

# **A Primal-Dual Algorithm for Convex Nonsmooth Optimization on Riemannian Manifolds**

Ronny Bergmann

joint work with

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

IFIP TC 7, 2021

Minisymposium: Non-Smooth First-order Methods, Convex, and Non-convex  
Quito, Ecuador & virtual

August 28 – September 3, 2021

# The Model

We consider a minimization problem

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

- ▶  $\mathcal{M}, \mathcal{N}$  are (high-dimensional) Riemannian Manifolds
  - ▶  $F: \mathcal{M} \rightarrow \overline{\mathbb{R}}$  nonsmooth, (locally, geodesically) convex
  - ▶  $G: \mathcal{N} \rightarrow \overline{\mathbb{R}}$  nonsmooth, (locally) convex
  - ▶  $\Lambda: \mathcal{M} \rightarrow \mathcal{N}$  nonlinear
  - ▶  $\mathcal{C} \subset \mathcal{M}$  strongly geodesically convex.
- ④ 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]

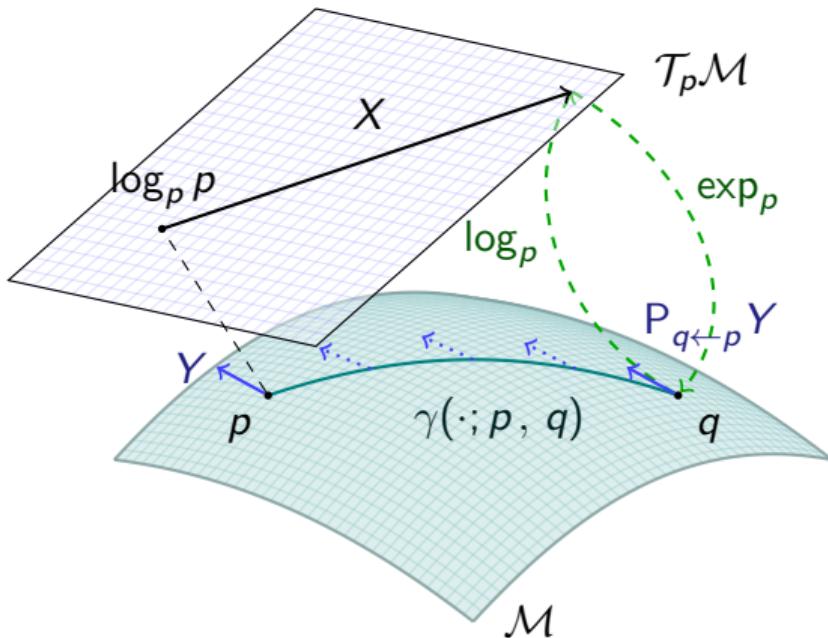
But on a Riemannian manifold  $\mathcal{M}$ :  no duality theory!

## Goals of this talk.

Formulate Duality on a Manifold

Derive a Riemannian Chambolle–Pock Algorithm (RCPA)

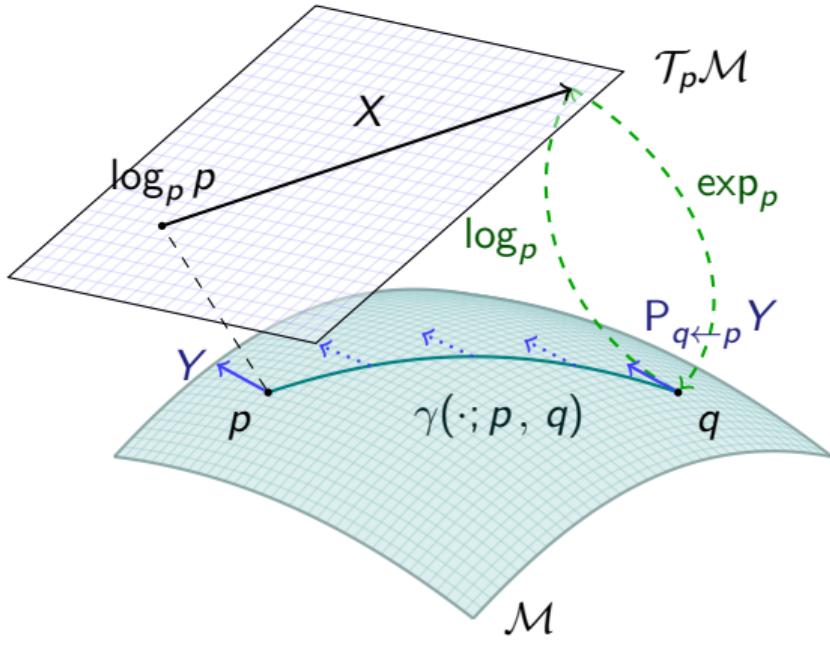
## A $d$ -dimensional Riemannian manifold $\mathcal{M}$



