



NTNU

Nonsmooth, nonconvex Optimization on Riemannian Manifolds

Ronny Bergmann

Workshop “[From Modeling and Analysis to Approximation](#)”,

Hasenwinkel,

September 4, 2023

Motivation

The Rayleigh Quotient

When minimizing the **Rayleigh quotient** for a symmetric $A \in \mathbb{R}^{n \times n}$

$$\arg \min_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^T A x}{\|x\|^2}$$

⚠ Any eigenvector x^* to the smallest EV λ is a minimizer

👎 no isolated minima **and** Newton's method diverges

💡 Constrain the problem to unit vectors $\|x\| = 1$!

classic constrained optimization (ALM, EPM,...)

Today Utilize the geometry of the sphere

⬆ unconstrained optimization $\arg \min_{p \in \mathbb{S}^{n-1}} p^T A p$


⌘ adapt unconstrained optimization to **Riemannian manifolds**.


The Generalized Rayleigh Quotient

More general. Find a basis for the space of eigenvectors to $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k$:

$$\arg \min_{X \in \text{St}(n, k)} \text{tr}(X^T A X), \quad \text{St}(n, k) := \{X \in \mathbb{R}^{n \times k} \mid X^T X = I\},$$

 a problem on the **Stiefel** manifold $\text{St}(n, k)$

 Invariant under rotations within a k -dim subspace.

 Find the best subspace!

$$\arg \min_{\text{span}(X) \in \text{Gr}(n, k)} \text{tr}(X^T A X), \quad \text{Gr}(n, k) := \{\text{span}(X) \mid X \in \text{St}(n, k)\},$$

 a problem on the **Grassmann** manifold $\text{Gr}(n, k) = \text{St}(n, k)/O(k)$.

Optimization on Riemannian Manifolds

We are looking for **numerical algorithms** to find

$$\arg \min_{p \in \mathcal{M}} f(p)$$

where

- ▶ \mathcal{M} is a Riemannian manifold
- ▶ $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is a function
- ⚠ f might be **nonsmooth** and/or **nonconvex**
- ⚠ \mathcal{M} might be **high-dimensional**

A Riemannian Manifold \mathcal{M}

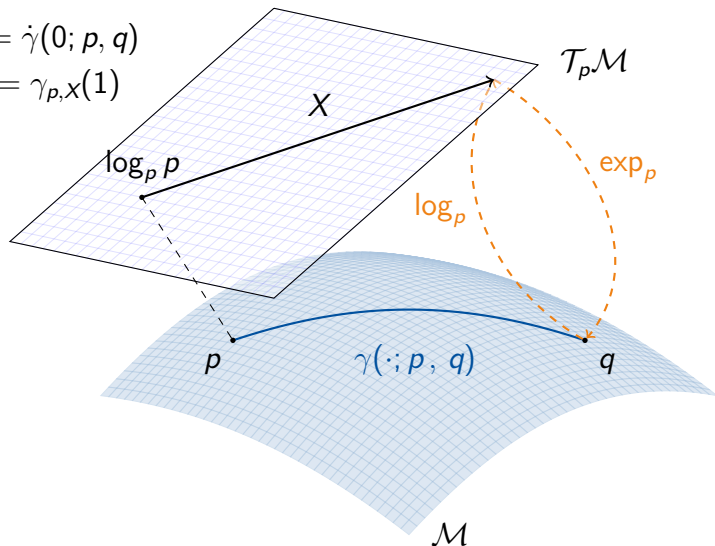
A d -dimensional Riemannian manifold can be informally defined as a set \mathcal{M} covered with a “suitable” collection of charts, that identify subsets of \mathcal{M} with open subsets of \mathbb{R}^d and a continuously varying inner product on the tangent spaces.

