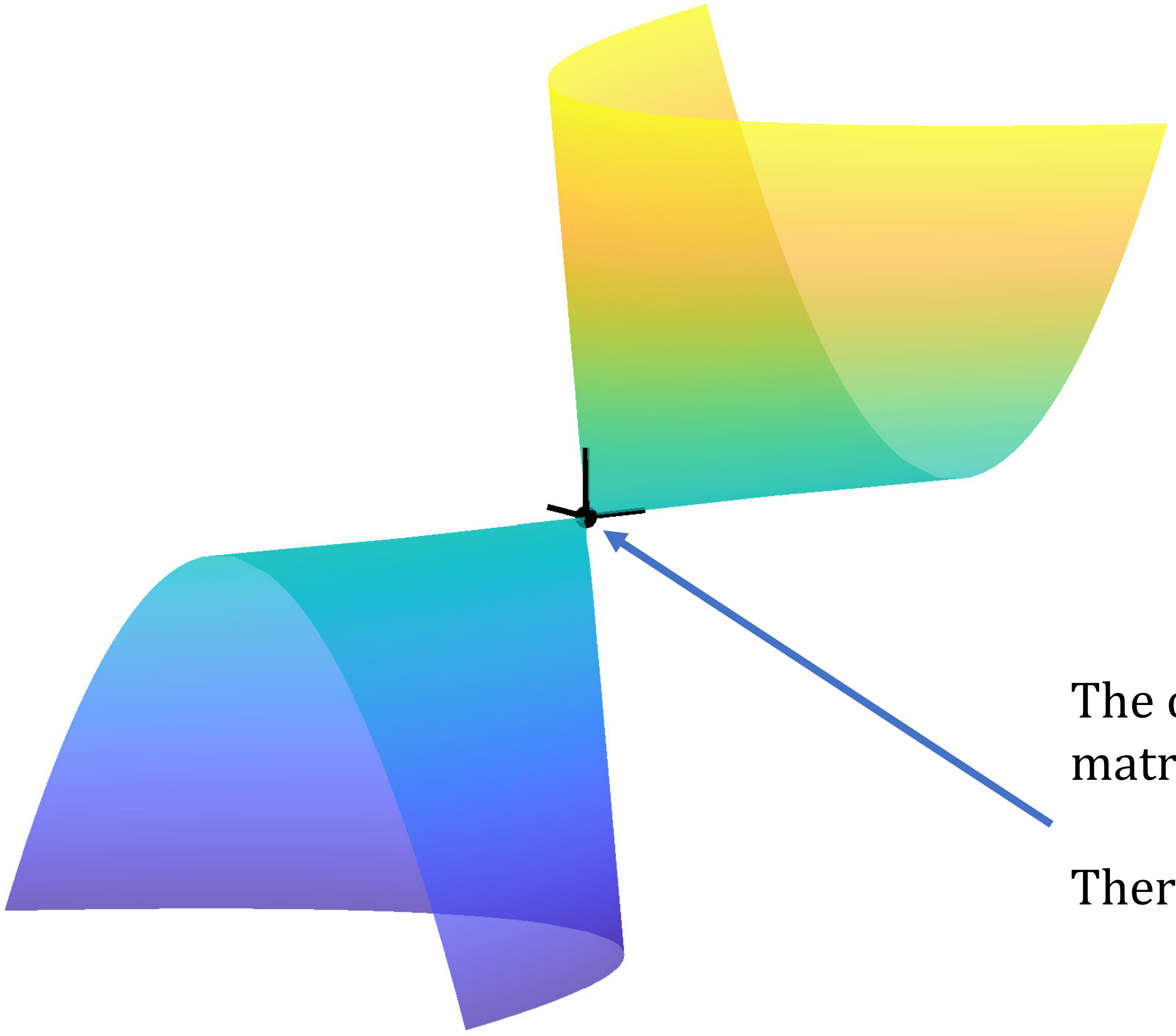
$$\{X \in \mathbf{R}^{2\times 2} : X = X^\top \text{ and } \operatorname{rank}(X) = 1\}$$

However, the following set is not a smooth manifold:

$$\{X \in \mathbf{R}^{2\times 2} : X = X^\top \text{ and } \operatorname{rank}(X) \leq 1\}$$

Let’s do a proof by picture.

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



The origin is the only matrix of rank zero.

There, the set is not smooth.

Optimization under rank constraints

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

Fixed-rank matrices form a smooth embedded submanifold of $\mathbf{R}^{m \times n}$.