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 continuously 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**  $P_{q \leftarrow p} Y$

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

# Musical Isomorphisms

[Lee 2003]

The dual space  $\mathcal{T}_p^*\mathcal{M}$  of a tangent space  $\mathcal{T}_p\mathcal{M}$  is called **cotangent space**. We denote by  $\langle \cdot, \cdot \rangle$  the duality pairing.

We define the **musical isomorphisms**

- ▶  $\flat: \mathcal{T}_p\mathcal{M} \ni X \mapsto X^\flat \in \mathcal{T}_p^*\mathcal{M}$  via  $\langle X^\flat, Y \rangle = (X, Y)_p$  for all  $Y \in \mathcal{T}_p\mathcal{M}$
  - ▶  $\sharp: \mathcal{T}_p^*\mathcal{M} \ni \xi \mapsto \xi^\sharp \in \mathcal{T}_p\mathcal{M}$  via  $(\xi^\sharp, Y)_p = \langle \xi, Y \rangle$  for all  $Y \in \mathcal{T}_p\mathcal{M}$ .
- $\Rightarrow$  inner product and parallel transport on/between  $\mathcal{T}_p^*\mathcal{M}$

# Convexity

[Sakai 1996; Udrişte 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  $F: \mathcal{C} \rightarrow \overline{\mathbb{R}}$  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 Euclidean Fenchel Conjugate

Let  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be proper and convex.

We define the **Fenchel conjugate**  $f^*: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  of  $f$  by

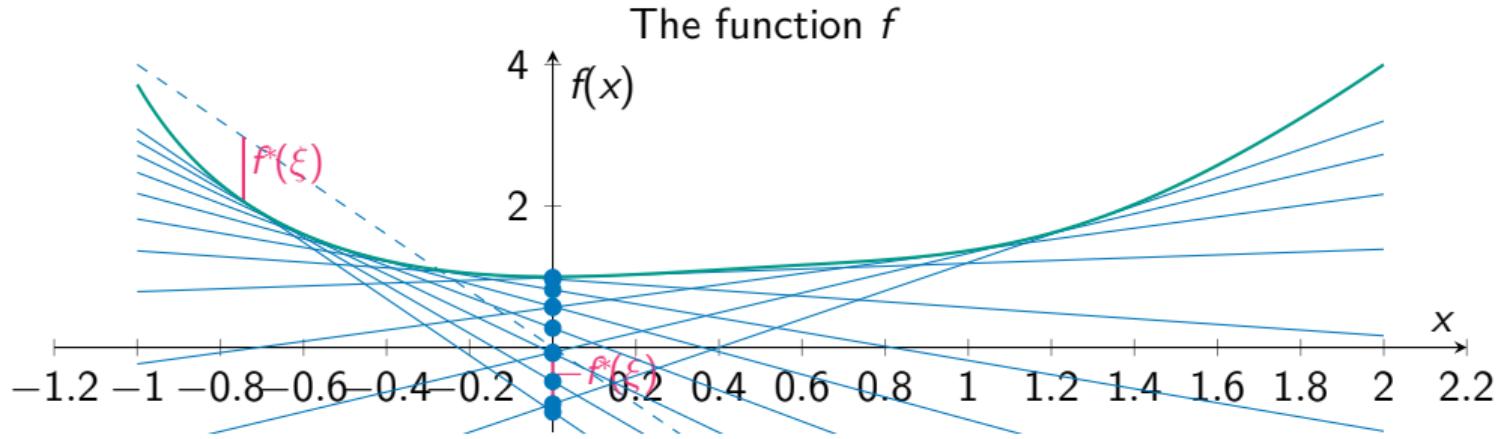
$$f^*(\xi) := \sup_{x \in \mathbb{R}^n} \langle \xi, x \rangle - f(x) = \sup_{x \in \mathbb{R}^n} \begin{pmatrix} \xi \\ -1 \end{pmatrix}^\top \begin{pmatrix} x \\ f(x) \end{pmatrix}$$

