NTNU
Norwegian University of Science and Technology

Fenchel Duality Theory and a Primal-Dual Algorithm on Riemannian Manifolds

Ronny Bergmann

joint work with
R. Herzog, M. Silva Louzeiro, D. Tenbrinck, J. Vidal-Núñez.

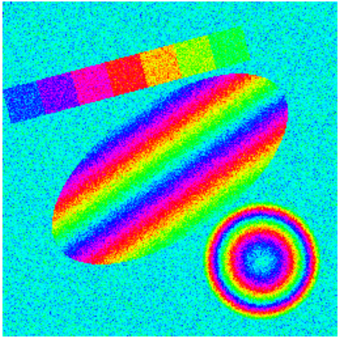
40nt International Conference on Machine Learning
Workshop on Duality Principles for Modern Machine Learning

Honolulu, Hawaii, July 29, 2023
travel support by RIKEN

Manifold-valued Signal & Image Processing

Tasks in image processing are often phrased as an optimisation problem.
Here. The pixel take values on a manifold

- phase-valued data ($\mathbb{S}^1$)
- wind-fields, GPS ($\mathbb{S}^2$)
- DT-MRI ($\mathcal{P}(3)$)
- EBSD, (grain) orientations ($\mathrm{SO}(n)$)



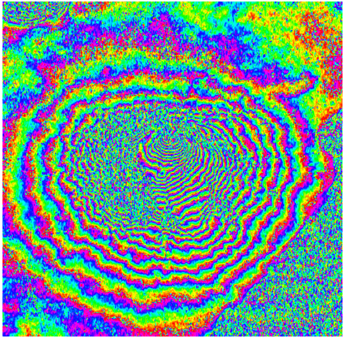
Artificial noisy phase-valued data.

Tasks. Denoising, Inpainting, labeling (classification), deblurring,...

Manifold-valued Signal & Image Processing

Tasks in image processing are often phrased as an optimisation problem.
Here. The pixel take values on a manifold

- phase-valued data ($\mathbb{S}^1$)
- wind-fields, GPS ($\mathbb{S}^2$)
- DT-MRI ($\mathcal{P}(3)$)
- EBSD, (grain) orientations ($\mathrm{SO}(n)$)



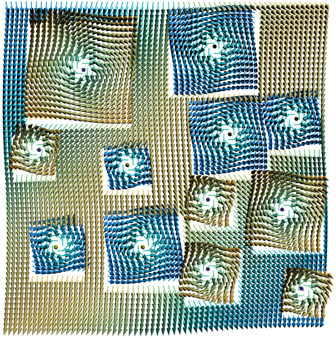
InSAR-Data of Mt. Vesuvius.
[Rocca, Prati, and Guarnieri 1997]

Tasks. Denoising, Inpainting, labeling (classification), deblurring,...

Manifold-valued Signal & Image Processing

Tasks in image processing are often phrased as an optimisation problem.
Here. The pixel take values on a manifold

- phase-valued data ($\mathbb{S}^1$)
- wind-fields, GPS ($\mathbb{S}^2$)
- DT-MRI ($\mathcal{P}(3)$)
- EBSD, (grain) orientations ($\mathrm{SO}(n)$)



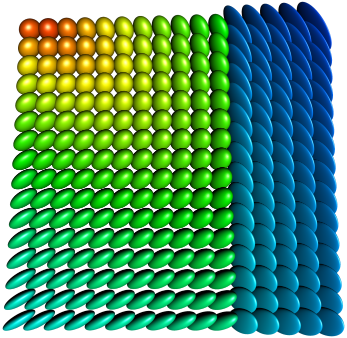
Artificial noisy data on the sphere $\mathbb{S}^2$.

Tasks. Denoising, Inpainting, labeling (classification), deblurring,...

Manifold-valued Signal & Image Processing

Tasks in image processing are often phrased as an optimisation problem.
Here. The pixel take values on a manifold

- phase-valued data ($\mathbb{S}^1$)
- wind-fields, GPS ($\mathbb{S}^2$)
- DT-MRI ($\mathcal{P}(3)$)
- EBSD, (grain) orientations ($\mathrm{SO}(n)$)



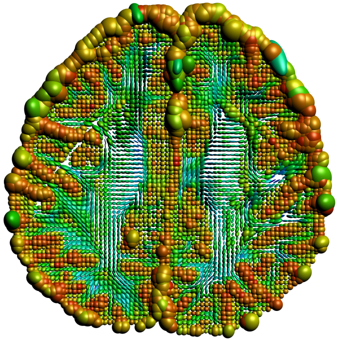
Artificial diffusion data,
each pixel is a symmetric positive matrix.

Tasks. Denoising, Inpainting, labeling (classification), deblurring,...

