NTNU

Norges teknisk-naturvitenskapelige universitet

Nonsmooth Optimization on Riemannian manifolds

Ronny Bergmann

joint work with

O. P. Ferreira, R. Herzog, H. Jasa, E. M. Santos, and J. C. O. Souza.

Kolloquium über Angewandte Mathematik

Institut für Numerische und Angewandte Mathematik,

Göttingen, June 11, 2024.

Nonsmooth Optimization on Riemannian Manifolds

We are looking for numerical algorithms to find

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} f(p)$$

where

- $\mathcal{M}$ is a Riemannian manifold
- $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ is a function

⚠ f might be nonsmooth and/or nonconvex

⚠ $\mathcal{M}$ might be high-dimensional

A Riemannian Manifold $\mathcal{M}$

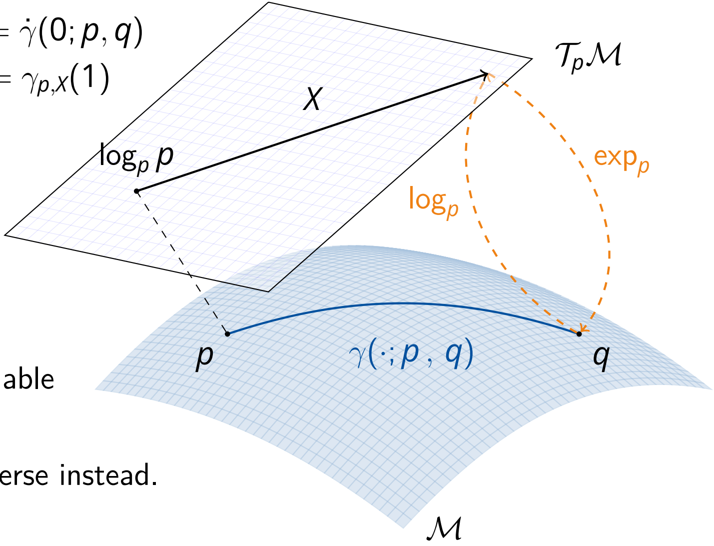
A d-dimensional Riemannian manifold can be informally defined as a set $\mathcal{M}$ covered with a "suitable" collection of charts, that identify subsets of $\mathcal{M}$ with open subsets of $\mathbb{R}^d$ and a continuously varying inner product on the tangent spaces.

[Absil, Mahony, and Sepulchre 2008]

A Riemannian Manifold $\mathcal{M}$

Notation.

- Logarithmic map $\log_p q = \dot\gamma(0; p, q)$
- Exponential map $\exp_p X = \gamma_{p,X}(1)$
- Geodesic $\gamma(\cdot; p\,,\,q)$
- Tangent space $\mathcal{T}_p\mathcal{M}$
- inner product $(\cdot\,,\,\cdot)_p$



Numerics.

$\exp_p$ and $\log_p$ maybe not available efficiently/ in closed form

$\Rightarrow$ use a retraction and its inverse instead.

(Geodesic) Convexity

[Sakai 1996; Udriște 1994]

A set $\mathcal{C} \subset \mathcal{M}$ is called (strongly geodesically) convex
if for all $p, q \in \mathcal{C}$ the geodesic $\gamma(\cdot; p\,, q)$ is unique and lies in $\mathcal{C}$.

A function $f\colon \mathcal{C} \to \overline{\mathbb{R}}$ is called (geodesically) convex
if for all $p, q \in \mathcal{C}$ the composition $f(\gamma(t; p\,, q)), t \in [0, 1]$, is convex.

The Riemannian Subdifferential

Let $\mathcal{C}$ be a convex set.

The subdifferential of f at $p \in \mathcal{C}$ is given by [O. Ferreira and Oliveira 2002; Lee 2003; Udriște 1994]

$$\partial_{\mathcal{M}} f(p) := \Bigl\{ \xi \in \mathcal{T}_p^* \mathcal{M} \,\Big|\, f(q) \geq f(p) + \langle \xi \,, \log_p q \rangle_p \ \text{ for } q \in \mathcal{C} \Bigr\},$$

where

- $\mathcal{T}_p^*\mathcal{M}$ is the dual space of $\mathcal{T}_p\mathcal{M}$, also called cotangent space
- $\langle \cdot \,, \cdot \rangle_p$ denotes the duality pairing on $\mathcal{T}_p^*\mathcal{M} \times \mathcal{T}_p\mathcal{M}$

