NTNU
Norwegian University of Science and Technology

Nonsmooth, nonconvex Optimization on Riemannian Manifolds

Ronny Bergmann

Workshop
From Modeling and Analysis to Approximation and Fast Algorithms,
Hasenwinkel, September 4, 2023

Motivation

The Rayleigh Quotient

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

$$\operatorname*{arg\,min}_{x\in\mathbb{R}^n\setminus\{0\}} \frac{x^\mathrm{T}Ax}{\|x\|^2}$$

- 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 $\operatorname*{arg\,min}_{p\in\mathbb{S}^{n-1}} p^\mathrm{T}Ap$
- adapt unconstrained optimization to Riemannian manifolds.

The Generalized Rayleigh Quotient

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

$$\operatorname*{arg\,min}_{X \in \mathrm{St}(n,k)} \mathrm{tr}(X^\mathsf{T} A X), \qquad \mathrm{St}(n,k) := \left\{ X \in \mathbb{R}^{n \times k} \,\middle|\, X^\mathsf{T} X = I \right\},$$

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

Invariant under rotations within a k-dim subspace.

Find the best subspace!

$$\operatorname*{arg\,min}_{\mathrm{span}(X) \in \mathrm{Gr}(n,k)} \mathrm{tr}(X^\mathsf{T} A X), \qquad \mathrm{Gr}(n,k) := \left\{ \mathrm{span}(X) \,\middle|\, X \in \mathrm{St}(n,k) \right\},$$

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

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 $\mathcal{M}$

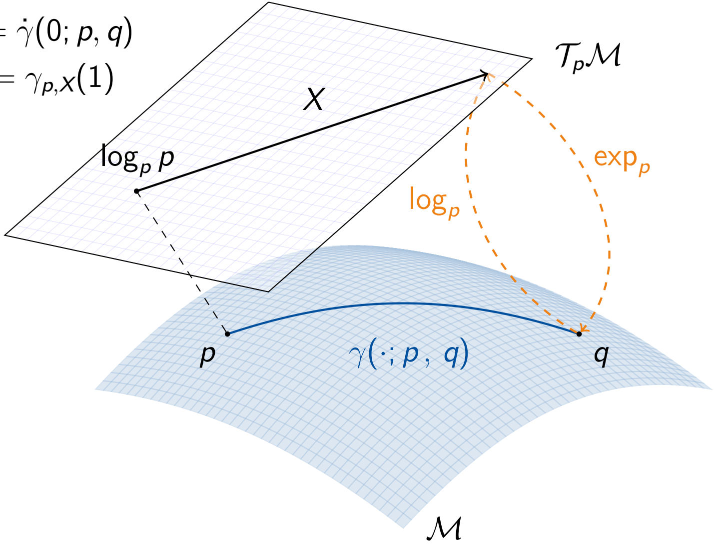
A d-dimensional Riemannian manifold can be informally defined as a set $\mathcal{M}$ covered with a "suitable" collection of charts, that identify subsets of $\mathcal{M}$ with open subsets of $\mathbb{R}^d$ and a continuously varying inner product on the tangent spaces.

[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 $(\cdot, \cdot)_p$



(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

The subdifferential of f at $p \in \mathcal{C}$ is given by [Lee 2003; Udriște 1994]

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

where

- $\mathcal{T}_p^*\mathcal{M}$ is the dual space of $\mathcal{T}_p\mathcal{M}$,
- $\langle \cdot \,, \cdot \rangle_p$ 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]

$$\flat\colon \mathcal{T}_p\mathcal{M} \to \mathcal{T}_p^*\mathcal{M} \quad \text{and} \quad \sharp\colon \mathcal{T}_p^*\mathcal{M} \to \mathcal{T}_p\mathcal{M}$$

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 $f\colon \mathcal{M} \to \mathbb{R}$ and a $\lambda > 0$ we define the proximal map as

[Moreau 1965; Rockafellar 1970; O. Ferreira and Oliveira 2002]

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

Properties.

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

Nonsmooth splittings

Splitting Methods & Algorithms

For $\operatorname*{arg\,min}_{p\in\mathcal{M}} f(p) + g(p)$ 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 equivalend 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 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by

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

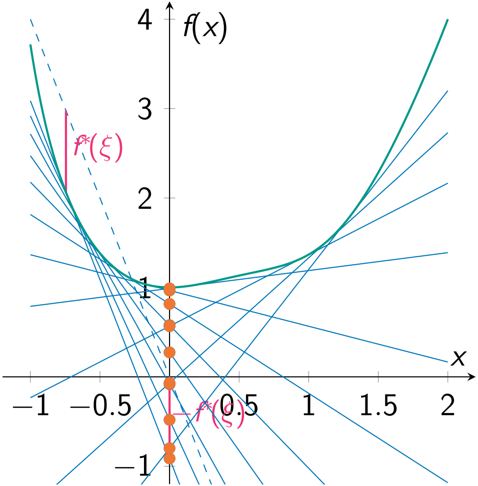
- given $\xi \in \mathbb{R}^n$: maximize the distance between $\xi^{\mathsf{T}}\cdot$ and f
- can also be written in the epigraph

