Nonlinear Fenchel conjugates and the Riemannian Difference of Convex Algorithm

Ronny Bergmann

joint work with
O. P. Ferreira, R. Herzog, A. Schiela, E. M. Santos, and J. C. O. Souza

MOP Research Seminar

Saarbrücken & online, February 20, 2025.

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
- 💡 f has some “nice structure”

The Fenchel Conjugate

The Fenchel conjugate of a function $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by

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

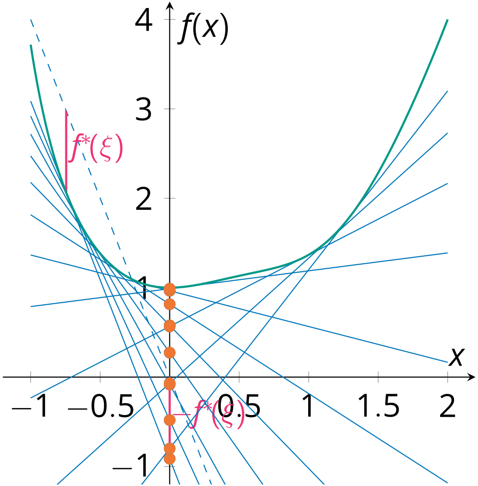
- given $\xi \in \mathbb{R}^n$: maximize the distance between $\xi^{\mathrm{T}}\cdot$ and f
- can also be written in the epigraph

The Fenchel biconjugate reads

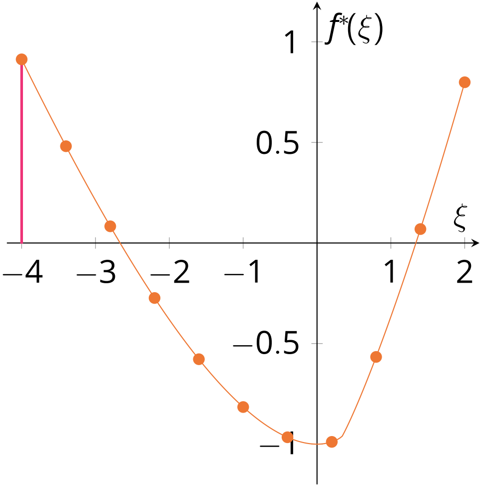
$$f^{**}(x) = (f^*)^*(x) = \sup_{\xi\in\mathbb{R}^n} \langle \xi, x\rangle - f^*(\xi).$$

Illustration of the Fenchel Conjugate

The function f



The Fenchel conjugate f^*



Applications of the Fenchel conjugate

The Fenchel conjugate is at the core of nonsmooth optimization

$$\operatorname*{arg\,min}_{x\in\mathbb{R}^n} f(x) + g(Kx)$$

as a so-called splitting method

- primal-dual (PD) algorithms [Esser, Zhang, Chan, 2010; Chambolle, Pock, 2011]
- PD with non-linear operators K [Valkonen, 2014; Mom, Langer, Sixou, 2022]
- several variants: hybrid gradient, primal/dual relaxed, linearized,...
- To derive the Difference of Convex algorithm (g concave)

Recently this has been generalised Riemannian manifolds using

- a tangent space approach [RB, Herzog, Silva Louzeiro, Tenbrinck, Vidal-Núñez, 2021]
- a tangent bundle approach [Silva Louzeiro, RB, Herzog, 2022]
- Busemann functions [de Carvalho Bento, Neto, Melo, 2023]

Formulate a framework for Fenchel conjugates on nonlinear spaces.

Nonlinear Fenchel conjugates

The Nonlinear Fenchel Conjugate

[Schiela, Herzog, RB, 2024]

In the Fenchel conjugate we use linear test functions $\varphi(x) = \langle \xi, x \rangle$.

Use arbitrary test functions

Let $\mathcal{M}$ be a set. We define the domain of the sum (difference) of two extended real-valued functions $f, g \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ as

$$\mathcal{D}(f \pm g) \coloneqq \{x \in \mathcal{M} \,|\, f(x) \pm g(x) \text{ is defined}\}.$$

Definition

The nonlinear Fenchel conjugate of $f \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ is defined as

$$\begin{aligned} f^{\circledast} \colon \mathcal{P}_{\pm\infty}(\mathcal{M}) &\to \mathbb{R}_{\pm\infty} \\ \varphi &\mapsto f^{\circledast}(\varphi) \coloneqq \sup\{\varphi(x) - f(x) \,|\, x \in \mathcal{D}(\varphi - f)\}. \end{aligned}$$

A few properties

The following properties carry over to the nonlinear case,
just being a bit careful with the domain of the test functions.