NTNU

The Riemannian Convex Bundle Method

The ε-Subdifferential

Let $\varepsilon > 0$.
The $\varepsilon-$subdifferential of a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ reads

$$\partial_\varepsilon f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + s^\mathrm{T}(y - x) - \varepsilon \text{ for all } y \in \mathbb{R}^n \right\}$$

Let $\varepsilon > 0$ and $\mathcal{C} \subset \mathcal{M}$ be a convex set.
The $\varepsilon-$subdifferential of a convex function $f\colon \mathcal{C} \to \mathbb{R}$ reads

$$\partial_\varepsilon f(x) = \left\{ X \in \mathcal{T}_p\mathcal{M} \,\middle|\, f(q) \geq f(p) + (X, \log_p q) - \varepsilon \text{ for all } q \in \mathcal{C} \right\}$$

Clearly in both cases $\partial f(x) = \partial_0 f(x) \subset \partial_\varepsilon f(x)$

The Riemannian Convex Bundle Method

[RB, Herzog, and Jasa 2024]

- Given $f\colon \mathcal{C} \to \mathbb{R}$ on a (geodesically) convex set $\mathcal{C} \subset \mathcal{M}$
- collect
  - subgradients $X_{q^{(k)}} \in \partial f(q^{(k)})$
  - stabilisation centers $p^{(k)}$ ("best" iterates)
- use this information to
  - determine the next descent direction $d^{(k)} \in \mathcal{T}_{p^{(k)}}\mathcal{M}$ by solving a QP in $\mathcal{T}_{p^{(k)}}\mathcal{M}$
  - where $d^{(k)} \in \partial_{c^{(k)}} f(p^{(k)})$
- we stop when both
  - the approximation $\partial_{c^{(k)}} f(p^{(k)})$ of $\partial f(p^{(k)})$ is "good enough"
  - $\|d^{(k)}\|$ is "small enough"

Aproximating the ε-Subdifferential

For $f\colon \mathbb{R}^n \to \mathbb{R}$, given $x^{(0)}, \ldots, x^{(k)} \in \mathbb{R}^n$, and $s^{(j)} \in \partial f(x^{(j)})$,
define the linearization errors

$$e_j^{(k)} := f(x^{(k)}) - f(x^{(j)}) - (s^{(j)})^T(x^{(k)} - x^{(j)}), \qquad j = 0, \ldots, k.$$

Then (Geiger and Kanzow 2002, Theorem 6.68)

$$s^{(j)} \in \partial_{e_j^{(k)}} f(x^{(k)})$$

and we can characterize an inner approximation $G_\varepsilon^{(k)} \subseteq \partial_\varepsilon f(x^{(k)})$ as

$$G_\varepsilon^{(k)} := \left\{ \sum_{j=0}^{k} \lambda_j s^{(j)} \;\middle|\; \sum_{j=0}^{k} \lambda_j\, e_j^{(k)} \leq \varepsilon,\ \sum_{j=0}^{k} \lambda_j = 1,\ \lambda_j \geq 0 \text{ for all } j = 0, \ldots, k \right\}$$

Challenge on manifolds.
How can we take into account curvature in the error terms?

Curvature Correction

Let $\Omega \in \mathbb{R}$ be an upper bound on the curvature. Define

[RB, Herzog, and Jasa 2024]

$$c_j^{(k)} := f(p^{(k)}) - f(p^{(j)}) - \left(X_{p^{(j)}} \,,\, \log_{p^{(j)}} p^{(k)} \right) \quad \text{if } \Omega \leq 0,$$
$$c_j^{(k)} := f(p^{(k)}) - f(p^{(j)}) + \|X_{p^{(j)}}\| \|\log_{p^{(j)}} p^{(k)}\| \quad \text{if } \Omega > 0.$$

Then we get

$$G_\varepsilon^{(k)} := \left\{ \sum_{j=0}^{k} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow p^{(j)}} X_{p^{(j)}} \,\middle|\, \sum_{j=0}^{k} \lambda_j \, e_j^{(k)} \leq \varepsilon, \ \sum_{j=0}^{k} \lambda_j = 1, \ \lambda_j \geq 0, j = 0, \ldots, k \right\}$$

with $G_\varepsilon^{(k)} \subseteq \partial_\varepsilon f(p^{(k)})$, and $\mathrm{P}_{p^{(k)} \leftarrow p^{(j)}} X_{p^{(j)}} \in \partial_{c_j^{(k)}} f(p^{(k)})$.

