Nonlinear Fenchel conjugates

Ronny Bergmann
joint work with
Anton Schiela and Roland Herzog

MathICSE seminar

Lausanne, November 29, 2024.

The Fenchel Conjugate

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

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

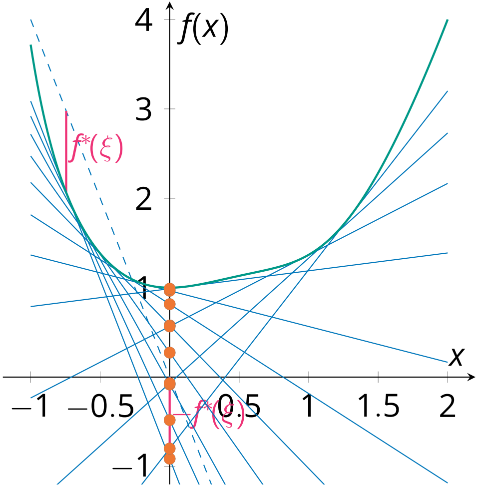
- given $\xi \in \mathbb{R}^n$: maximize the distance between $\xi^{\mathsf{T}}\cdot$ and f
- can also be written in the epigraph

The Fenchel biconjugate reads

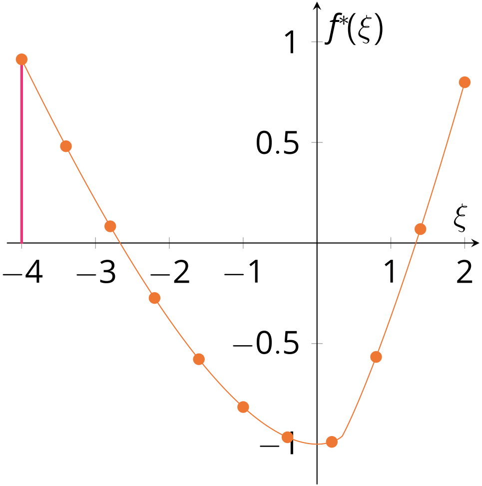
$$f^{**}(x) = (f^*)^*(x) = \sup_{\xi\in\mathbb{R}^n} \langle \xi\,, x\rangle - f^*(\xi).$$

Illustration of the Fenchel Conjugate

The function f



The Fenchel conjugate f^*



The Chambolle–Pock Algorithm

From the pair of primal-dual problems [Chambolle, 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

$$\begin{aligned} -K^*\hat{\xi} &\in \partial f(\hat{x}) \\ K\hat{x} &\in \partial g^*(\hat{\xi}) \end{aligned}$$

The Chambolle–Pock Algorithm

From the pair of primal-dual problems [Chambolle, 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

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

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,...

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.

The Nonlinear Fenchel Conjugate

[Schiela, Herzog, RB, 2024]

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

Use 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 $f, g \in \mathcal{P}_{\pm\infty}(\mathcal{M})$.

[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 saw K^* the adjoint or dual map 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

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

$$\lfloor f \rfloor_{\mathcal{F}}(x) := \sup\left\{\varphi(x) \,\middle|\, \varphi \in \widetilde{\mathcal{F}},\ \varphi \leqslant f\right\}.$$

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

In short we write: $\lfloor f \rfloor_{\mathcal{F}} = \sup\left\{\varphi \,\middle|\, \varphi \in \widetilde{\mathcal{F}},\ \varphi \leqslant f\right\}$

Some properties of $\mathcal{F}$ regilarization

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. For $\mathcal{M}$ 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.
   - choosing $\mathcal{F} = C(\mathcal{M})$, $\text{sup-cl}(\mathcal{F}) \coloneqq \left\{ \lfloor f \rfloor_{\mathcal{F}} \,\middle|\, f \in \mathcal{P}_{\pm\infty}(\mathcal{M}) \right\}$ consists of the cone of lower semi-continuous functions in $\mathcal{P}_\infty(\mathcal{M})$
3. alternate 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

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

   Generalizes duality pairing instead of the set of test functions.

