Fenchel Duality Theory on Riemannian Manifolds

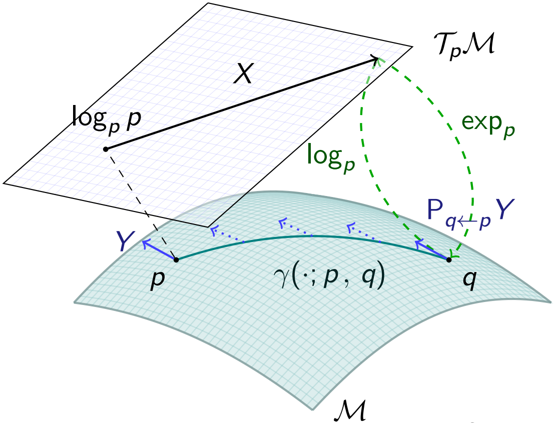
Ronny Bergmann

joint work with
Roland Herzog, Maurício Silva Louzeiro, Daniel Tenbrinck, José Vidal-Núñez.

IFIP TC 7, 2021
Minisymposium: Optimisation & Manifolds
Quito, Ecuador & virtual
August 28 – September 3, 2021

Norwegian University of Science and Technology

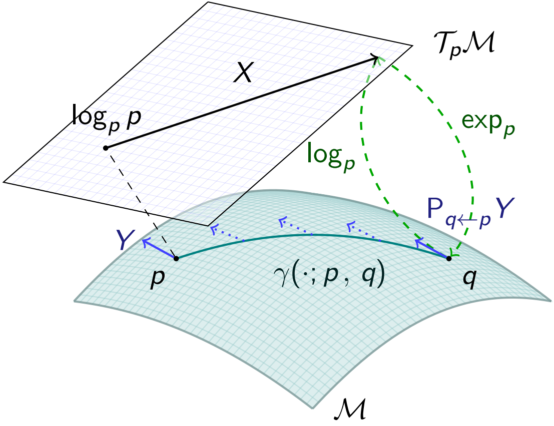
A d-dimensional Riemannian manifold $\mathcal{M}$



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 continuous varying inner product on the tangent spaces.

[Absil, Mahony, and Sepulchre 2008]

A d-dimensional Riemannian manifold $\mathcal{M}$



Geodesic $\gamma(\cdot; p, q)$

a shortest path between $p, q \in \mathcal{M}$

Tangent space $\mathcal{T}_p\mathcal{M}$ at p

with inner product $(\cdot, \cdot)_p$

Logarithmic map $\log_p q = \dot\gamma(0; p, q)$

"speed towards q"

Exponential map $\exp_p X = \gamma_{p,X}(1)$,

where $\gamma_{p,X}(0) = p$ and $\dot\gamma_{p,X}(0) = X$

Parallel transport $\mathrm{P}_{q\leftarrow p}Y$

from $\mathcal{T}_p\mathcal{M}$ along $\gamma(\cdot; p, q)$ to $\mathcal{T}_q\mathcal{M}$

The Model

We consider a minimization problem

$$\operatorname*{arg\,min}_{p\in\mathcal{C}} F(p) + G(\Lambda(p))$$

- $\mathcal{M}, \mathcal{N}$ are (high-dimensional) Riemannian Manifolds
- $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ nonsmooth, (locally, geodesically) convex
- $G\colon \mathcal{N} \to \overline{\mathbb{R}}$ nonsmooth, (locally) convex
- $\Lambda\colon \mathcal{M} \to \mathcal{N}$ nonlinear
- $\mathcal{C} \subset \mathcal{M}$ strongly geodesically convex.

➔ In image processing:
choose a model, such that finding a minimizer yields the reconstruction

Splitting Methods & Algorithms

On a Riemannian manifold $\mathcal{M}$ we have

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

On $\mathbb{R}^n$ PDRA is known to be equivalent to [O'Connor and Vandenberghe 2018; Setzer 2011]

- 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]

But on a Riemannian manifold $\mathcal{M}$: ⚠ no duality theory!