[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 $(\cdot, \cdot)_p$



(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 $F: \mathcal{C} \rightarrow \overline{\mathbb{R}}$ 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

The **subdifferential** of f at $p \in \mathcal{C}$ is given by

[Lee 2003; Udriște 1994]

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

where

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

Musical Isomorphisms

Using the tangent space $\mathcal{T}_p\mathcal{M}$ and its dual $\mathcal{T}_p^*\mathcal{M}$,
the inner product $(\cdot, \cdot)_p$ and the duality pairing $\langle \cdot, \cdot \rangle$,

the **musical isomorphisms** are

[Lee 2003]

$$\flat: \mathcal{T}_p\mathcal{M} \rightarrow \mathcal{T}_p^*\mathcal{M} \quad \text{and} \quad \sharp: \mathcal{T}_p^*\mathcal{M} \rightarrow \mathcal{T}_p\mathcal{M}$$

such that for any $X, Y \in \mathcal{T}_p\mathcal{M}$ and $\xi \in \mathcal{T}_p^*\mathcal{M}$ we have

$$\langle X^\flat, Y \rangle = (X, Y)_p \quad \text{and} \quad (\xi^\sharp, Y)_p = \langle \xi, Y \rangle$$

The Proximal Map

For a function $f: \mathcal{M} \rightarrow \mathbb{R}$ and a $\lambda > 0$ we define the **proximal map** as
[Moreau 1965; Rockafellar 1970; O. Ferreira and Oliveira 2002]

$$\text{prox}_{\lambda f}(p) := \arg \min_{q \in \mathcal{M}} d_{\mathcal{M}}(q, p)^2 + \lambda f(q).$$

Properties.

- ▶ Minimizer p^* of $f \Leftrightarrow$ fix point of the prox $\text{prox}_{\lambda f}(p^*) = p^*$
- ▶ If f is proper, convex, lsc.: arg min unique.
- ▶ proximal point algorithm (PPA): $p^{(k+1)} = \text{prox}_{\lambda f}(p^{(k)})$ converges to p^*

Nonsmooth splittings

Splitting Methods & Algorithms

For $\arg \min_{p \in \mathcal{M}} f(p) + g(p)$ we can use

- ▶ Cyclic Proximal Point Algorithm (CPPA) [Bačák 2014]
- ▶ (parallel) Douglas–Rachford Algorithm (PDRA) [RB, Persch, and Steidl 2016]

which are for $\mathcal{M} = \mathbb{R}^n$ also equivalent to [Setzer 2011; O'Connor and Vandenberghe 2018]

- ▶ Primal-Dual Hybrid Gradient Algorithm (PDHGA) [Esser, Zhang, and Chan 2010]
- ▶ Chambolle-Pock Algorithm (CPA) [Chambolle and Pock 2011; Pock, Cremers, Bischof, and Chambolle 2009]

Challenge.

These rely on the dual space of \mathbb{R}^n , which \mathcal{M} does not have.
[More precisely](#). They employ the Fenchel conjugate.

The Fenchel Conjugate

The **Fenchel conjugate** of a function $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is given by

$$f^*(\xi) := \sup_{x \in \mathbb{R}^n} \langle \xi, x \rangle - f(x) = \sup_{x \in \mathbb{R}^n} \begin{pmatrix} \xi \\ -1 \end{pmatrix}^T \begin{pmatrix} x \\ f(x) \end{pmatrix}$$

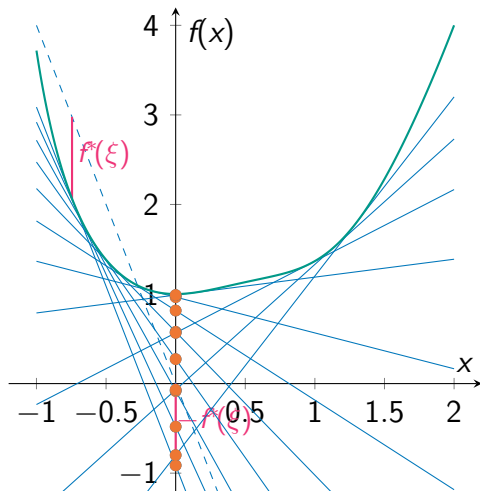
- ▶ given $\xi \in \mathbb{R}^n$: maximize the distance between $\xi^T \cdot$ and f
- ▶ **can also** be written in the epigraph