The Fenchel biconjugate reads

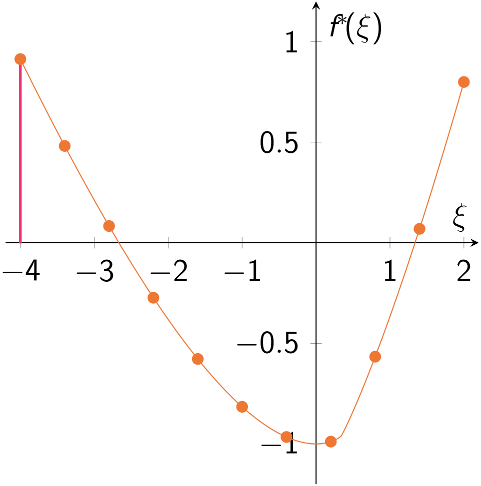
$$f^{**}(x) = (f^*)^*(x) = \sup_{\xi\in\mathbb{R}^n} \langle \xi , x\rangle - f^*(\xi).$$

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, lsc $\Rightarrow f^{**} = f$.
- Fenchel–Young inequality.

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

- For a proper, convex function f

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

- 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 $f\colon \mathcal{C} \to \overline{\mathbb{R}}$ is given by

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

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

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

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

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 $f \circ \exp_m$

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

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

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

NTNU

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]

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

we obtain for f, g proper convex, lsc the

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

$$x^{(k+1)} = \operatorname{prox}_{\sigma f}\big(x^{(k)} - \sigma K^*\bar{\xi}^{(k)}\big)$$
$$\xi^{(k+1)} = \operatorname{prox}_{\tau g,^*}\big(\xi^{(k)} + \tau K x^{(k+1)}\big)$$
$$\bar{\xi}^{(k+1)} = \xi^{(k+1)} + \theta(\xi^{(k+1)} - \xi^{(k)})$$

Saddle Point Formulation on Manifolds

On manifolds, we consider for

$$\min_{p\in\mathcal{M}} f(p) + g(\Lambda p), \qquad \Lambda\colon \mathcal{M} \to \mathcal{N},$$

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: $\quad \xi_n^{(k+1)} \leftarrow \operatorname{prox}_{\tau g_n^*}\Bigl(\xi_n^{(k)} + \tau\bigl(\log_n \Lambda(\bar{p}^{(k)})\bigr)^\flat\Bigr)$

5: $\quad p^{(k+1)} \leftarrow \operatorname{prox}_{\sigma f}\Bigl(\exp_{p^{(k)}}\Bigl(\mathrm{P}_{p^{(k)} \leftarrow m}\bigl(-\sigma D\Lambda(m)^*[\xi_n^{(k+1)}]\bigr)^\sharp\Bigr)\Bigr)$

6: $\quad \bar{p}^{(k+1)} \leftarrow \exp_{p^{(k+1)}}\bigl(-\theta \log_{p^{(k+1)}} p^{(k)}\bigr)$

7: $\quad k \leftarrow k + 1$

8: **end while**

Output: $p^{(k)}$

Difference of Convex

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 $h_k(x) := h(x^{(k)}) + \langle x - x^{(k)}, y^{(k)} \rangle$ 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

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

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

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

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

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 2023]

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

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

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

Theorem.

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

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 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.

NTNU

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

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

$\Rightarrow$ 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 $p^0 \in \operatorname{dom}(g)$, 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

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

  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.

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 $\bar p \in \operatorname{dom}(g)$ 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. 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 $\displaystyle\lim_{k\to\infty} d(p^{(k)}, p_{k+1}) = 0$.

Software

ManifoldsBase.jl

[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)`

Manifolds.jl

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

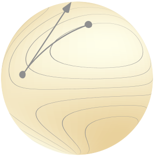
[Axen, Baran, RB, and Rzecki 2023]

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, Projective Spaces
- (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)`
$\Rightarrow$ 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.