The Riemannian Subproblem

Let $k \in \mathbb{N}$ and $j \in \{0, \dots, k\} = J^{(k)}$ and $X_{p^{(j)}} \in \partial f(p^{(j)})$, $p^{(j)} \in \mathbb{R}^n$

For a coefficients $\lambda_j \geq 0$ with $\sum_j \lambda_j = 1$, we have

$$\sum_{j \in J^{(k)}} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}} \in \partial_\varepsilon f(p^{(k)}) \qquad \text{if and only if} \qquad \sum_{j \in J^{(k)}} \lambda_j c_j^{(k)} \leq \varepsilon$$

Solving the constrained quadratic problem

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J^{(k)}|}} \quad & \frac{1}{2} \Bigl\lVert \sum_{j \in J^{(k)}} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}} \Bigr\rVert^2 + \sum_{j \in J^{(k)}} \lambda_j e_j^{(k)} c_j^{(k)} \\ \text{s. t.} \quad & \sum_{j \in J^{(k)}} \lambda_j = 1, \quad \lambda_j \geq 0 \quad \text{for all } j \in J^{(k)} \end{aligned}$$

yields the new search direction

$$d^{(k)} := - \sum_{j \in J^{(k)}} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}}.$$

The Riemannian Convex Bundle Method

Input: $p^{(0)} = q^{(0)} \in \mathcal{C}$, $g^{(0)} = X_{p_0} \in \partial f(p^{(0)})$, $m \in (0,1)$,
$\varepsilon^{(0)} = e^{(0)} c^{(0)} = 0$, $J^{(0)} = \{0\}$, and $k = 0$.

1: **while** not converged **do**
2: Set $k = k + 1$
3: Compute a solution $\lambda^{(k)} \in \mathbb{R}^{|J^{(k)}|}$ of the subproblem.
4: Set
$$g^{(k)} := \sum_{j \in J^{(k)}} \lambda_j^{(k)} \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}}, \qquad \varepsilon^{(k)} := \sum_{j \in J^{(k)}} \lambda_j^{(k)} e_j^{(k)} c_j^{(k)},$$
$$d^{(k)} := -g^{(k)}, \qquad \xi^{(k)} := -\|g^{(k)}\|^2 - \varepsilon^{(k)},$$
5: Set $q^{(k+1)} = \exp_{p^{(k)}} d^{(k)}$ [1] and take $X_{q^{(k+1)}} \in \partial f(q^{(k+1)})$,
6: If $f(q^{(k+1)}) \leq f(p^{(k)}) + m\xi^{(k)}$ set $p^{(k+1)} = q^{(k+1)}$ else $p^{(k+1)} = p^{(k)}$
7: Update $J^{(k+1)} = \{j \in J^{(k)} \mid \lambda_j^{(k)} > 0\} \cup \{k+1\}$, and $c_j^{(k+1)}$
8: **end while**

Output: $p^{(k_*)}$ from the final $k_* \in \mathbb{N}$.

[1] Perform a backtracking if $q^{(k+1)} \notin \operatorname{int}(\operatorname{dom} f)$ or equal to $p^{(k)}$

Convergence

Theorem (Geiger and Kanzow 2002, Theorem 6.80)
Let the solution set $S = \{x^* \in \mathbb{R}^n \,|\, f(x^*) = \inf_{x\in\mathbb{R}^n} f(x)\}$ of the minimization problem be nonempty. Then every sequence $\{x^{(k)}\}$ generated by the bundle method algorithm converges to a minimizer of f.

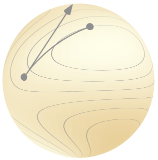
On Hadamard manifolds ($\Omega \leq 0$) we have the analogous, if

[RB, Herzog, and Jasa 2024]

1. the backtracking step size $t^{(k)} > m$ for all $k \geq k_*$, if a finite number of serious steps k_* occur
2. no accumulation point of $p^{(k)}$ is allowed to lie on the boundary of $\mathcal{C}$

Numerical Examples

Manopt.jl



Goal. Provide optimization algorithms on Riemannian manifolds.

Features. Given a `Problem p` and a `SolverState s`,
implement `initialize_solver!(p, s)` and `step_solver!(p, s, i)`
$\Rightarrow$ an algorithm in the `Manopt.jl` interface

Highlevel interfaces like `gradient_descent(M, f, grad_f)`
on any manifold `M` from `Manifolds.jl`.