- ▶ interpretation: maximize the distance of  $\xi^\top x$  to  $f$
- ⇒ extremum seeking problem on the epigraph

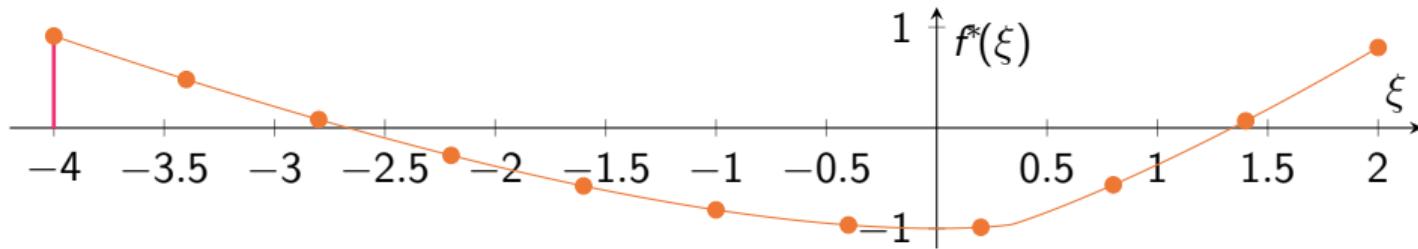
The Fenchel biconjugate reads

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

# Illustration of the Fenchel Conjugate



The Fenchel conjugate  $f^*$



# 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  $F: \mathcal{C} \rightarrow \overline{\mathbb{R}}$ .

The  $m$ -Fenchel conjugate  $F_m^*: \mathcal{T}_m^*\mathcal{M} \rightarrow \overline{\mathbb{R}}$  is defined by

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

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

Let  $m' \in \mathcal{C}$ .

The  $mm'$ -Fenchel-biconjugate  $F_{mm'}^{**}: \mathcal{C} \rightarrow \overline{\mathbb{R}}$  is given by

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

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

# Saddle Point Formulation

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

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

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

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

For Optimality Conditions and the Dual Problem: What's  $\Lambda^*$ ?

**Approach.** Linearization:  $\Lambda(p) \approx \exp_{\Lambda(m)} D\Lambda(m)[\log_m p]$

[Valkonen 2014]

# The exact Riemannian Chambolle–Pock Algorithm (eRCPA)

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

- 1:  $k \leftarrow 0$
- 2:  $\bar{p}^{(0)} \leftarrow p^{(0)}$
- 3: **while** not converged **do**
- 4:    $\xi_n^{(k+1)} \leftarrow \text{prox}_{\tau G_n^*} \xi_n^{(k)} + \tau (\log_n \Lambda(\bar{p}^{(k)}))^\flat$
- 5:    $p^{(k+1)} \leftarrow \text{prox}_{\sigma F} \exp_{p^{(k)}} \left( P_{m \leftarrow} p^{(k)} - \sigma D\Lambda(m)^* [\xi_n^{(k+1)}]^\sharp \right)$
- 6:    $\bar{p}^{(k+1)} \leftarrow \exp_{p^{(k+1)}} (-\theta \log_{p^{(k+1)}} p^{(k)})$
- 7:    $k \leftarrow k + 1$
- 8: **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  $\gamma$
- ▶ 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  $\Lambda$ , i.e., adopt the Euclidean case from  
 $\log_n \Lambda(\bar{p}^{(k)}) \rightarrow P_{n \leftarrow \Lambda(m)} D\Lambda(m) [\log_m \bar{p}^{(k)}]$  [Valkonen 2014]
- ▶ choose  $n \neq \Lambda(m)$  introduces a parallel transport