Manifold-valued Signal & Image Processing

Tasks in image processing are often phrased as an optimisation problem.
Here. The pixel take values on a manifold

- phase-valued data ($\mathbb{S}^1$)
- wind-fields, GPS ($\mathbb{S}^2$)
- DT-MRI ($\mathcal{P}(3)$)
- EBSD, (grain) orientations ($\mathrm{SO}(n)$)



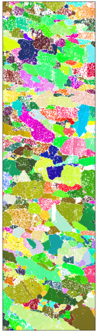
DT-MRI of the human brain.
Camino Profject: cmic.cs.ucl.ac.uk/camino

Tasks. Denoising, Inpainting, labeling (classification), deblurring,...

Manifold-valued Signal & Image Processing

Tasks in image processing are often phrased as an optimisation problem.
Here. The pixel take values on a manifold

- phase-valued data ($\mathbb{S}^1$)
- wind-fields, GPS ($\mathbb{S}^2$)
- DT-MRI ($\mathcal{P}(3)$)
- EBSD, (grain) orientations ($\mathrm{SO}(n)$)



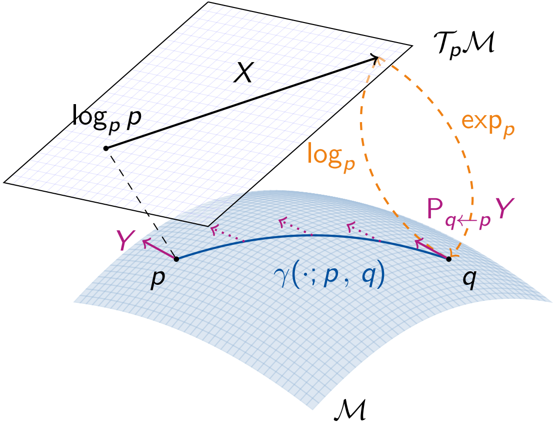
Grain orientations in EBSD data.
MTEX toolbox: mtex-toolbox.github.io

Tasks. Denoising, Inpainting, labeling (classification), deblurring,...

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]



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 $(\cdot\,,\,\cdot)_p$
- parallel transport $\mathcal{P}_{q\leftarrow p}X$

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 [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]

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

Goals of this talk.
Formulate Duality on a Manifold
Derive a Riemannian Chambolle–Pock Algorithm (RCPA)

The Euclidean Fenchel Conjugate

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

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

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

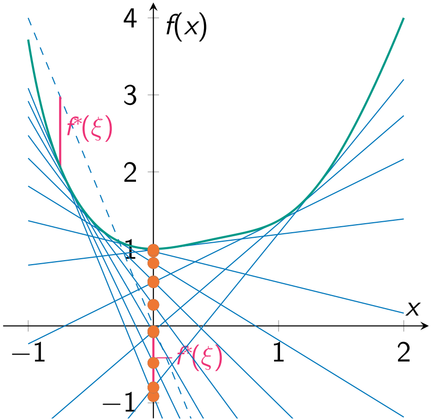
$\Rightarrow$ extremum seeking problem on the epigraph

The Fenchel biconjugate reads

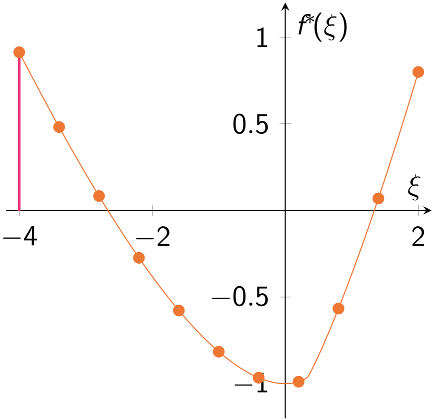
$$f^{**}(x) = (f^*)^*(x) = \sup_{\xi\in\mathbb{R}^n} \langle \xi, x\rangle - f^*(\xi).$$

Illustration of the Fenchel Conjugate

The function f



The Fenchel conjugate f^*



The Riemannian $m-$Fenchel Conjugate

[RB, Herzog, Silva Louzeiro, Tenbrinck, and Vidal-Núñez 2021]
alternative approaches: [Ahmadi Kakavandi and Amini 2010; Silva Louzeiro, RB, and Herzog 2022]

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

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

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

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

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

Properties of the m-Fenchel Conjugate

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

- F_m^* is convex on $\mathcal{T}_m^*\mathcal{M}$
- $F(p) \leq G(p)$ 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, then $F_{mm}^{**} = F$ 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

Proximal Map

For $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ and $\lambda > 0$ we define the Proximal Map as
[Moreau 1965; Rockafellar 1970; Ferreira and Oliveira 2002]

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

! For a minimizer u^* of f we have $\operatorname{prox}_{\lambda f}(u^*) = u^*$.