Goals of this talk.

Formulate Duality (dualities) on a Manifold

To cover different properties.

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 Subdifferential

[Lee 2003; Udriște 1994]

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

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

where

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

The Euclidean Fenchel Conjugate

Let $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex.
We define the Fenchel conjugate $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ of f by

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

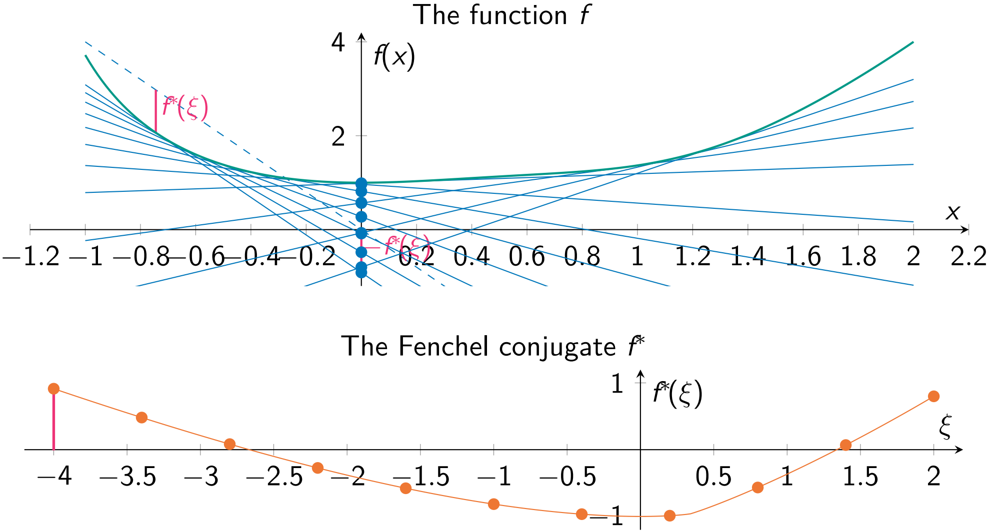
- interpretation: maximize the distance of $\xi^{\mathrm{T}} x$ to f

$\Rightarrow$ extremum seeking problem on 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



Properties of the Fenchel Conjugate

[Rockafellar 1970]

- The Fenchel conjugate f^* is convex (even if f is not)
- If $f(x) \leq g(x)$ holds for all $x \in \mathbb{R}^n$ then $f^*(\xi) \geq g^*(\xi)$ holds for all $\xi \in \mathbb{R}^n$
- If $g(x) = f(x + b)$ for some $b \in \mathbb{R}$ holds for all $x \in \mathbb{R}^n$
  then $g^*(\xi) = f^*(\xi) - \xi^\mathsf{T} b$ holds for all $\xi \in \mathbb{R}^n$
- If $g(x) = \lambda f(x)$, for some $\lambda > 0$, holds for all $x \in \mathbb{R}^n$
  then $g^*(\xi) = \lambda f^*(\xi/\lambda)$ holds for all $\xi \in \mathbb{R}^n$
- f^{**} is the largest convex, lsc function with $f^{**} \leq f$
- especially the Fenchel–Moreau theorem:
  f convex, proper, lsc $\Rightarrow f^{**} = f$.

Properties of the Fenchel Conjugate II

The Fenchel–Young inequality holds, i.e.,

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

We can characterize subdifferentials

- 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]
alternative approach: [Ahmadi Kakavandi and Amini 2010]

Idea: Introduce a point on $\mathcal{M}$ to "act as" 0.
Let $m \in \mathcal{C} \subset \mathcal{M}$ be given and $F\colon \mathcal{C} \to \overline{\mathbb{R}}$.
The m-Fenchel conjugate $F_m^*\colon \mathcal{T}_m^*\mathcal{M} \to \overline{\mathbb{R}}$ is defined by

$$F_m^*(\xi_m) \coloneqq \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \bigl\{\langle \xi_m\,, X\rangle - F(\exp_m X)\bigr\},$$