Suppose that $\mathcal{M}$ is a set and $f, g \in \mathcal{P}_{\pm\infty}(\mathcal{M})$.

[Schiela, Herzog, RB, 2024]

1. For $\alpha > 0$ and $\beta \in \mathbb{R}$,
$$\alpha f^{\circledast}(\varphi) + \beta = (\alpha f)^{\circledast}(\alpha\,\varphi + \beta) = (\alpha f - \beta)^{\circledast}(\alpha\,\varphi).$$
2. If $\mathcal{D}(f - \psi) = \mathcal{D}(\varphi + \psi) = \mathcal{M}$, then
$$(f - \psi)^{\circledast}(\varphi) = f^{\circledast}(\varphi + \psi).$$
3. If $\mathcal{D}(f + g) = \mathcal{D}(\varphi + \psi) = \mathcal{M}$ and $f^{\circledast}(\varphi) + g^{\circledast}(\psi)$ is defined, then
$$(f + g)^{\circledast}(\varphi + \psi) \leqslant f^{\circledast}(\varphi) + g^{\circledast}(\psi).$$
4. $\varphi \geqslant \psi$ and $f \leqslant g$ implies $f^{\circledast}(\varphi) \geqslant g^{\circledast}(\psi)$.
5. $f^{\circledast}$ is convex on $\mathcal{P}_{\infty}(\mathcal{M})$.

The Fenchel-Young inequality

An important inequality in the classical case is the Fenchel-Young inequality

$$f(x) + f^*(\xi) \geq \langle \xi, x \rangle$$

This carries over to the nonlinear case, with a bit of carefulness as to when the sum is defined.

Theorem (Fenchel-Young inequality)

Suppose that $f, \varphi \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ and $x \in \mathcal{M}$.
The Fenchel-Young inequalities

- $f^{\circledast}(\varphi) \geqslant \varphi(x) - f(x)$
- $f(x) \geqslant \varphi(x) - f^{\circledast}(\varphi)$
- $\varphi(x) \leqslant f(x) + f^{\circledast}(\varphi)$

hold, provided that the respective right-hand side is defined in $\mathbb{R}_{\pm\infty}$.

Nonlinear dual map

Motivation. In the classical case, we often need the adjoint K^* of K.

Definition

Suppose $\mathcal{M}$ and $\mathcal{N}$ are two non-empty sets and $A\colon \mathcal{M} \to \mathcal{N}$ is some map. The map

$$
\begin{aligned}
A^{\circledast}\colon \mathcal{P}_{\pm\infty}(\mathcal{N}) &\to \mathcal{P}_{\pm\infty}(\mathcal{M}) \\
\psi &\mapsto A^{\circledast}(\psi) := \psi \circ A
\end{aligned}
$$

is called the dual or adjoint map of A, or the pullback by A.

- $A^{\circledast}(\alpha\,\psi_1 + \psi_2) = \alpha\,A^{\circledast}(\psi_1) + A^{\circledast}(\psi_2)$ is a homomorphism
- If A is bijective, then $(f \circ A^{-1})^{\circledast} = f^{\circledast} \circ A^{\circledast}$
- more generally: defining $(f \bullet A^{-1})(y) := \inf_{x \in A^{-1}(y)} f(x)$,

  we obtain $(f \bullet A^{-1})^{\circledast} = f^{\circledast} \circ A^{\circledast}$.

Motivation: The biconjugate

- approximate f its maximal convex, lsc. minorant
- linear setting: Γ-regularization, the pointwise suppremum of continuous affine functions. [Ch. I.3 Ekeland, Temam, 1999]

$\Rightarrow$ $f^{**} \in \mathcal{P}_{\pm\infty}(V)$ coincides with Γ-regularization of f, i. e. the largest convex lsc. minorant of $f \in \mathcal{P}_{\pm\infty}(V)$

- **Fenchel-Moreau**: [Thm. 13.32 Bauschke, Combettes, 2011]

  $f \in \mathcal{P}_{\infty}(V)$ is convex, lsc. $\Leftrightarrow f^{**} = f$.

Nonlinear case.
Find a suitable subset $\mathcal{F} \subset \mathcal{P}_{\pm\infty}(\mathcal{M})$
as a generalization for affine functions.

? Can we state a biconjugation theorem as well?