The **Fenchel biconjugate** reads

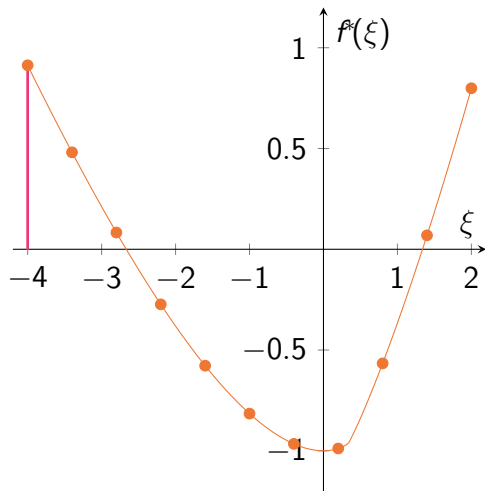
$$f^{**}(x) = (f^*)^*(x) = \sup_{\xi \in \mathbb{R}^n} \langle \xi, x \rangle - f^*(\xi).$$

Illustration of the Fenchel Conjugate

The function f



The Fenchel conjugate f^*



Properties of the Fenchel Conjugate

[Rockafellar 1970]

- ▶ The Fenchel conjugate f^* is **convex** (even if f is not)
- ▶ f^{**} is the largest convex, lsc function with $f^{**} \leq f$
- ▶ If $f(x) \leq g(x)$ for all $x \in \mathbb{R}^n \Rightarrow f^*(\xi) \geq g^*(\xi)$ for all $\xi \in \mathbb{R}^n$
- ▶ **Fenchel–Moreau Theorem.** f convex, proper, l. s. c. $\Rightarrow f^{**} = f$.
- ▶ **Fenchel–Young inequality.**

$$f(x) + f^*(\xi) \geq \xi^T x \quad \text{for all } x, \xi \in \mathbb{R}^n$$

- ▶ For a proper, convex function f

$$\xi \in \partial f(x) \Leftrightarrow f(x) + f^*(\xi) = \xi^T x$$

- ▶ For a proper, convex, lsc function f , then

$$\xi \in \partial f(x) \Leftrightarrow x \in \partial f^*(\xi)$$

The (Riemannian) m -Fenchel Conjugate

[RB, Herzog, Silva Louzeiro, Tenbrinck, and Vidal-Núñez 2021]

Idea. Localize to $\mathcal{C} \subset \mathcal{M}$ around a point m which “acts as” 0.

The m -Fenchel conjugate of a function $f: \mathcal{C} \rightarrow \overline{\mathbb{R}}$ is given by

$$f_m^*(\xi_m) := \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{ \langle \xi_m, X \rangle - f(\exp_m X) \},$$

where $\mathcal{L}_{\mathcal{C},m} := \{X \in \mathcal{T}_m \mathcal{M} \mid q = \exp_m X \in \mathcal{C} \text{ and } \|X\|_p = d(q, p)\}$.

Let $m' \in \mathcal{C}$. The mm' -Fenchel-biconjugate $F_{mm'}^{**}: \mathcal{C} \rightarrow \overline{\mathbb{R}}$ is given by

$$F_{mm'}^{**}(p) = \sup_{\xi_{m'} \in \mathcal{T}_{m'}^* \mathcal{M}} \{ \langle \xi_{m'}, \log_{m'} p \rangle - F_m^*(P_{m \leftarrow m'} \xi_{m'}) \},$$

where usually we only use the case $m = m'$.