All provide debug output, recording, cache & counting capabilities,
as well as a library of step sizes and stopping criteria.

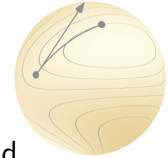
Manopt family.

manoptjl.org [RB 2022]

manopt.org [Boumal, Mishra, Absil, and Sepulchre 2014]

pymanopt.org [Townsend, Koep, and Weichwald 2016]

List of Algorithms in Manopt.jl



Derivatve Free Nelder-Mead, Particle Swarm, CMA-ES

Subgradient-based Subgradient Method, Convex Bundle Method, Proximal Bundle Method

Gradient-based Gradient Descent, Conjugate Gradient, Stochastic, Momentum, Nesterov, Averaged, ...
Quasi-Newton with (L-)BFGS, DFP, Broyden, SR1,...
Levenberg-Marquard

Hessian-based Trust Regions, Adaptive Regularized Cubics (ARC)

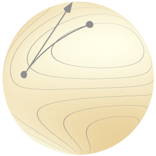
nonsmooth Chambolle-Pock, Douglas-Rachford, Cyclic Proximal Point

constrained Augmented Lagrangian, Exact Penalty, Frank-Wolfe

nonconvex Difference of Convex Algorithm, DCPPA

manoptjl.org/stable/solvers/

The Convex Bundle Method in Manopt.jl



In `Manopt.jl` a solver call looks like[2]

```
p = convex_bundle_method(M, f, ∂f, p0;
        diameter = δ, k_max = Ω, m = 10⁻³, kwargs...
    )
```

where

- `M` is a Riemannian manifold
- `f` is the objective function
- `∂f` is a subgradient of the objective function
- `p0` is an initial point on the manifold

The default stopping criterion for the algorithm is set to

$$-\xi^{(k)} \leq 10^{-8}.$$

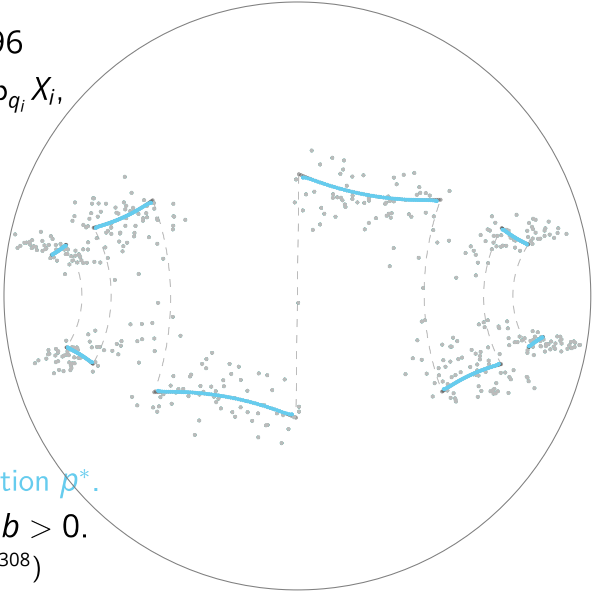
[2] full documentation: manoptjl.org/stable/solvers/convex_bundle_method/

Denoising a Signal on Hyperbolic Space $\mathcal{H}^2$

- signal $q \in \mathcal{M}$, $(\mathcal{H}^2)^n$, $n = 496$
- noisy signal $\bar{q} \in \mathcal{M}$, $\bar{q}_i = \exp_{q_i} X_i$, $\sigma = 0.1$
- ROF Model:

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} \frac{1}{n} \mathrm{d}_{\mathcal{M}}(p,q)^2 + \alpha \sum_{i=1}^{n-1} \mathrm{d}_{\mathcal{H}^2}(p_i, p_{i+1})$$

- Setting $\alpha = 0.05$ yields reconstruction p^*.
- in RCBM: set $\operatorname{diam}(\operatorname{dom} f) = b > 0$.
  (in practice: $b =$`floatmax()`$\approx 10^{308}$)



Algorithms for Denoising a Signal

- Riemannian Convex Bundle Method (RCBM) [RB, Herzog, and Jasa 2024]
- Proximal Bundle Algorithm (PBA) [Hoseini Monjezi, Nobakhtian, and Pouryayevali 2021]
- Subgradient Method (SGM) [O. Ferreira and Oliveira 1998]
- Cyclic Proximal Point Algorithm (CPPA) [Bačák 2014]