$\mathcal{F}$-regularization

[Schiela, Herzog, RB, 2024]

Suppose that $\emptyset \neq \mathcal{F} \subseteq \mathcal{P}_{\pm\infty}(\mathcal{M})$ and denote by

$$\widetilde{\mathcal{F}} := \{\varphi + c \,|\, \varphi \in \mathcal{F},\ c \in \mathbb{R}\}$$

the set of all φ that result from a shift of elements of $\mathcal{F}$.

We define the $\mathcal{F}$-regularization of $f \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ as

$$\lfloor f \rfloor_{\mathcal{F}}(x) := \sup\bigl\{\varphi(x) \,\big|\, \varphi \in \widetilde{\mathcal{F}},\ \varphi \leqslant f\bigr\}.$$

💡 $\lfloor f \rfloor_{\mathcal{F}}$ is the pointwise supremum of all minorants of f taken from $\mathcal{F}$ and its constant shifts.

In short we write: $\lfloor f \rfloor_{\mathcal{F}} = \sup\bigl\{\varphi \,\big|\, \varphi \in \widetilde{\mathcal{F}},\ \varphi \leqslant f\bigr\}$

Some properties of $\mathcal{F}$-regularization

[Schiela, Herzog, RB, 2024]

1. $f \leqslant g$ and $\mathcal{F} \subseteq \mathcal{G}$ implies $\lfloor f \rfloor_{\mathcal{F}} \leqslant \lfloor g \rfloor_{\mathcal{G}}$.
2. For $\varphi \in \mathcal{F}$ and $c \in \mathbb{R}$ we have $\lfloor f + \varphi + c \rfloor_{\mathcal{F}} = \lfloor f \rfloor_{\mathcal{F}} + \varphi + c$.
3. $\lfloor f \rfloor_{\mathcal{F}} \leqslant f$, thus $f \leqslant \lfloor f \rfloor_{\mathcal{F}} \Leftrightarrow \lfloor f \rfloor_{\mathcal{F}} = f$
4. $f \in \mathcal{F} \Rightarrow \lfloor f \rfloor_{\mathcal{F}} = f$.
5. $\mathcal{F} \subseteq \mathcal{G}$ implies $\lfloor \lfloor f \rfloor_{\mathcal{G}} \rfloor_{\mathcal{F}} = \lfloor f \rfloor_{\mathcal{F}}$.
6. if $\mathcal{F}$ is a convex cone we obtain for $\alpha_1, \alpha_2 > 0$ and $f_1, f_2 \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ with $\lfloor f_1 \rfloor_{\mathcal{F}} \not\equiv -\infty$ and $\lfloor f_2 \rfloor_{\mathcal{F}} \not\equiv -\infty$ we obtain

$$\alpha_1 \lfloor f_1 \rfloor_{\mathcal{F}} + \alpha_2 \lfloor f_2 \rfloor_{\mathcal{F}} \leqslant \lfloor \alpha_1 f_1 + \alpha_2 f_2 \rfloor_{\mathcal{F}} \leqslant \alpha_1 f_1 + \alpha_2 f_2$$

Examples

1. If $\mathcal{M}$ is a locally convex linear topological space
   - $\mathcal{F} = \mathcal{M}^*$ is its topological dual space
   - $\widetilde{\mathcal{F}}$ is the space of all continuous affine functions
   - $\lfloor f \rfloor_{\mathcal{M}^*}$ is the pointwise supremum over all affine minorants of f.
2. Suppose that $\mathcal{M}$ is a metric space.
   - Then lower semi-continuous functions $f \in \mathcal{P}_\infty(\mathcal{M})$ can be written as the pointwise supremum of continuous functions
   - For $\mathcal{F} = C(\mathcal{M})$ the set sup-cl$(\mathcal{F}) \coloneqq \left\{ \lfloor f \rfloor_{\mathcal{F}} \,\middle|\, f \in \mathcal{P}_{\pm\infty}(\mathcal{M}) \right\}$ consists of the cone of lower semi-continuous functions in $\mathcal{P}_\infty(\mathcal{M})$
3. alternative generalization: the c-conjugate [Martínez-Legaz, 2005]

   For a coupling function $c\colon \mathcal{M} \times \mathcal{N} \to \mathbb{R}_{\pm\infty}$ defined as

$$f^c(y) \coloneqq \sup_{x \in \mathcal{M}} c(x,y) - f(x) \quad \text{for } y \in \mathcal{N}.$$

   This generalizes duality pairing instead of the set of test functions.