$$D\Lambda(m)^*[\xi_n^{(k+1)}] \rightarrow D\Lambda(m)^*[P_{\Lambda(m) \leftarrow n} \xi_n^{(k+1)}]$$

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

# The Linearized RCPA with Dual Relaxation

We introduce for ease of notation

$$\tilde{p}^{(k)} = \exp_{p^{(k)}} \left( P_{p^{(k)} \leftarrow m} - (\sigma(D\Lambda(m))^* [\bar{\xi}_n^{(k)}])^\sharp \right)$$

for the linearized Riemannian Chambolle Pock  
with dual relaxed

$$\bar{\xi}_n^{(k)} \leftarrow \xi_n^{(k)} + \theta(\xi_n^{(k)} - \xi_n^{(k-1)}).$$

Especially for  $\theta = 1$  we obtain

$$\bar{\xi}_n^{(k)} = 2\xi_n^{(k)} - \xi_n^{(k-1)}.$$

# A Conjecture

We define

$$C(k) := \frac{1}{\sigma} d^2(p^{(k)}, \tilde{p}^{(k)}) + \langle \bar{\xi}_n^{(k)}, D\Lambda(m)[\zeta_k] \rangle,$$

where

$$\zeta_k = P_{m \leftarrow p^{(k)}} (\log_{p^{(k)}} p^{(k+1)} - P_{p^{(k)} \leftarrow \tilde{p}^{(k)}} \log_{\tilde{p}^{(k)}} \hat{p}) - \log_m p^{(k+1)} + \log_m \hat{p},$$

and  $\hat{p}$  is a minimizer of the primal problem.

## Remark.

For  $\mathcal{M} = \mathbb{R}^d$ :  $\zeta_k = \tilde{p}^{(k)} - p^{(k)} = -\sigma(D\Lambda(m))^*[\bar{\xi}_n^{(k)}] \Rightarrow C(k) = 0$ .

## Conjecture.

Assume  $\sigma\tau < \|D\Lambda(m)\|^2$ . Then  $C(k) \geq 0$  for all  $k > K$ ,  $K \in \mathbb{N}$ .

# Convergence of the IRCPA

## Theorem.

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

Let  $\mathcal{M}, \mathcal{N}$  be Hadamard. Assume that the linearized problem

$$\min_{p \in \mathcal{M}} \max_{\xi_n \in \mathcal{T}_n^* \mathcal{N}} \langle (D\Lambda(m))^*[\xi_n], \log_m p \rangle + F(p) - G_n^*(\xi_n).$$

has a saddle point  $(\hat{p}, \hat{\xi}_n)$ .

Choose  $\sigma, \tau$  such that

$$\sigma\tau < \|D\Lambda(m)\|^2$$

and assume that  $C(k) \geq 0$  for all  $k > K$ . Then it holds

1. the sequence  $(p^{(k)}, \xi_n^{(k)})$  remains bounded,
2. there exists a saddle-point  $(p', \xi'_n)$  such that  $p^{(k)} \rightarrow p'$  and  $\xi_n^{(k)} \rightarrow \xi'_n$ .