- For f proper, convex, lsc:
  - the proximal map is unique.
  - Proximal-Point-Algorithm:
    $p_k = \operatorname{prox}_{\lambda f}(p_{k-1})$ converges to $\operatorname{arg\,min} f$

The Chambolle-Pock Algorithm

[Chambolle and Pock 2011]

From the pair of primal-dual problems

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

we obtain for f, g proper convex, lsc the
optimality conditions (OC) for a solution $(\hat{x}, \hat{\xi})$ as ,
Chambolle–Pock Algorithm. with $\sigma > 0$, $\tau > 0$, $\theta \in \mathbb{R}$ reads

$$\begin{aligned}
\partial f \quad &\ni -K^*\hat{\xi} \\
\partial g^*(\hat{\xi}) &\ni K\hat{x} \\
\bar{\xi}^{(k+1)} &= \xi^{(k+1)} + \theta(\xi^{(k+1)} - \xi^{(k)})
\end{aligned}$$

The Exact Riemannian Chambolle–Pock Algorithm

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

Assume. $f(p) = F(p) + G(\Lambda(p))$, with $\Lambda\colon \mathcal{M} \to \mathcal{N}$.

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

$k \leftarrow 0$

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

while not converged **do**

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

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

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

$\quad k \leftarrow k + 1$

end while

Output: $p^{(k)}$

Generalizations & Variants of the RCPA

Classically

[Chambolle and Pock 2011]

- change $\sigma = \sigma_k$, $\tau = \tau_k$, $\theta = \theta_k$ during the iterations
- introduce an acceleration γ
- relax dual $\bar{\xi}$ instead of primal $\bar{p}$ (switches lines 4 and 5)

Furthermore we

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

- introduce the IRCPA: linearize Λ, i. e., adopt the Euclidean case from [Valkonen 2014]

$$\log_n \Lambda(\bar{p}^{(k)}) \quad \to \quad \mathrm{P}_{n \leftarrow \Lambda(m)} D\Lambda(m)[\log_m \bar{p}^{(k)}]$$

- choose $n \neq \Lambda(m)$ introduces a parallel transport

$$D\Lambda(m)^*[\xi_n^{(k+1)}] \quad \to \quad D\Lambda(m)^*[\mathrm{P}_{\Lambda(m) \leftarrow n} \xi_n^{(k+1)}]$$

- change $m = m^{(k)}$, $n = n^{(k)}$ during the iterations

ManifoldsBase.jl & Manifolds.jl

`ManifoldsBase.jl` is an interface for Riemannian manifolds `M`

- `inner(M, p, X, Y)` $(X, Y)_p$
- `exp(M, p, X)` and `log(M, p, q)`,
- more general:
  `retract(M, p, X, m)`,
  where `m` is a retraction method
- embeddings as decorator
- ☺ mutating variants, e. g.
  `exp!(M, q, p, X)`
  works in place of `q`

🔗 juliamanifolds.github.io/ManifoldsBase.jl/
🔗 juliamanifolds.github.io/Manifolds.jl/

`Manifolds.jl` is a Library of manifolds

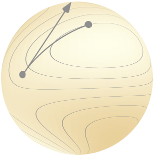
- Circle, (unit) Sphere & Torus
- Fixed Rank Matrices
- (Symplectic) Stiefel & Grassmann
- Hyperbolic space & Rotations
- Symmetric positive definite matrices
- ...and many more

as well as generically

- power & product manifold
- tangent & vector bundles
- Lie groups, connections, metrics,...

[Axen, Baran, RB, and Rzecki 2021]

Manopt.jl: Optimisation on Manifolds in Julia



Goal. Optimisation algorithms on Riemannian manifolds, based on `ManifoldsBase.jl` ⇒ works with any manifold from `Manifolds.jl`.

Features.

- generic algorithm framework:
  With `Problem p` and a `SolverState s`
  - `initialize_solver!(p, s)`
  - `step_solver!(p, s, i)`: ith step
- run algorithm: call `solve(p, s)`
- generic debug and recording
- step sizes and stopping criteria.

Manopt Family.

- manoptjl.org [RB 2022]
- manopt.org [Boumal, Mishra, Absil, and Sepulchre 2014]
- pymanopt.org [Townsend, Koep, and Weichwald 2016]

Algorithms.