Properties of the m -Fenchel Conjugate

- ▶ f_m^* is convex on $\mathcal{T}_m^*\mathcal{M}$
- ▶ If $f(p) \leq g(p)$ for all $p \in \mathcal{C} \Rightarrow f_m^*(\xi_m) \geq g_m^*(\xi_m)$ for all $\xi_m \in \mathcal{T}_m^*\mathcal{M}$
- ▶ Fenchel–Moreau Theorem $f \circ \exp_m$ convex (on $\mathcal{T}_m\mathcal{M}$), proper, lsc, $\Rightarrow f_{mm}^{**} = f$ on \mathcal{C} .
- ▶ Fenchel-Young inequality: For a proper, convex function $f \circ \exp_m$

$$\xi_p \in \partial_{\mathcal{M}} f(p) \Leftrightarrow f(p) + f_m^*(P_{m \leftarrow p} \xi_p) = \langle P_{m \leftarrow p} \xi_p, \log_m p \rangle.$$

- ▶ For a proper, convex, lsc function $f \circ \exp_m$

$$\xi_p \in \partial_{\mathcal{M}} f(p) \Leftrightarrow \log_m p \in \partial f_m^*(P_{m \leftarrow p} \xi_p).$$

The Chambolle–Pock Algorithm

From the pair of primal-dual problems

[Chambolle and Pock 2011]

$$\min_{x \in \mathbb{R}^n} f(x) + g(Kx), \quad K \text{ linear,}$$

$$\max_{\xi \in \mathbb{R}^m} -f^*(-K^*\xi) - g^*(\xi)$$

we obtain for f, g proper convex, lsc the optimality conditions of a solution $(\hat{x}, \hat{\xi})$ as

$$-K^*\hat{\xi} \in \partial f(\hat{x})$$

$$K\hat{x} \in \partial g^*(\hat{\xi})$$

The Chambolle–Pock Algorithm

From the pair of primal-dual problems

[Chambolle and Pock 2011]

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) + g(Kx), \quad K \text{ linear,} \\ \max_{\xi \in \mathbb{R}^m} \quad & -f^*(-K^*\xi) - g^*(\xi) \end{aligned}$$

we obtain for f, g proper convex, lsc the

Chambolle–Pock Algorithm. with $\sigma > 0$, $\tau > 0$, $\theta \in \mathbb{R}$ reads

$$\begin{aligned} x^{(k+1)} &= \text{prox}_{\sigma f}(x^{(k)} - \sigma K^* \bar{\xi}^{(k)}) \\ \xi^{(k+1)} &= \text{prox}_{\tau g^*}(\xi^{(k)} + \tau K x^{(k+1)}) \\ \bar{\xi}^{(k+1)} &= \xi^{(k+1)} + \theta(\xi^{(k+1)} - \xi^{(k)}) \end{aligned}$$

Saddle Point Formulation on Manifolds

On manifolds, we consider for

$$\min_{p \in \mathcal{M}} f(p) + g(\Lambda p), \quad \Lambda: \mathcal{M} \rightarrow \mathcal{N},$$

where f is geodesically convex, and $g \circ \exp_n$ is convex for some $n \in \mathcal{N}$.

Saddle point formulation. Using the n -Fenchel conjugate g_n^* of g :

$$\min_{p \in \mathcal{C}} \max_{\xi_n \in \mathcal{T}_n^* \mathcal{N}} \langle \xi_n, \log_n \Lambda(p) \rangle + f(p) - g_n^*(\xi_n).$$

But. Λ is inherently nonlinear and inside a logarithmic map \Rightarrow no adjoint.