$\mathcal{F}$-biconjugates

[Schiela, Herzog, RB, 2024]

- We denote the restriction of the conjugate $f^{\circledast} \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ to $\mathcal{F}$ by

$$f^{\circledast}|_{\mathcal{F}} \colon \mathcal{F} \to \mathbb{R}_{\pm\infty}$$

- Let the evaluation (Dirac) functions be given by

$$\delta_x \colon \mathcal{P}_{\pm\infty}(\mathcal{M}) \to \mathbb{R}_{\pm\infty}, \qquad \varphi \mapsto \delta_x(\varphi) \coloneqq \varphi(x).$$

➜ $\delta_x|_{\mathcal{F}}$, $\mathcal{F} \subset \mathcal{P}_{\pm\infty}(\mathcal{M})$ linear, is a linear function and continuous.

Definition

Suppose that $\mathcal{F}$ is a linear subspace of $\mathcal{P}(\mathcal{M})$.
We define the $\mathcal{F}$-biconjugate $f_{\mathcal{F}}^{\circledast\circledast}$ of $f \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ as

$$f_{\mathcal{F}}^{\circledast\circledast} \colon \mathcal{M} \to \mathbb{R}_{\pm\infty}, \qquad x \mapsto (f^{\circledast}|_{\mathcal{F}})^{\circledast}(\delta_x).$$

Note. We employ the embedding of $\mathcal{M}$ into the dual space of $\mathcal{F}$ via

$$J_{\mathcal{M}\to\mathcal{F}'} \colon \mathcal{M} \to \mathcal{F}', \qquad x \mapsto \delta_x.$$

$\mathcal{F}$-biconjugate theorem

Remember.
For the classical Fenchel biconjugate the set $\mathcal{F}$ are all affine functions and $\lfloor f \rfloor_{\mathcal{F}}$ is largest convex lsc. minorant of $f \in \mathcal{P}_{\pm\infty}(V)$

Theorem [Schiela, Herzog, RB, 2024]

Suppose that $\mathcal{F}$ is a linear subspace of $\mathcal{P}(\mathcal{M})$. Thn, the $\mathcal{F}$-biconjugate satisfies $f_{\mathcal{F}}^{\circledast\circledast} = \lfloor f \rfloor_{\mathcal{F}}$ for all $f \in \mathcal{P}_{\pm\infty}(\mathcal{M})$.

➜ If $f = \lfloor f \rfloor_{\mathcal{F}}$, or in other words f agrees with the pointwise supremum of all minorants from $\mathcal{F}$, then we recover f from its $\mathcal{F}$-biconjugate.

Motivation: The subdifferential

With the Fenchel conjugate $f^*\colon V^* \to \mathbb{R}_{\pm\infty}$ of a proper, convex, lsc. function $f\colon V \to \mathbb{R}_{\pm\infty}$ on a vector space V we have

$$\xi \in \partial f(x) \quad \text{if and only if} \quad x \in \partial f^*(\xi)$$

➜ we can characterize both subdifferentials.

Nonlinear case.
We need "more structure on $\mathcal{M}$" to define a subdifferential of f.

In practice/numerics we use Riemannian manifolds.

In the following we consider a manifold $\mathcal{M}$, that is locally homeomorphic to a Banach space $\mathcal{X}$, or a Banach manifold for short.

The viscosity Fréchet Subdifferential

A function $f \in \mathcal{P}_{\pm\infty}(\mathcal{M})$ is lower semi-continuous at $x \in \mathcal{M}$ if,
$\forall \varepsilon > 0 \;\; \exists$ a neighbourhood $\mathcal{U}$ of x s.t. that $f(y) \geqslant f(x) - \varepsilon$ for all $y \in \mathcal{U}$.
We denote by $\operatorname{lsc}_\infty(\mathcal{M})$ the set of all functions that are lower semi-continuous at every $x \in \mathcal{M}$.

Definition

Suppose that $\mathcal{M}$ is a C^1-Banach manifold, $f \in \operatorname{lsc}_\infty(\mathcal{M})$, $x \in \mathcal{M}$ and $f(x) \neq +\infty$.
The (viscosity) Fréchet subdifferential $\partial_F f(x)$ of f is defined as follows:

$$\partial_F f(x) := \left\{ \varphi'(x) \,\middle|\, \varphi \in C^1(\mathcal{M}),\ f - \varphi \text{ attains a local minimum at } x \right\} \subseteq \mathcal{T}_x^* \mathcal{M},$$

where $\mathcal{T}_x^*\mathcal{M} := (\mathcal{T}_x\mathcal{M})^*$ denotes the cotangent space at x.
In case $f(x) = +\infty$, we set $\partial_F f(x) := \emptyset$.