$\mathcal{F}$ biconjugates

- 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
$$\delta_x \colon \mathcal{P}_{\pm\infty}(\mathcal{M}) \to \mathbb{R}_{\pm\infty}, \qquad \varphi \mapsto \delta_x(\varphi) \coloneqq \varphi(x).$$
- The restriction $\delta_x|_{\mathcal{F}}$ to $\mathcal{F}$ is a linear function on $\mathcal{F}$ and continuous.

Definition

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

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

Note. We employ theembedding 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
Suppose that $\mathcal{F}$ is a linear subspace of $\mathcal{P}(\mathcal{M})$. 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

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

can define the subdifferential.

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

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

Infimal convolution 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.
With 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 group operation
- $(\mathcal{M}, \cdot)$ is a Lie group.

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

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

Consider the linear space of group homomorphisms

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

Then the get the relation

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

Chambolle-Pock algorithm

Special case: Test functions on $\mathcal{T}_x\mathcal{M}$

For a $x \in \mathcal{M}$ consider a neighbourhood V of the origin in the tangent space $\mathcal{T}_x\mathcal{M}$ on which the exponential map $\exp_x$ is a diffeomorphism to $\mathcal{V} := \exp_x(V) \subseteq \mathcal{M}$.

As set of test functions we use [Ahmadi Kakavandi, Amini, 2010; Silva Louzeiro, RB, Herzog, 2022]

$$\mathcal{F}_x := \left\{ x^* \circ \exp_x^{-1} \in C^\infty(\mathcal{V}, \mathbb{R}) \,\middle|\, x^* \in \mathcal{T}_x^*\mathcal{M} \right\}$$

We also consider a localised version of the nonlinear conjugate

$$(f + \iota_{\mathcal{V}})^{\circledast}(\varphi) = \sup_{y \in \mathcal{V}} \{ \varphi(y) - f(y) \} \quad \text{for } \varphi \in \mathcal{F}_x.$$

This indeed agrees with the classical Fenchel conjugate on the tangent space as $f_x(x^*) := (f \circ \exp_x + \iota_V)^*(x^*)$

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, Vidal-Núñez, 2021; Chambolle, 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 σ, τ, $\theta > 0$

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

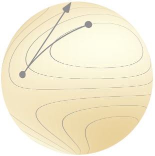
6: $\quad \bar{p}^{(k+1)} \leftarrow \exp_{p^{(k+1)}}\bigl(-\theta \log_{p^{(k+1)}} p^{(k)}\bigr)$

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

8: **end while**

Output: $p^{(k)}$

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.

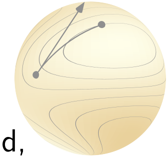
Manopt family.

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

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/

Riemannian Chambolle-Pock in Manopt.jl

To call the exact Riemannian Chambolle-Pock algorithm in `Manopt.jl`:

`ChambollePock(M, N, F, p, X, m, n, prox_f, prox_g_n,` $D\Lambda^*$`; kwargs..`

- `M, N` are the manifolds f and g, resp., are defined on
- `F` is the objective function $f+g$
- `p,n,m` are the initial, Fenchel conjugate base, and linearization point, resp.
- `X` is the initial tangent vector
- `prox_f, prox_g_n` are the proximal maps of f and g_n^*, resp.
- $D\Lambda^*$ is the adjoint of the linearization of Λ

Summary

The 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
- Infimal convolution

➜ Unified framework for the existing generalisations
and hence for nonsmooth optimization on Riemannian manifolds.

Example Chambolle-Pock algorithm on Riemannian manifolds
and its implementation in `Manopt.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; 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.

Boumal, N.; B. Mishra; P.-A. Absil; R. Sepulchre (2014). "Manopt, a Matlab toolbox for optimization on manifolds". *Journal of Machine Learning Research* 15, pp. 1455–1459. URL: https://www.jmlr.org/papers/v15/boumal14a.html.

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.

Townsend, J.; N. Koep; S. Weichwald (2016). *Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation*. arXiv: 1603.03236.

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

ronnybergmann.net/talks/2024-Lausanne-Nonlinear-Fenchel-Conjugates.pdf