Approach. Linearization: Choose m such that $n = \Lambda(m)$ and

[Valkonen 2014]

$$\Lambda(p) \approx \exp_{\Lambda(m)} D\Lambda(m)[\log_m p].$$

The exact Riemannian Chambolle–Pock Algorithm

[RB, Herzog, Silva Louzeiro, Tenbrinck, and Vidal-Núñez 2021; Chambolle and Pock 2011]

Input: $m, p^{(0)} \in \mathcal{C} \subset \mathcal{M}$, $n = \Lambda(m)$, $\xi_n^{(0)} \in \mathcal{T}_n^* \mathcal{N}$, and $\sigma, \tau, \theta > 0$

1: $k \leftarrow 0$

2: $\bar{p}^{(0)} \leftarrow p^{(0)}$

3: **while** not converged **do**

4: $\xi_n^{(k+1)} \leftarrow \text{prox}_{\tau g_n^*} \left(\xi_n^{(k)} + \tau \left(\log_n \Lambda(\bar{p}^{(k)}) \right)^\flat \right)$

5: $p^{(k+1)} \leftarrow \text{prox}_{\sigma f} \left(\exp_{p^{(k)}} \left(P_{p^{(k)} \leftarrow m} \left(-\sigma D\Lambda(m)^* [\xi_n^{(k+1)}] \right)^\sharp \right) \right)$

6: $\bar{p}^{(k+1)} \leftarrow \exp_{p^{(k+1)}} \left(-\theta \log_{p^{(k+1)}} p^{(k)} \right)$

7: $k \leftarrow k + 1$

8: **end while**

Output: $p^{(k)}$

Difference of Convex

Difference of Convex

We aim to solve

$$\arg \min_{p \in \mathcal{M}} f(p)$$

where

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

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

- ▶ $g, h: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ are convex, lower semicontinuous, and proper

The Euclidean DCA

Idea 1. At x_k , approximate $h(x)$ by its affine minorization $h_k(x) := h(x^{(k)}) + \langle x - x^{(k)}, y^{(k)} \rangle$ for some $y^{(k)} \in \partial h(x^{(k)})$.

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

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

$$y^{(k)} \in \arg \min_{y \in \mathbb{R}^n} \left\{ h^*(y) - g^*(y^{(k-1)}) - \langle y - y^{(k-1)}, x^{(k)} \rangle \right\}$$

Idea 3. Reformulate 2 using a proximal map \Rightarrow DCPPA

On manifolds:

[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

[Silva Louzeiro, RB, and Herzog 2022]

Definition

Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$. The **Fenchel conjugate** of f is the function $f^*: \mathcal{T}^*\mathcal{M} \rightarrow \overline{\mathbb{R}}$ defined by

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

The Dual Difference of Convex Problem

Given the Difference of Convex problem

$$\arg \min_{p \in \mathcal{M}} g(p) - h(p)$$

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

[RB, O. P. Ferreira, Santos, and Souza 2023]

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

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

[RB, O. P. Ferreira, Santos, and Souza 2023]

Theorem.

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

The Dual Difference of Convex Problem

The primal and dual Difference of convex problem

$$\arg \min_{p \in \mathcal{M}} g(p) - h(p) \quad \text{and} \quad \arg \min_{(p, \xi) \in T^* \mathcal{M}} h^*(p, \xi) - g^*(p, \xi)$$

are equivalent in the following sense.

Theorem.

[RB, O. P. Ferreira, Santos, and Souza 2023]

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 linearization of h at some point $p^{(k)}$:

With $\xi \in \partial h(p^{(k)})$ we get

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

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

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

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

Note. On \mathbb{R}^n the function h_k is linear.

On a manifold h_k is not necessarily **convex**, even on a Hadamard manifold.

The Riemannian DC Algorithm

[RB, O. P. Ferreira, Santos, and Souza 2023]

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

- 1: Set $k = 0$.
- 2: **while** not converged **do**
- 3: Take $X_k \in \partial_{\mathcal{M}}h(p_k)$
- 4: Compute the next iterate p^{k+1} as

$$p_{k+1} \in \arg \min_{p \in \mathcal{M}} g(p) - (X_k, \log_{p_k} p)_{p_k}. \quad (*)$$

- 5: Set $k \leftarrow k + 1$
- 6: **end while**

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

Convergence of the Riemannian DCA

[RB, O. P. Ferreira, Santos, and Souza 2023]

Let $\{p_k\}_{k \in \mathbb{N}}$ and $\{X_k\}_{k \in \mathbb{N}}$ be the iterates and subgradients of the RDCA.