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

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

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

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 $F(p) \leq G(p)$ holds for all $p \in \mathcal{C}$
  then $F_m^*(\xi_m) \geq G_m^*(\xi_m)$ holds for all $\xi_m \in \mathcal{T}_m^*\mathcal{M}$
- If $G(p) = F(p) + a$ for some $a \in \mathbb{R}$ holds for all $p \in \mathcal{C}$
  then $G_m^*(\xi_m) = F_m^*(\xi_m) - a$ holds for all $\xi_m \in \mathcal{T}_m^*\mathcal{M}$
- If $G(p) = \lambda F(p)$, for some $\lambda > 0$, holds for all $p \in \mathcal{C}$
  then $G_m^*(\xi_m) = \lambda F_m^*(\xi_m/\lambda)$ holds for all $\xi_m \in \mathcal{T}_m^*\mathcal{M}$
- It holds $F_{mm}^{**} \leq F$ on $\mathcal{C}$
- especially the Fenchel-Moreau theorem:
  If $F \circ \exp_m$ convex (on $\mathcal{T}_m\mathcal{M}$), proper, lsc, then $F_{mm}^{**} = F$ on $\mathcal{C}$.

Properties of the m-Fenchel Conjugate II

The Fenchel–Young inequality holds, i.e.,

$$F(p) + F_m^*(\xi_m) \geq \langle \xi_m \,, \log_m p \rangle \quad \text{for all} \quad p \in \mathcal{C}, \xi_m \in \mathcal{T}_m^*\mathcal{M}$$

We can characterize subdifferentials

- 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 $F \circ \exp_m$

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

Saddle Point Formulation

Let F be geodesically convex, $G \circ \exp_n$ be convex (on $\mathcal{T}_n\mathcal{N}$).

From

$$\min_{p\in\mathcal{C}} F(p) + G(\Lambda(p))$$

we derive the saddle point formulation for the n-Fenchel conjugate of G as

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

But $\Lambda\colon \mathcal{M} \to \mathcal{N}$ is a non-linear operator!

For Optimality Conditions and the Dual Prolem: What's Λ^*?

Approach. Linearization:

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

[Valkonen 2014]

Optimality Conditions for the Saddle Point Problem

The first order opimality conditions for a saddle point of the exact saddle point problem

$$(\widehat{p}, \widehat{\xi}_n) \in \mathcal{C} \times \mathcal{T}^*_n\mathcal{N}$$

can be formally derived as

$$\begin{aligned} D * \Lambda(\widehat{p})\big[D * \log_n(\Lambda(\widehat{p}))[\widehat{\xi}_n]\big] &\in \partial_{\mathcal{M}} F(\widehat{p}) \\ \log_n \Lambda(\hat{p}) &\in \partial G^*_n(\widehat{\xi}_n) \end{aligned}$$

Advantage. By only linearizing for the adjoint, we stay closer to the original problem.

The m-Fenchel Conjuagte (I) – Summary

- most properties carry over
- exception a shift property $g(x) = f(x + b)$ which depends on linearity
- yields a Riemannian Chambolle–Pock algorithm

But! We need convexity of $F \circ \exp_m$ for Fenchel Moreau.

NTNU

The Riemannian Fenchel Conjugate (II)

[RB, Herzog, and Silva Louzeiro 2021]

Let $\mathcal{M}$ be a Hadamard manifold and $F\colon \mathcal{M} \to \overline{\mathbb{R}}$.
The Fenchel conjugate of F is the function $F^*\colon \mathcal{T}^*\mathcal{M} \to \overline{\mathbb{R}}$ defined by

$$F^*(p,\xi) := \sup_{q\in\mathcal{M}} \bigl\{\langle \xi\,,\log_p q\rangle - F(q)\bigr\} \quad \text{for } (p,\xi) \in \mathcal{T}^*\mathcal{M}.$$

and the biconjugate

$$F^{**}(p) := \sup_{(q,\xi)\in\mathcal{T}^*\mathcal{M}} \bigl\{\langle \xi\,,\log_q p\rangle - F^*(q,\xi)\bigr\} \quad \text{for } p \in \mathcal{M}.$$

Remarks on the Alternate Definition

- The domain is now the whole cotangent bundle $\mathcal{T}^*\mathcal{M}$.