However, this is not a closed set.

Issue for optimization: sequences might “converge” outside the manifold, at which point we lose all control.

Optimization under rank constraints

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

Closure: bounded-rank matrices form an algebraic variety in $\mathbf{R}^{m \times n}$.

Issue for optimization: this is no longer smooth.

If iterates remain comfortably on rank-k manifold, fine.
But if they converge to lesser-rank matrices, bad things can happen.

Even computing stationary points is tricky

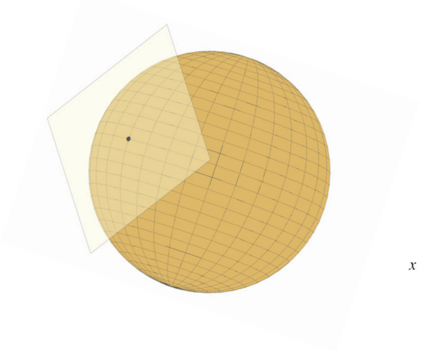
$$\min_{X \in \mathbf{R}^{m \times n}} f(X) \quad \text{subject to} \quad \text{rank}(X) \leq k$$

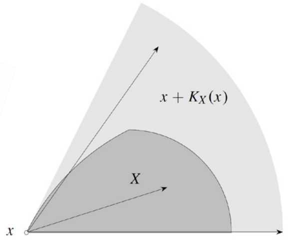
There exist f and X_0 for which a projected gradient descent method* with Armijo backtracking produces iterates $X_1, X_2, X_3, \ldots$ such that:

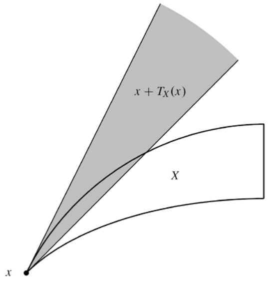
1. $\text{rank}(X_n) = k$ for all n,
2. Some stationarity measure goes to zero as $n \to \infty$,
3. The sequence converges to a feasible matrix X,
4. Yet the limit X is not stationary.

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

Apocalypses 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.