# The $\ell^2$ -TV Model

[Rudin, Osher, and Fatemi 1992; Lellmann, Strelakowski, 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

$$\arg \min_{p \in \mathcal{M}} \frac{1}{\alpha} F(p) + G(\Lambda(p)), \quad \alpha > 0,$$

with

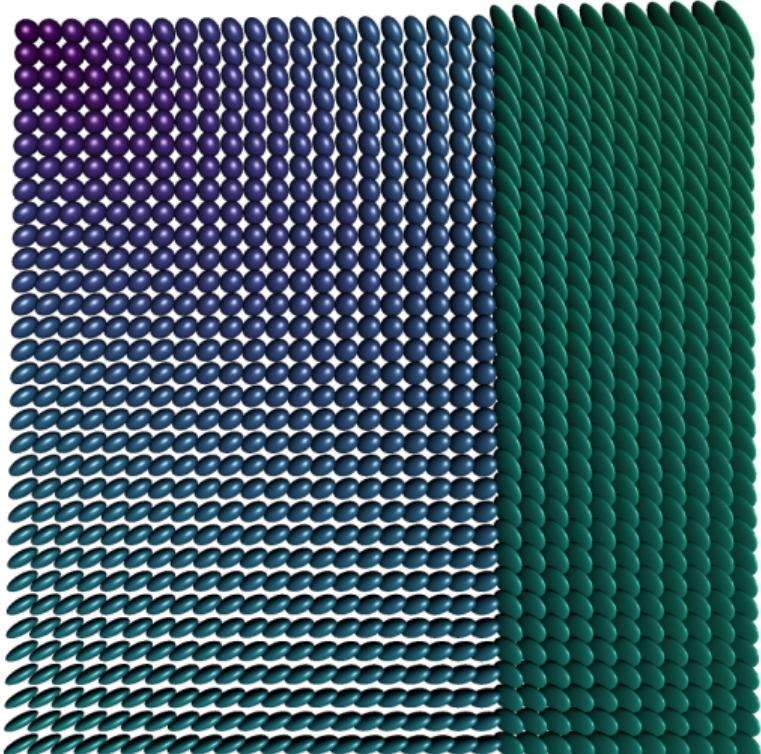
- ▶ data term  $F(p) = \frac{1}{2} d_{\mathcal{M}}^2(p, f)$
- ▶ “forward differences”  $\Lambda: \mathcal{M} \rightarrow (\mathcal{T}\mathcal{M})^{d_1-1, d_2-1, 2}$ ,

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

- ▶ prior  $G(X) = \|X\|_{g,q,1}$  similar to a collaborative TV

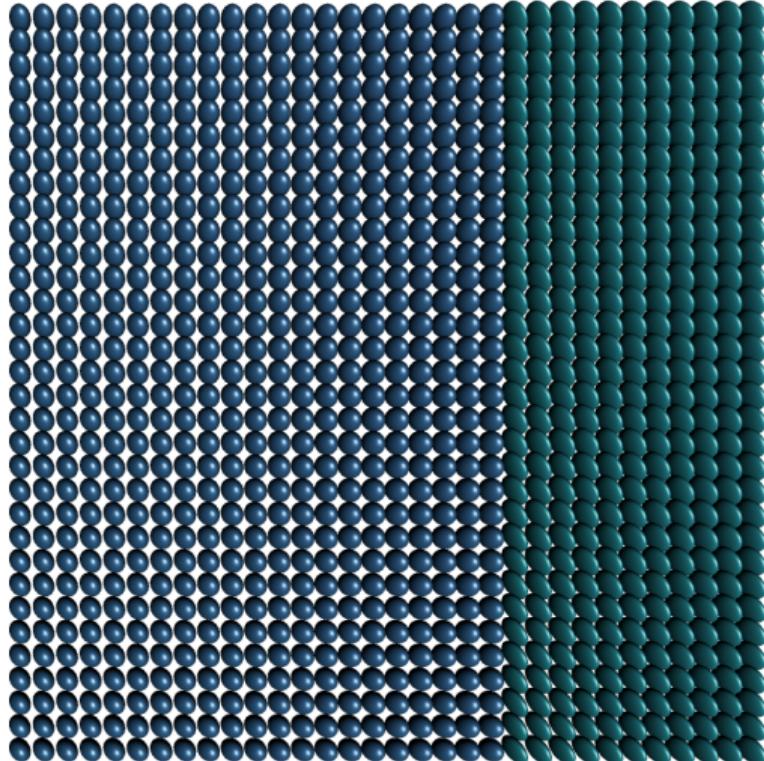
[Duran, Moeller, Sbert, and Cremers 2016]