Subdifferential Classification

Theorem [Schiela, Herzog, RB, 2024]

Suppose that $\mathcal{M}$ is a C^1-Banach manifold.
Let $x \in \mathcal{M}$, f be lower semicontinuous at every $x \in \mathcal{M}$ and $\varphi \in C^1(\mathcal{M})$.

1. *If $f^{\circledast}(\varphi) = \varphi(x) - f(x)$, i. e. we have equality in the Fenchel-Young inequality,*
   then $\varphi'(x) \in \partial_F f(x)$ and the Dirac function $\delta_x \in \partial(f^{\circledast}|_{C^1(\mathcal{M})})(\varphi)$.
2. *Conversely, if $\delta_x \in \partial(f^{\circledast}|_{C^1(\mathcal{M})})(\varphi)$, then $f^{\circledast}(\varphi) = \varphi(x) - f(x)$.*

Motivation: Infimal convolution

NTNU

Infimal convolution on a vector space $\mathcal{M} = V$ is defined as

$$(f \star_{\inf} g)(x) := \inf_{y \in \mathcal{M}} \{f(y) + g(x-y)\}\,.$$

The infimal convolution formula shows that [Prop. 13.21 Bauschke, Combettes, 2011]

$$(f \star_{\inf} g)^* = f^* + g^*$$

Nonlinear case.
We need even "slightly more structure" to generalise infimal convolution, a way to define "$x - y \in \mathcal{M}$" to be precise.

? Can we then get the same result for the nonlinear Fenchel conjugate? And what is a suitable restriction of $\mathcal{P}_{\pm\infty}(\mathcal{M})$ then?

Using Lie groups

Let

- $\mathcal{M}$ be a Riemannian manifold
- $\cdot\colon \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ be a smooth group operation
- ➜ $(\mathcal{M}, \cdot)$ is a Lie group.

We generalize infimal convolution to functions $f, g \in \mathcal{P}_\infty(\mathcal{M})$ as [Bachir, 2015]

$$(f \star_{\inf} g)(x) := \inf_{y \in \mathcal{M}} f(x \cdot y^{-1}) + g(y) = \inf_{z \in \mathcal{M}} f(z) + g(z^{-1} \cdot x).$$

Consider the linear space of group homomorphisms

$$\mathcal{H} := \operatorname{Hom}((\mathcal{M}, \cdot), (\mathbb{R}, +))$$

Then we get the relation [Schiela, Herzog, RB, 2024]

$$(f \star_{\inf} g)^{\circledast}(\varphi) = f^{\circledast}(\varphi) + g^{\circledast}(\varphi) \quad \text{for all } \varphi \in \mathcal{H}.$$

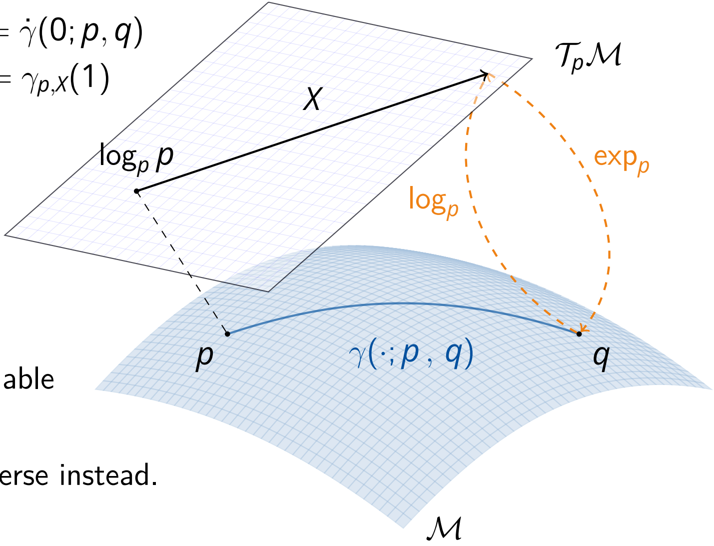
NTNU

The Riemannian Difference of Convex Algorithm

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 [Ferreira, 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 $\mathcal{T}_p^*\mathcal{M} \times \mathcal{T}_p\mathcal{M}$
- numerically we use musical isomorphisms $X = \xi^\flat \in \mathcal{T}_p\mathcal{M}$ to obtain a subset of $\mathcal{T}_p\mathcal{M}$

NTNU

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

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

$\Rightarrow$ 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, Souza, 2020; Souza, Oliveira, 2015]

In the Euclidean case, all three models are equivalent.