Theorem.

If \bar{p} is a cluster point of $\{p_k\}_{k \in \mathbb{N}}$, then $\bar{p} \in \text{dom}(g)$ 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. 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 $\sum_{k=0}^{\infty} d^2(p_k, p_{k+1}) < \infty$, so in particular $\lim_{k \rightarrow \infty} d(p_k, p_{k+1}) = 0$.

Software



[Axen, Baran, RB, and Rzecki 2023]

Goal. Provide an interface to implement and use Riemannian manifolds.

Interface `AbstractManifold` to model manifolds

Functions like `exp(M, p, X)`, `log(M, p, X)` or `retract(M, p, X, method)`.

Decorators for implicit or explicit specification of an embedding, a metric, or a group,

Efficiency by providing in-place variants like `exp!(M, q, p, X)`



[Axen, Baran, RB, and Rzecki [2023](#)]

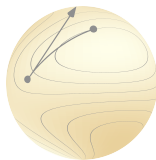
Goal. Provide a library of Riemannian manifolds, that is efficiently implemented and well-documented

Meta. generic implementations for $\mathcal{M}^{n \times m}$, $\mathcal{M}_1 \times \mathcal{M}_2$, vector- and tangent-bundles, esp. $T_p\mathcal{M}$, or Lie groups

Library. Implemented functions for

- ▶ Circle, Sphere, Torus, Hyperbolic
- ▶ (generalized, symplectic) Stiefel, (generalized) Grassmann, Rotations
- ▶ symmetric positive definite matrices
- ▶ multinomial, symmetric, symplectic matrices
- ▶ Tucker & Oblique manifold, Kendall's Shape space

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)`
⇒ an algorithm in the `Manopt.jl` interface

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

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

Manopt family.



manoptjl.org

[RB 2022]



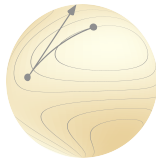
manopt.org

[Boumal, Mishra, Absil, and Sepulchre 2014]



pymanopt.org

[Townsend, Koep, and Weichwald 2016]



Algorithms.

Cost-based Nelder-Mead, Particle Swarm

Subgradient-based Subgradient Method

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

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

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

constrained Augmented Lagrangian, Exact Penalty, Frank-Wolfe

nonconvex Difference of Convex Algorithm, DCPPA

Implementation of the DCA

The algorithm is implemented and released in Julia using `Manopt.jl`¹. 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 $\partial h(M, p)$.

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

¹see https://manoptjl.org/stable/solvers/difference_of_convex/

Numerical Examples

The ℓ^2 -TV Model

[Rudin, Osher, and Fatemi 1992; Lellmann, Strekalovskiy, Koetter, and Cremers 2013; Weinmann, Demaret, and Storath 2014]

For a manifold-valued image $q \in \mathcal{M}$, $\mathcal{M} = \mathcal{N}^{d_1, d_2}$, we compute

$$\arg \min_{p \in \mathcal{M}} \frac{1}{2\alpha} d_{\mathcal{M}}^2(p, q) + \|\lambda(p)\|_{g,s,1}$$

with

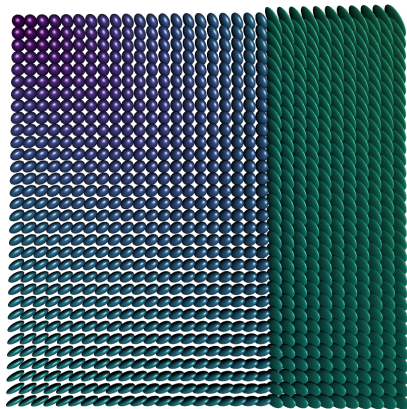
► “forward differences” $\Lambda: \mathcal{M} \rightarrow (T\mathcal{M})^{d_1-1, d_2-1, 2}$,

$$p \mapsto \Lambda(p) = \left((\log_{p_i} p_{i+e_1}, \log_{p_i} p_{i+e_2}) \right)_{i \in \{1, \dots, d_1-1\} \times \{1, \dots, d_2-1\}}$$