Algorithm	Iter.	Time (sec.)	Objective	Error
RCBM	3417	51.393	1.7929×10^{-3}	3.3194×10^{-4}
PBA	15 000	102.387	1.8153×10^{-3}	4.3874×10^{-4}
SGM	15 000	99.604	1.7920×10^{-3}	3.3080×10^{-4}
CPPA	15 000	94.200	1.7928×10^{-3}	3.3230×10^{-4}

The Riemanniann Median on $\mathcal{S}^d$

- Consider the $d-$dimensional sphere $\mathcal{M} = \mathcal{S}^d$
- $\bar{p}$ north pole
- $B_r(p)$ (geodesic) ball around p with radius r.
- $n = 1000$ Gaussian random data points $q^{(1)}, \ldots, q^{(n)} \in B_{\frac{\pi}{8}}(\bar{p})$
- Riemannian median on $B_{\frac{\pi}{8}}(\bar{p})$:

$$f(p) = \begin{cases} \frac{1}{n}\sum_{j=1}^{n} d_{\mathcal{M}}(p, q^{(j)}) & \text{if } p \in B_{\frac{\pi}{8}}(\bar{p}), \\ +\infty & \text{otherwise.} \end{cases}$$

Solve

$$p^* := \operatorname*{arg\,min}_{p \in \mathcal{S}^d} f(p)$$

for different manifold-dimensions d.

Algorithms for the Riemanniann Median on $\mathcal{S}^d$

	RCBM			PBA		
Dimension	Iter.	Time (sec.)	Objective	Iter.	Time (sec.)	Objective
2	19	6.50×10^{-3}	0.192 89	20	5.30×10^{-3}	0.192 89
4	28	1.01×10^{-2}	0.198 81	23	5.99×10^{-3}	0.198 81
32	58	2.29×10^{-2}	0.195 76	28	1.13×10^{-2}	0.195 76
1024	48	3.91×10^{-1}	0.197 75	40	3.31×10^{-1}	0.197 75
32 768	43	7.54	0.192 90	21	4.16	0.192 90

	SGM		
Dimension	Iter.	Time (sec.)	Objective
2	5000	1.14	0.192 89
4	3270	8.09×10^{-1}	0.198 81
32	5000	2.18	0.195 76
1024	122	9.75×10^{-1}	0.197 75
32 768	172	5.25×10^{1}	0.192 90

The Riemannian Difference of Convex Algorithm

Difference of Convex

We aim to solve

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} f(p)$$

where

- $\mathcal{M}$ is a Riemannian manifold
- $f\colon \mathcal{M} \to \mathbb{R}$ is a difference of convex function, i. e. of the form

$$f(p) = g(p) - h(p)$$

- $g, h\colon \mathcal{M} \to \overline{\mathbb{R}}$ are convex, lower semicontinuous, and proper

The Euclidean DCA

Idea 1. At x_k, approximate $h(x)$ by its affine minorization

$$h_k(x) \coloneqq h(x^{(k)}) + \langle x - x^{(k)}, y^{(k)} \rangle \text{ for some } y^{(k)} \in \partial h(x^k)$$

$$\Rightarrow \text{ iteratively minimize } g(x) - h_k(x) = g(x) - h(x^{(k)}) - \langle x - x^{(k)}, y^{(k)} \rangle$$

Idea 2. Using duality theory finding a new $y^{(k)} \in \partial h(x^{(k)})$ is equivalent to

$$y^{(k)} \in \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \Bigl\{ h^*(y) - g^*(y^{(k-1)}) - \langle y - y^{(k-1)}, x^{(k)} \rangle \Bigr\}$$

Idea 3. Reformulate 2 using a proximal map $\Rightarrow$ DCPPA
on manifolds this was done in [Almeida, Neto, Oliveira, and Souza 2020; Souza and Oliveira 2015]

In the Euclidean case, all three models are equivalent.

A Fenchel Duality on a Hadamard Manifold

Let

- $T\mathcal{M} = \dot{\bigcup}_p T_p\mathcal{M}$ denote the tangent bundle
- analogously $T^*\mathcal{M}$ denotes the cotangent bundle
- $\mathcal{M}$ be a Hadamard manifold (non-positive sectional curvature).

Definition [Silva Louzeiro, RB, and Herzog 2022]

Let $f\colon \mathcal{M} \to \overline{\mathbb{R}}$.