A Fenchel Duality on a Hadamard Manifold

NTNU

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, 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

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

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, Ferreira, Santos, 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, Ferreira, Santos, Souza, 2024]

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

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, Ferreira, Santos, 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, Ferreira, Santos, Souza, 2024]

Input: An initial point $p^{(0)} \in \operatorname{dom}(g)$, 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

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

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.

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, Ferreira, Santos, 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, Ferreira, Santos, 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 $\displaystyle\lim_{k\to\infty} d(p^{(k)}, p^{(k+1)}) = 0$.

A Numerical Example

The Difference of Convex Algorithm in Manopt.jl

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 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

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

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 $\mathbb{R}^2$

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

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

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 $\mathcal{M}$.

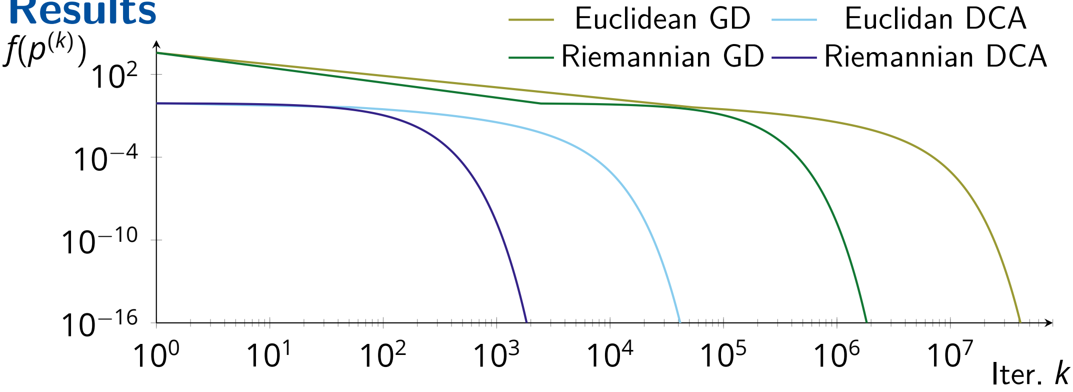
For DCA third we split f into $f(x) = g(x) - h(x)$ with

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

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

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

Nonlinear Fenchel Conjugate generalises the Fenchel conjugate.
A lot of properties can be proven more generally as well:

- Fenchel-Young inequality
- Biconjugate theorem
- Subdifferential classification on manifolds
- Infimal convolution on Lie groups

The Difference of Convex Algorithm to solve a nonsmooth, nonconvex problems on manifold of the form

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

- Relation to Fenchel Duality on Hadamard manifolds
- Convergence on Hadamard manifolds
- available in `Manopt.jl` for all manifolds form `Manifolds.jl`

Selected References

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

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

RB; R. Herzog; M. Silva Louzeiro; D. Tenbrinck; J. Vidal-Núñez (2021). "Fenchel duality theory and a primal-dual algorithm on Riemannian manifolds". *Foundations of Computational Mathematics* 21.6, pp. 1465–1504. DOI: 10.1007/s10208-020-09486-5. arXiv: 1908.02022.

De Carvalho Bento, G.; J. C. Neto; Í. D. L. Melo (2023). "Fenchel conjugate via Busemann function on Hadamard manifolds". *Applied Mathematics & Optimization* 88.3. DOI: 10.1007/s00245-023-10060-y.

Schiela, A.; R. Herzog; RB (2024). *Nonlinear Fenchel conjugates*. arXiv: 2409.04492.

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

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

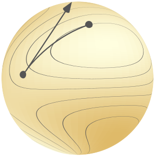
ronnybergmann.net/talks/2025-Saarbruecken-Nonlinear-Fenchel-Conjugates.pdf

NTNU

Norges teknisk-naturvitenskapelige universitet

More on Manopt.jl

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.

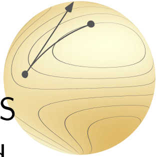
Manopt family.

manoptjl.org [RB, 2022]

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

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

List of Algorithms in Manopt.jl



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

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, Interior Point Newton

nonconvex Difference of Convex Algorithm, DCPPA

manoptjl.org/stable/solvers/