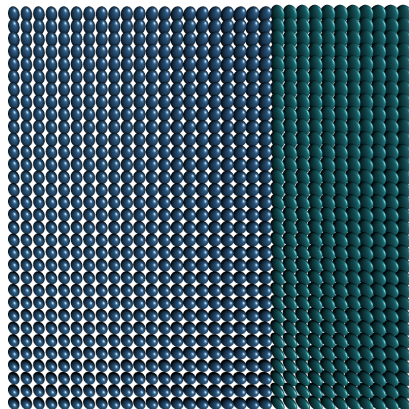
► $\|X\|_{g,s,1}$ similar to a collaborative TV, [Duran, Moeller, Sbert, and Cremers 2016]

⇒ anisotropic TV ($s = 1$) and isotropic TV ($s = 2$)

Numerical Example for a $\mathcal{P}(3)$ -valued Image



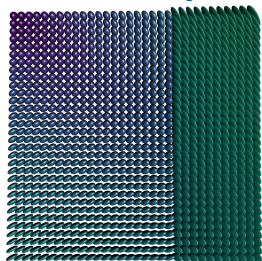
$\mathcal{P}(3)$ -valued data.



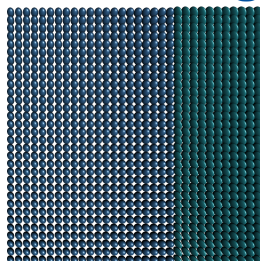
anisotropic TV, $\alpha = 6$.

- ▶ in each pixel we have a symmetric positive definite matrix
- ▶ Applications: denoising/inpainting e.g. of DT-MRI data

Numerical Example for a $\mathcal{P}(3)$ -valued Image



$\mathcal{P}(3)$ -valued data.



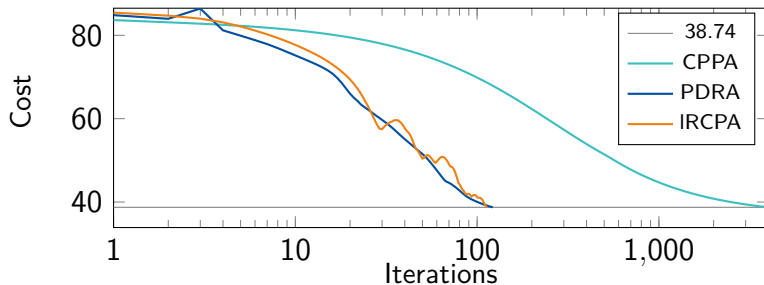
anisotropic TV, $\alpha = 6$.

Approach. CPPA as benchmark

[Bačák 2014; RB, Persch, and Steidl 2016; RB, Herzog, Silva Louzeiro, Tenbrinck, and Vidal-Núñez 2021]

	CPPA	PDRA	IRCPA
parameters	$\lambda_k = \frac{4}{k}$	$\lambda = 0.58$ $\beta = 0.93$	$\sigma = \tau = 0.4$ $\gamma = 0.2, m = l$
iterations	4000	122	113
runtime	1235 s.	380 s.	96.1 s.

Numerical Example for a $\mathcal{P}(3)$ -valued Image



Approach. CPPA as benchmark

[Bačák 2014; RB, Persch, and Steidl 2016; RB, Herzog, Silva Louzeiro, Tenbrinck, and Vidal-Núñez 2021]

	CPPA	PDRA	IRCPA
parameters	$\lambda_k = \frac{4}{k}$	$\lambda = 0.58$ $\beta = 0.93$	$\sigma = \tau = 0.4$ $\gamma = 0.2, m = l$
iterations	4000	122	113
runtime	1235 s.	380 s.	96.1 s.