The Fenchel conjugate of f is the function $f^*\colon \mathcal{T}^*\mathcal{M} \to \overline{\mathbb{R}}$ defined by

$$f^*(p,\xi) := \sup_{q\in\mathcal{M}} \Bigl\{ \langle \xi, \log_p q \rangle - f(q) \Bigr\}, \qquad (p,\xi) \in \mathcal{T}^*\mathcal{M}.$$

The Dual Difference of Convex Problem

Given the Difference of Convex problem

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} g(p) - h(p)$$

and the Fenchel duals g^* and h^*,
we can state the dual difference of convex problem as

[RB, O. P. Ferreira, Santos, and Souza 2024]

$$\operatorname*{arg\,min}_{(p,\xi)\in T^*\mathcal{M}} \; h^*(p,\xi) - g^*(p,\xi).$$

On $\mathcal{M} = \mathbb{R}^n$ this indeed simplifies to the classical dual problem.

Theorem. [RB, O. P. Ferreira, Santos, and Souza 2024]

$$\inf_{(q,X)\in\mathcal{T}^*\mathcal{M}} \Bigl\{h^*(q,X) - g^*(q,X)\Bigr\} = \inf_{p\in\mathcal{M}} \{g(p) - h(p)\}\,.$$

NTNU

The Dual Difference of Convex Problem

The primal and dual Difference of Convex problem

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} g(p) - h(p) \qquad \text{and} \qquad \operatorname*{arg\,min}_{(p,\xi)\in T^*\mathcal{M}} h^*(p,\xi) - g^*(p,\xi)$$

are equivalent in the following sense.

Theorem. [RB, O. P. Ferreira, Santos, and Souza 2024]

If p^* is a solution of the primal problem, then $(p^*, \xi^*) \in T^*\mathcal{M}$ is a solution for the dual problem for all $\xi^* \in \partial_{\mathcal{M}} h(p^*) \cap \partial_{\mathcal{M}} g(p^*)$.

If $(p^*, \xi^*) \in T^*\mathcal{M}$ is a solution of the dual problem for some $\xi^* \in \partial_{\mathcal{M}} h(p^*) \cap \partial_{\mathcal{M}} g(p^*)$, then p^* is a solution of the primal problem.

Derivation of the Riemannian DCA

We consider the first order Taylor approximation of h at some point $p^{(k)}$:
With $\xi \in \partial h(p^{(k)})$ we set

$$h_k(p) := h(p^{(k)}) + \langle \xi \,, \log_{p^{(k)}} p \rangle_{p^{(k)}}$$

Using musical isomorphisms we identify $X = \xi^\sharp \in T_p\mathcal{M}$,
where we call X a subgradient. Locally h_k minorizes h, i. e.

$$h_k(q) \leq h(q) \quad \text{locally around} \quad p^{(k)}$$

$\Rightarrow$ Use $-h_k(p)$ as upper bound for $-h(p)$ in $f = g - h$.

Note. On $\mathbb{R}^n$ the function h_k is linear.
On a manifold h_k is nonlinear and not even necessarily convex, even on a Hadamard manifold.

The Riemannian DC Algorithm

[RB, O. P. Ferreira, Santos, and Souza 2024]

Input: An initial point $p^{(0)} \in \operatorname{dom}(g)$, g and $\partial_{\mathcal{M}} h$

Set $k = 0$.
while not converged **do**
    Take $X^{(k)} \in \partial_{\mathcal{M}} h(p^{(k)})$
    Compute the next iterate $p^{(k+1)}$ as

$$p^{(k+1)} \in \operatorname*{arg\,min}_{p \in \mathcal{M}} g(p) - \bigl(X^{(k)}\,,\, \log_{p^{(k)}} p\bigr)_{p^{(k)}}. \qquad (*)$$

    Set $k \leftarrow k + 1$
end while

Note. In general the subproblem $(*)$ can not be solved in closed form. But an approximate solution yields a good candidate.

For example: Given g, $p^{(k)}$, and $X^{(k)}$ and $\operatorname{grad} g \Rightarrow$ Gradient descent.

Convergence of the Riemannian DCA

Let $\{p^{(k)}\}_{k\in\mathbb{N}}$ and $\{X^{(k)}\}_{k\in\mathbb{N}}$ be the iterates and subgradients of the RDCA.

Theorem. [RB, O. P. Ferreira, Santos, and Souza 2024]

If $\bar{p}$ is a cluster point of $\{p^{(k)}\}_{k\in\mathbb{N}}$, then $\bar{p} \in \operatorname{dom}(g)$ and there exists a cluster point $\bar{X}$ of $\{X^{(k)}\}_{k\in\mathbb{N}}$ s. t. $\bar{X} \in \partial g(\bar{p}) \cap \partial h(\bar{p})$.