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

Gray pictures: Ruszczyński, *Nonlinear Optimization*, 2006

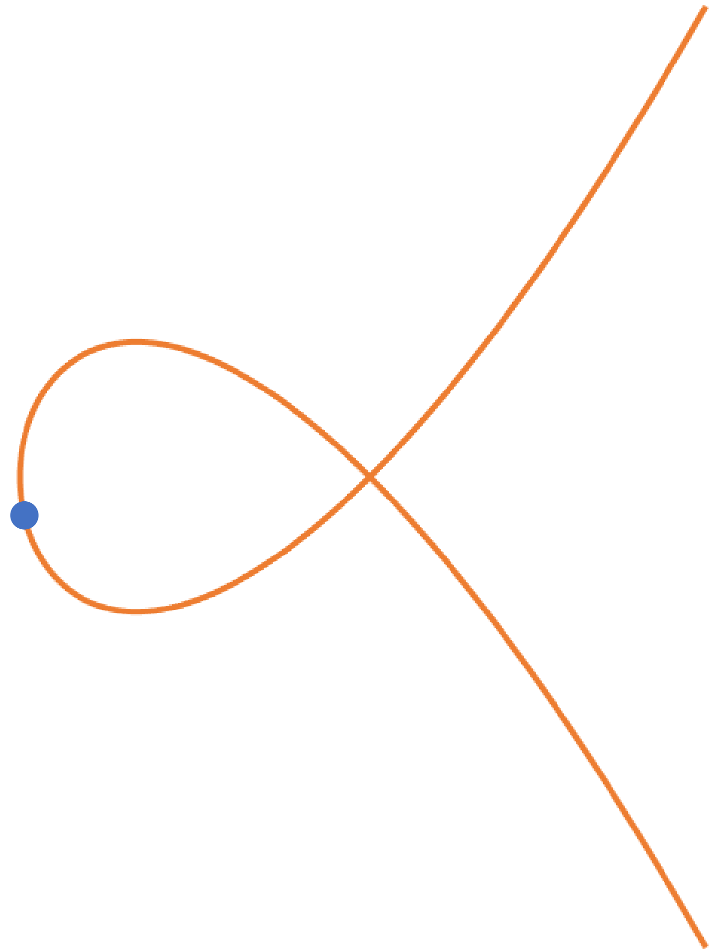
Apocalypses in general

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

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

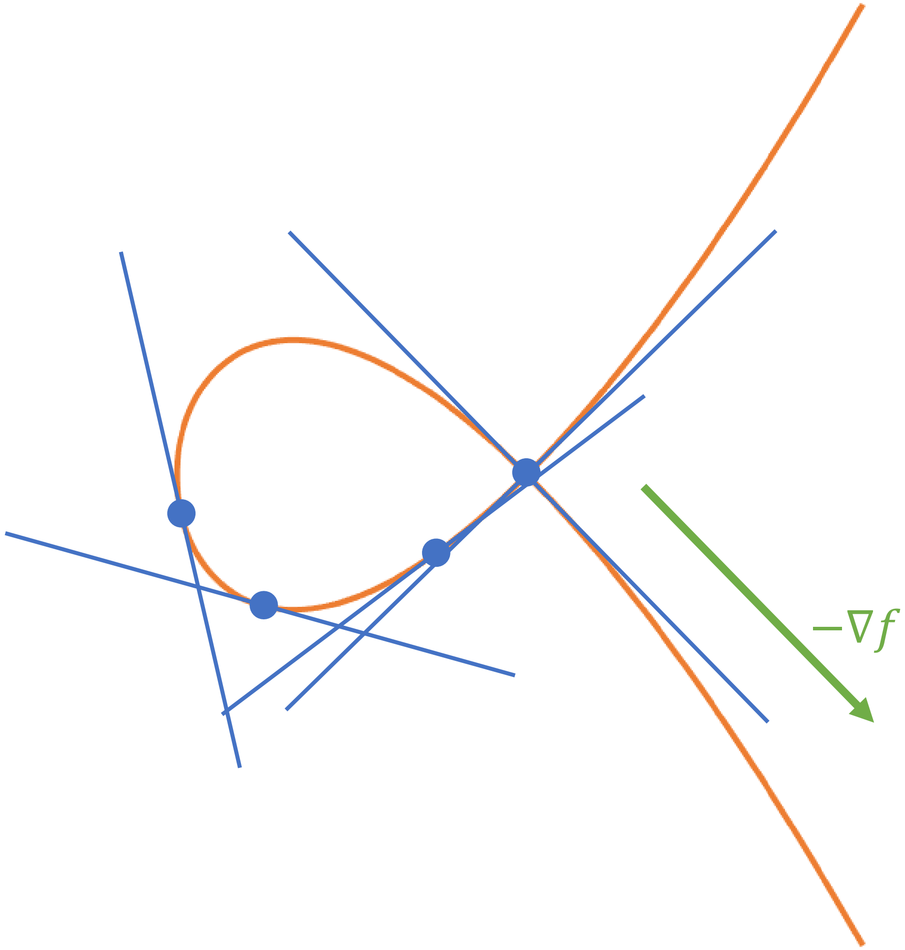
x is stationary if $\left\|\mathrm{Proj}_{\mathrm{T}_x\mathcal{X}}\left(-\nabla f(x)\right)\right\| = 0$.

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.



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.

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.

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 positives notes

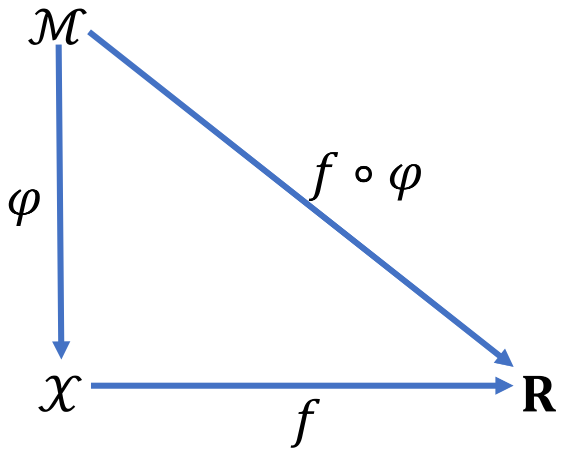
To our knowledge, no optimization algorithm iterating on the bounded-rank variety guarantees convergence to stationary points in all cases, even under generous assumptions on f.

We believe apocalypses are the main obstacle to that goal.

Two positive notes:

1. There are no apocalypses on convex sets nor on manifolds with boundaries.
2. We can lift bounded-rank matrices to a smooth manifold and work there.

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

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

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

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

Summary

Optimization on non-smooth sets can be tricky due to apocalypses.

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

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

This can be done for many other nonsmooth sets: much to explore here.

If the lift has nice properties (e.g., 2-critical $\mapsto$ stationary),
this can help use converge to stationary points with certainty.