Theorem (Fenchel-Moreau-Theorem) *[RB, Herzog, and Silva Louzeiro 2021]*

Let $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ be a proper lsc convex function. Then $F^{**} = F$ holds.

Remarks on the Alternate Definition

- The domain is now the whole cotangent bundle $\mathcal{T}^*\mathcal{M}$.
- At first glance the dimension doubles – and is reduced again for the biconjugate

Theorem (Fenchel-Moreau-Theorem) *[RB, Herzog, and Silva Louzeiro 2021]*

Let $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ *be a proper lsc convex function. Then* $F^{**} = F$ *holds.*

Remarks on the Alternate Definition

- The domain is now the whole cotangent bundle $\mathcal{T}^*\mathcal{M}$.
- At first glance the dimension doubles – and is reduced again for the biconjugate
- using the congruence relation

$$(p,\xi) \sim (p',\xi') \quad \text{if and only if} \quad \langle \xi\,, \log_p q\rangle = \langle \xi'\,, \log_{p'} q\rangle \text{ holds for all } q \in \mathcal{M}$$

seems to reduce the dimension again (F^* is constant on $[(p,\xi)]$

Theorem (Fenchel-Moreau-Theorem) [RB, Herzog, and Silva Louzeiro 2021]

Let $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ be a proper lsc convex function. Then $F^{**} = F$ holds.

Remarks on the Alternate Definition

- The domain is now the whole cotangent bundle $\mathcal{T}^*\mathcal{M}$.
- At first glance the dimension doubles – and is reduced again for the biconjugate
- using the congruence relation

$$(p, \xi) \sim (p', \xi') \quad \text{if and only if} \quad \langle \xi \,, \log_p q \rangle = \langle \xi' \,, \log_{p'} q \rangle \text{ holds for all } q \in \mathcal{M}$$

  seems to reduce the dimension again (F^* is constant on $[(p, \xi)]$

$\Rightarrow$ On $\mathbb{R}^n$: $(p, \xi) \sim (p', \xi') \Leftrightarrow \langle \xi \,, p' \rangle = \langle \xi \,, p \rangle$
$\Rightarrow$ we obtain $F^*(\xi)$ as expected.

Theorem (Fenchel-Moreau-Theorem) *[RB, Herzog, and Silva Louzeiro 2021]*

Let $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ be a proper lsc convex function. Then $F^{**} = F$ holds.

Remarks on the Alternate Definition

- The domain is now the whole cotangent bundle $\mathcal{T}^*\mathcal{M}$.
- At first glance the dimension doubles – and is reduced again for the biconjugate
- using the congruence relation

$$(p, \xi) \sim (p', \xi') \quad \text{if and only if} \quad \langle \xi \,, \log_p q \rangle = \langle \xi' \,, \log_{p'} q \rangle \text{ holds for all } q \in \mathcal{M}$$

seems to reduce the dimension again (F^* is constant on $[(p, \xi)]$

$\Rightarrow$ On $\mathbb{R}^n$: $(p, \xi) \sim (p', \xi') \Leftrightarrow \langle \xi \,, p' \rangle = \langle \xi \,, p \rangle$
$\Rightarrow$ we obtain $F^*(\xi)$ as expected.

- the "pointwise", Fenchel-Young properties carry over (for fixed p).

Theorem (Fenchel-Moreau-Theorem) *[RB, Herzog, and Silva Louzeiro 2021]*

Let $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ be a proper lsc convex function. Then $F^{**} = F$ holds.