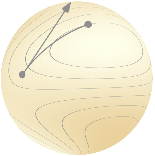
Manopt family.

manoptjl.org [RB 2022]

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

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

Manopt.jl



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

manoptjl.org

NTNU

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 `∂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

[1] see `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

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} \frac{1}{2\alpha} d^2_{\mathcal{M}}(p,q) + \|\Lambda(p)\|_{g,s,1}$$

with

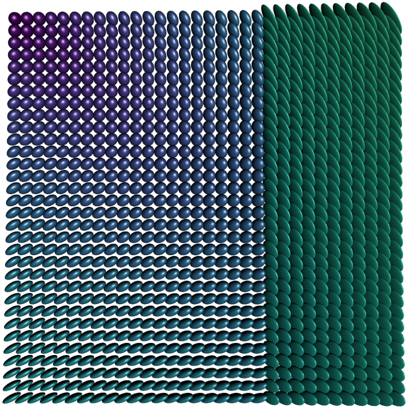
- "forward differences" $\Lambda\colon \mathcal{M} \to (T\mathcal{M})^{d_1-1,\ d_2-1,\ 2}$,

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

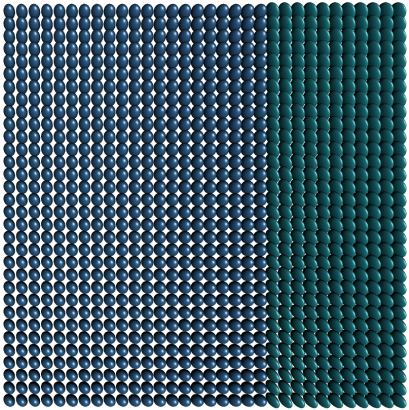
- $\|X\|_{g,s,1}$ similar to a collaborative TV, [Duran, Moeller, Sbert, and Cremers 2016]

$\Rightarrow$ anisotropic TV ($s = 1$) and isotropic TV ($s = 2$)

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



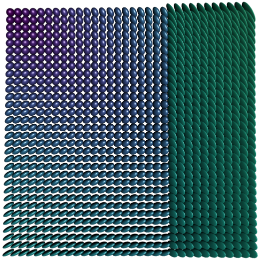
$\mathcal{P}(3)$-valued data.



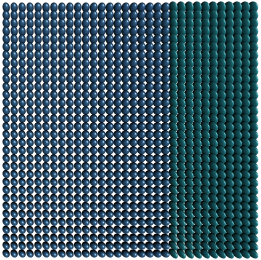
anisotropic TV, $\alpha = 6$.

- 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.



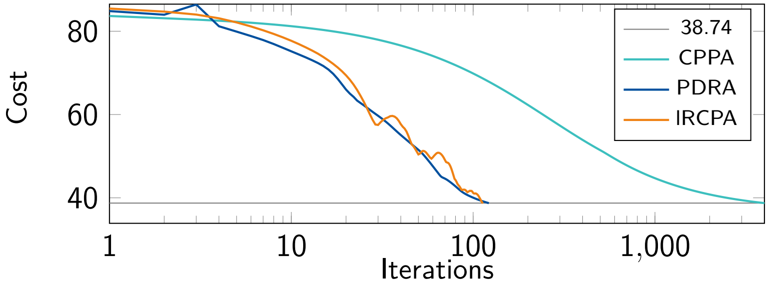
anisotropic TV, $\alpha = 6$.

Approach. CPPA as benchmark

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

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

Rosenbrock and First Order Methods

Problem. We consider the classical Rosenbrock example[2]

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

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

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

[2] available online in ManoptExamples.jl

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

NTNU

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 $\nabla f(p) = \begin{pmatrix} f_1'(p) & f_2'(p) \end{pmatrix}^{\mathsf{T}}$ using

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

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 $\mathcal{M}$.

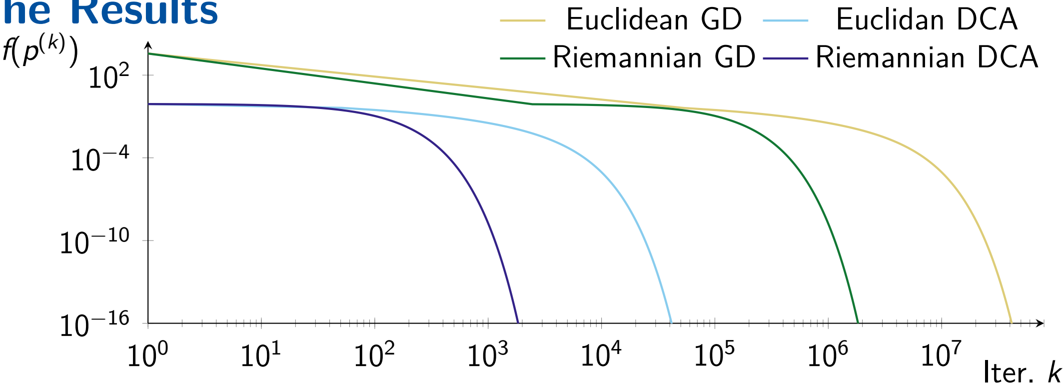
For DCA third we split f into $f(x) = g(x) - h(x)$ with

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

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 $\lVert \operatorname{grad} f(p^{(k)}) \rVert_p < 10^{-16}$.

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 pulication. 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.