Rosenbrock and First Order Methods

Problem. We consider the classical Rosenbrock example²

$$\arg \min_{x \in \mathbb{R}^2} a(x_1^2 - x_2)^2 + (x_1 - b)^2,$$

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

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

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 $(X, Y)_p = X^\top G_p Y$

The exponential and logarithmic map are given as

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

`Manifolds.jl`:

Implement these functions on `MetricManifold(\mathbb{R}^2 , RosenbrockMetric())`.

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

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

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

While we could implement this denoting $\nabla f(p) = (f'_1(p) \ f'_2(p))^T$ using

$$\left\langle \text{grad } f(q), \log_q p \right\rangle_q = (p_1 - q_1) f'_1(q) + (p_2 - q_2 - (p_1 - q_1)^2) f'_2(q),$$

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} .

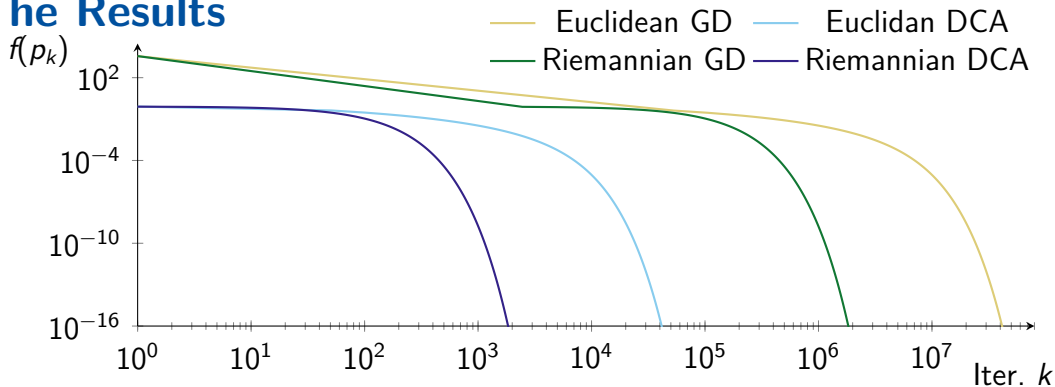
For DCA third we split f into $f(x) = g(x) - h(x)$ with

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

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 $\|\text{grad } f(p_k)\|_p < 10^{-16}$.

The Results









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

Summary

We considered two different ways to generalize the Fenchel conjugate to Riemannian manifolds and how they are used in

- ▶ Nonsmooth Riemannian Optimization:
 m -Fenchel Dual and the Chambolle-Pock algorithm
- ▶ Nonconvex Riemannian Optimization:
Fenchel Dual and the Difference of Convex algorithm
- ▶ Numerics in Julia:
`Manopt.jl` together with `ManifoldsBase.jl` & `Manifolds.jl`

Selected References

-  Axen, S. D., M. Baran, RB, and K. Rzecki (2023). “Manifolds.jl: An Extensible Julia Framework for Data Analysis on Manifolds”. In: *ACM Transactions on Mathematical Software*. Accepted for publication. DOI: 10.1145/3618296. arXiv: 2106.08777.
-  RB, O. P. Ferreira, E. M. Santos, and J. C. d. O. Souza (2023). *The difference of convex algorithm on Hadamard manifolds*. arXiv: 2112.05250.
-  RB, R. Herzog, M. Silva Louzeiro, D. Tenbrinck, and J. Vidal-Núñez (Jan. 2021). “Fenchel duality theory and a primal-dual algorithm on Riemannian manifolds”. In: *Foundations of Computational Mathematics* 21.6, pp. 1465–1504. DOI: 10.1007/s10208-020-09486-5. arXiv: 1908.02022.
-  Boumal, N. (2023). *An introduction to optimization on smooth manifolds*. Cambridge University Press. URL: <https://www.nicolasboumal.net/book>.
-  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. ISSN: 1052-6234, 1095-7189. 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.