$\Rightarrow$ Every cluster point of $\{p^{(k)}\}_{k\in\mathbb{N}}$, if any, is a critical point of f.

Proposition. [RB, O. P. Ferreira, Santos, and Souza 2024]

Let g be σ-strongly (geodesically) convex. Then

$$f(p^{(k+1)}) \leq f(p^{(k)}) - \frac{\sigma}{2} d^2(p^{(k)}, p^{(k+1)})$$

and $\displaystyle\sum_{k=0}^{\infty} d^2(p^{(k)}, p^{(k+1)}) < \infty$, so in particular $\displaystyle\lim_{k\to\infty} d(p^{(k)}, p^{(k+1)}) = 0$.

NTNU

A Numerical Example

The Difference of Convex Algorithm in Manopt.jl

The algorithm is implemented and released in Julia using `Manopt.jl`[3].
It can be used with any manifold from `Manifolds.jl`

A solver call looks like

```
q = difference_of_convex_algorithm(M, f, g, ∂h, p0)
```

where one has to implement `f(M, p)`, `g(M, p)`, and `∂h(M, p)`.

- a sub problem is generated if keyword `grad_g=` is set
- an efficient version of its cost and gradient is provided
- you can specify the sub-solver using `sub_state=`
  to also set up the specific parameters of your favourite algorithm

[3] see https://manoptjl.org/stable/solvers/difference_of_convex/

Rosenbrock and First Order Methods

Problem. We consider the classical Rosenbrock example[4]

$$\operatorname*{arg\,min}_{x\in\mathbb{R}^2} a\big(x_1^2 - x_2\big)^2 + \big(x_1 - b\big)^2,$$

where $a, b > 0$, usually $b = 1$ and $a \gg b$, here: $a = 2 \cdot 10^5$.

Known Minimizer $x^* = \begin{pmatrix} b \\ b^2 \end{pmatrix}$ with cost $f(x^*) = 0$.

Goal. Compare first-order methods, e. g. using the (Euclidean) gradient

$$\nabla f(x) = \begin{pmatrix} 4a(x_1^2 - x_2) \\ -2a(x_1^2 - x_2) \end{pmatrix} + \begin{pmatrix} 2(x_1 - b) \\ 0 \end{pmatrix}$$

[4] available online in ManoptExamples.jl

NTNU

A "Rosenbrock-Metric" on $\mathbb{R}^2$

In our Riemannian framework, we can introduce a new metric on $\mathbb{R}^2$ as

$$G_p := \begin{pmatrix} 1+4p_1^2 & -2p_1 \\ -2p_1 & 1 \end{pmatrix}, \text{ with inverse } G_p^{-1} = \begin{pmatrix} 1 & 2p_1 \\ 2p_1 & 1+4p_1^2 \end{pmatrix}.$$

We obtain $(X, Y)_p = X^{\mathrm{T}} G_p Y$

The exponential and logarithmic map are given as

$$\exp_p(X) = \begin{pmatrix} p_1 + X_1 \\ p_2 + X_2 + X_1^2 \end{pmatrix}, \qquad \log_p(q) = \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 - (q_1 - p_1)^2 \end{pmatrix}.$$

`Manifolds.jl`:
Implement these functions on `MetricManifold(ℝ^2, RosenbrockMetric())`.

The Riemannian Gradient w.r.t. the new Metric

Let $f\colon \mathcal{M} \to \mathbb{R}$. Given the Euclidean gradient $\nabla f(p)$, its Riemannian gradient $\operatorname{grad} f\colon \mathcal{M} \to T\mathcal{M}$ is given by

$$\operatorname{grad} f(p) = G_p^{-1} \nabla f(p).$$

While we could implement this denoting $\nabla f(p) = \begin{pmatrix} f_1'(p) & f_2'(p) \end{pmatrix}^\mathrm{T}$ using

$$\left\langle \operatorname{grad} f(q), \log_q p \right\rangle_q = (p_1 - q_1) f_1'(q) + (p_2 - q_2 - (p_1 - q_1)^2) f_2'(q),$$

but it is automatically done in `Manopt.jl`.