- Nelder-Mead, Particle Swarm
- Subgradient Method
- Gradient Descent
  CG, Stochastic, Momentum, …
- Quasi-Newton
  BFGS, DFP, Broyden, SR1, …
- Trust Regions
- Chambolle-Pock
- Douglas-Rachford, CPPA
- ALM, EPM, Frank-Wolfe,…
- Difference of Convex
  DCA, DCPPA

NTNU

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 $f \in \mathcal{M}$, $\mathcal{M} = \mathcal{N}^{d_1,d_2}$, we compute

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} \frac{1}{\alpha}F(p) + G(\Lambda(p)), \qquad \alpha > 0,$$

with

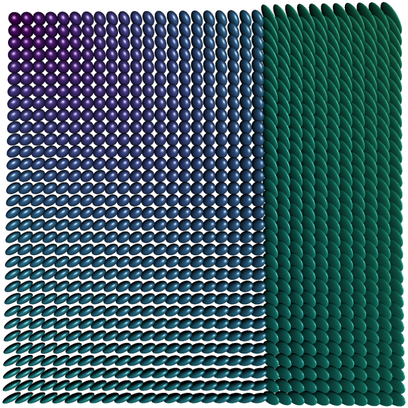
- data term $F(p) = \frac{1}{2}d^2_{\mathcal{M}}(p,f)$
- "forward differences" $\Lambda\colon \mathcal{M} \to (T\mathcal{M})^{d_1-1,\ d_2-1,\ 2}$,

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

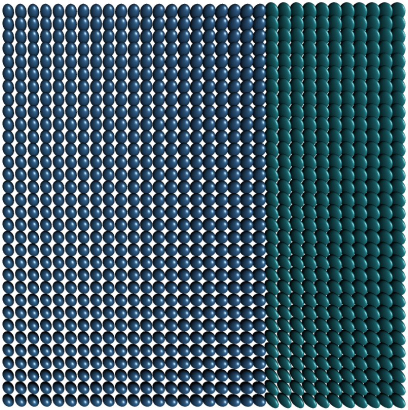
- prior $G(X) = \|X\|_{g,q,1}$ similar to a collaborative TV [Duran, Moeller, Sbert, and Cremers 2016]

$\Rightarrow$ $\operatorname{prox}_{\lambda G_n^*}$ given in closed form for $q = 1$ (anisotropic TV) and $q = 2$ (isotropic TV).

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



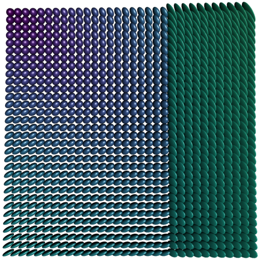
$\mathcal{P}(3)$-valued data.



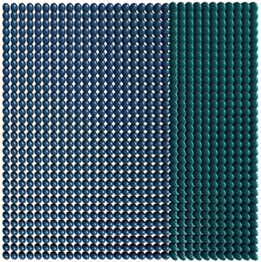
anisotropic TV, $\alpha = 6$.

- 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.

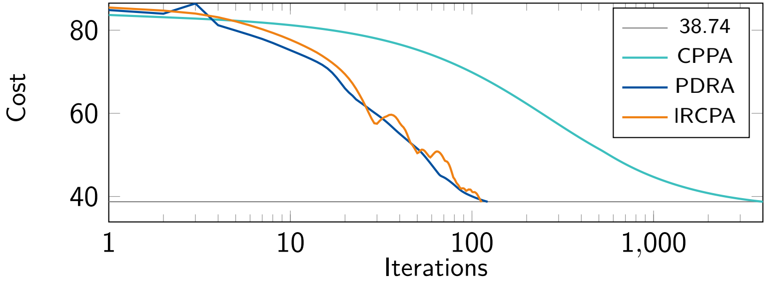


anisotropic TV, $\alpha = 6$.

Approach. CPPA as benchmark [Bačák 2014; RB, 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, 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.**

Summary

Summary.

- We introduced a duality framework on manifolds
- we introduced a Riemannian Chambolle–Pock algorithm
- We saw a Software framework for Optimisation algorithms on manifolds
- Numerical examples illustrates its performance
- Another model works with both functions being geodesically convex [Silva Louzeiro, RB, and Herzog 2022]

Outlook.

- Explore further areas where Duality can be used in non-Euclidean spaces
- Explore further connections between Duality-based algorithms
- look into further applications

Selected References

Axen, S. D., M. Baran, RB, and K. Rzecki (2021). *Manifolds.jl: An Extensible Julia Framework for Data Analysis on Manifolds*. arXiv: 2106.08777.

Bačák, M. (2014). "Computing medians and means in Hadamard spaces". In: *SIAM Journal on Optimization* 24.3, pp. 1542–1566. DOI: 10.1137/140953393.

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

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. arXiv: 1908.02022.

RB, J. Persch, and G. Steidl (2016). "A parallel Douglas Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds". In: *SIAM Journal on Imaging Sciences* 9.4, pp. 901–937. DOI: 10.1137/15M1052858.

Chambolle, A. and T. Pock (2011). "A first-order primal-dual algorithm for convex problems with applications to imaging". In: *Journal of Mathematical Imaging and Vision* 40.1, pp. 120–145. DOI: 10.1007/s10851-010-0251-1.

Silva Louzeiro, M., RB, and R. Herzog (June 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.