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



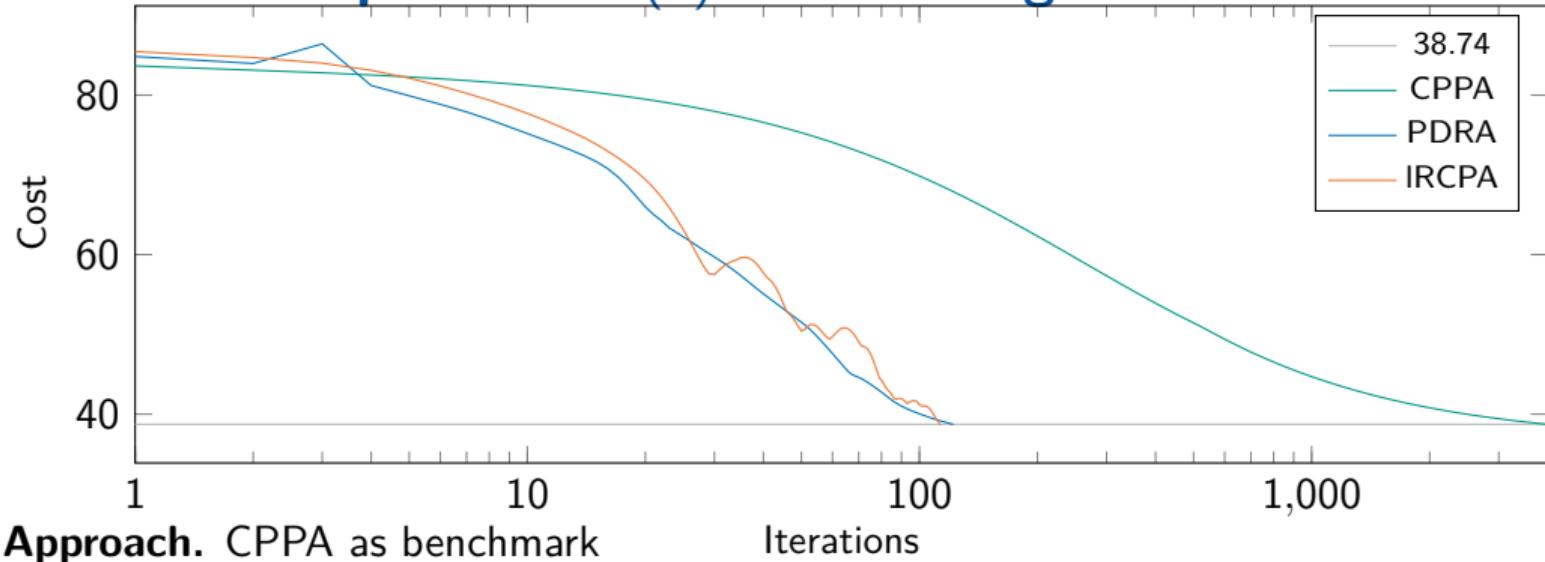
$\mathcal{P}(3)$ -valued data.

- ▶ in each pixel we have a symmetric positive definite matrix



anisotropic TV,  $\alpha = 6$ .