The Experiment Setup

Algorithms. We now compare

1. The Euclidean gradient descent algorithm on $\mathbb{R}^2$,
2. The Riemannian gradient descent algorithm on $\mathcal{M}$,
3. The Difference of Convex Algorithm on $\mathbb{R}^2$,
4. The Difference of Convex Algorithm on $\mathcal{M}$.

For DCA third we split f into $f(x) = g(x) - h(x)$ with

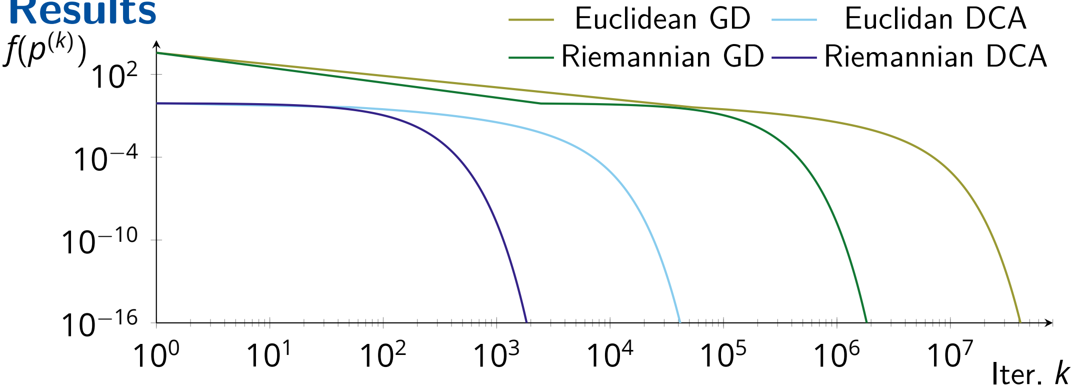
$$g(x) = a\left(x_1^2 - x_2\right)^2 + 2\left(x_1 - b\right)^2 \quad \text{and} \quad h(x) = \left(x_1 - b\right)^2.$$

Initial point. $p_0 = \frac{1}{10}\begin{pmatrix}1\\2\end{pmatrix}$ with cost $f(p_0) \approx 7220.81$.

Stopping Criterion.
$d_{\mathcal{M}}(p^{(k)}, p^{(k-1)}) < 10^{-16}$ or $\lVert \operatorname{grad} f(p^{(k)})\rVert_p < 10^{-16}$.

NTNU

Results



Algorithm	**Runtime** (sec.)	**# Iterations**
Euclidean GD	305.567	53 073 227
Euclidean DCA	58.268	50 588
Riemannian GD	18.894	2 454 017
Riemannian DCA	7.704	2 459

Summary

- Introduced the Convex Bundle Method on manifolds to solve

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} f(p)$$

  - Provide an inner approximation of $\partial_\varepsilon f(p)$
  - A quadratic sub problem in a tangent space
  - Convergence of the Method on Hadamard manifolds
- Introduced the Difference of Convex Algorithm to solve

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} g(p) - h(p)$$

  - Relation to Fenchel Duality on Hadamard manifolds
  - Convergence on Hadamard manifolds

Selected References

RB (2022). “Manopt.jl: Optimization on Manifolds in Julia”. In: *Journal of Open Source Software* 7.70, p. 3866. DOI: 10.21105/joss.03866.

RB, O. P. Ferreira, E. M. Santos, and J. C. d. O. Souza (2024). “The difference of convex algorithm on Hadamard manifolds”. In: *Journal of Optimization Theory and Applications*. DOI: 10.1007/s10957-024-02392-8. arXiv: 2112.05250.

RB, R. Herzog, and H. Jasa (2024). *The Riemannian convex bundle method*. arXiv: 2402.13670.

Geiger, C. and C. Kanzow (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. New York: Springer. DOI: 10.1007/978-3-642-56004-0.

Silva Louzeiro, M., RB, and R. Herzog (2022). “Fenchel Duality and a Separation Theorem on Hadamard Manifolds”. In: *SIAM Journal on Optimization* 32.2, pp. 854–873. DOI: 10.1137/21M1400699. arXiv: 2102.11155.

Souza, J. C. d. O. and P. R. Oliveira (2015). “A proximal point algorithm for DC fuctions on Hadamard manifolds”. In: *Journal of Global Optimization* 63.4, pp. 797–810. DOI: 10.1007/s10898-015-0282-7.

Interested in Numerical Differential Geometry? Join numdiffgeo.zulipchat.com!

ronnybergmann.net/talks/2024-Goettingen-Nonsmooth.pdf