Remarks on the Alternate Definition

- The domain is now the whole cotangent bundle $\mathcal{T}^*\mathcal{M}$.
- At first glance the dimension doubles – and is reduced again for the biconjugate
- using the congruence relation

$$(p,\xi) \sim (p',\xi') \quad \text{if and only if} \quad \langle \xi\,, \log_p q\rangle = \langle \xi'\,, \log_{p'} q\rangle \text{ holds for all } q \in \mathcal{M}$$

seems to reduce the dimension again (F^* is constant on $[(p,\xi)]$

$\Rightarrow$ On $\mathbb{R}^n$: $(p,\xi) \sim (p',\xi') \Leftrightarrow \langle \xi\,, p'\rangle = \langle \xi\,, p\rangle$
$\Rightarrow$ we obtain $F^*(\xi)$ as expected.

- the “pointwise”, Fenchel-Young properties carry over (for fixed p).
- Subdifferential property slightly changes: $\xi \in \partial F(p) \Leftrightarrow F^*(p,\xi) = -F(p)$.

Theorem (Fenchel-Moreau-Theorem) *[RB, Herzog, and Silva Louzeiro 2021]*

Let $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ be a proper lsc convex function. Then $F^{**} = F$ holds.

A comparison for the Translation property

We can not generalize

$$g(x) = f(x+b) \text{ for all } x \Rightarrow g^*(\xi) = f^*(\xi) - \xi^{\mathsf{T}} b \text{ for all } \xi$$

from $\mathbb{R}^n$ to (Hadamard) manifolds,
since the translation is “encoded into” both definitions:

For $\mathcal{M} = \mathbb{R}^n$ we get in both definitions

- $F_m^*(\xi_m) = F_0^*(\xi_m) - \langle \xi_m \,, m \rangle = F^*(\xi_m) - \langle \xi_m \,, m \rangle$
- $F^*(\xi_m, m) = F^*(\xi_m, 0) - \langle \xi_m \,, m \rangle = F^*(\xi_m) - \langle \xi_m \,, m \rangle$

Summary and Outlook

Summary.

- We introduced two frameworks for Fenchel duality on Riemannian manifolds
- The first yields a Riemannian Chambolle–Pock Algorithm
- Tue @ 15:15 BST (22:15 CEST) in MS Non-Smooth First-order Methods, Convex, and Non-convex
- ! Fenchel-Moreau depends on convexity of $F \circ \exp_m$
- The second duality yields a (geodesically) convex Fenchel-Moreau Theorem
- ! At first glance doubles dimension of the Domain for the Dual

Outlook.

- investigate equivalence classes
- derive a Riemannian Chambolle–Pock algorithm for the second defintion
- investigate further properties and algorithms

Selected References

- Absil, P.-A., R. Mahony, and R. Sepulchre (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. DOI: 10.1515/9781400830244.
- Ahmadi Kakavandi, B. and M. Amini (Nov. 2010). "Duality and subdifferential for convex functions on complete metric spaces". In: *Nonlinear Analysis: Theory, Methods & Applications* 73.10, pp. 3450–3455. DOI: 10.1016/j.na.2010.07.033.
- RB, R. Herzog, and M. Silva Louzeiro (2021). *Fenchel duality and a separation theorem on Hadamard manifolds*. arXiv: 2102.11155.
- 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*. DOI: 10.1007/s10208-020-09486-5.
- Rockafellar, R. T. (1970). *Convex Analysis*. Vol. 28. Princeton Mathematical Series. Princeton, New Jersey: Princeton University Press. URL: https://www.jstor.org/stable/j.ctt14bs1ff.
- Valkonen, T. (2014). "A primal–dual hybrid gradient method for nonlinear operators with applications to MRI". In: *Inverse Problems* 30.5, p. 055012. DOI: 10.1088/0266-5611/30/5/055012.

ronnybergmann.net/talks/2021-IFIP-Fenchel-Duality.pdf