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



Approach. CPPA as benchmark

Iterations

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

## Base point Effect on $\mathbb{S}^2$ -valued data

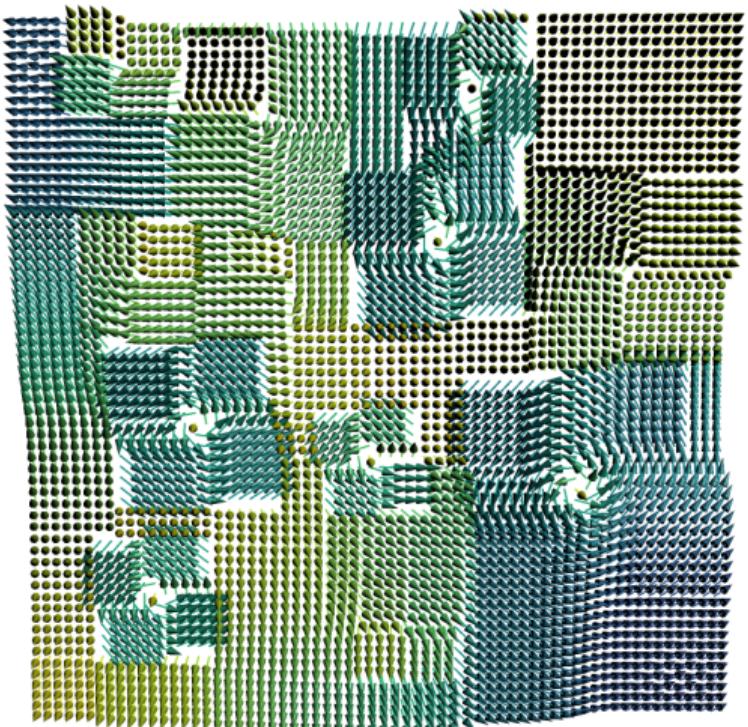


Original data

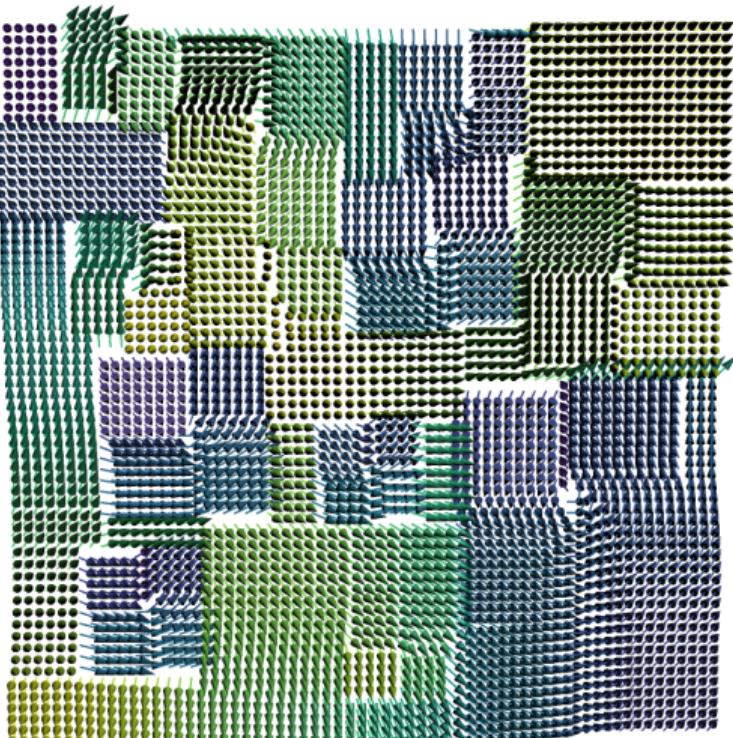


Original data

## Base point Effect on $\mathbb{S}^2$ -valued data

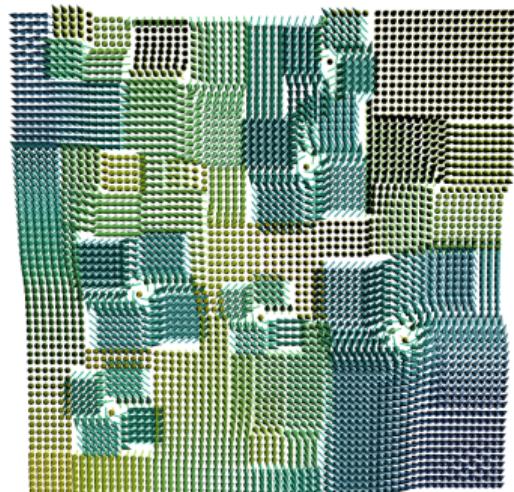


Result,  $m$  the mean (p. Px.)

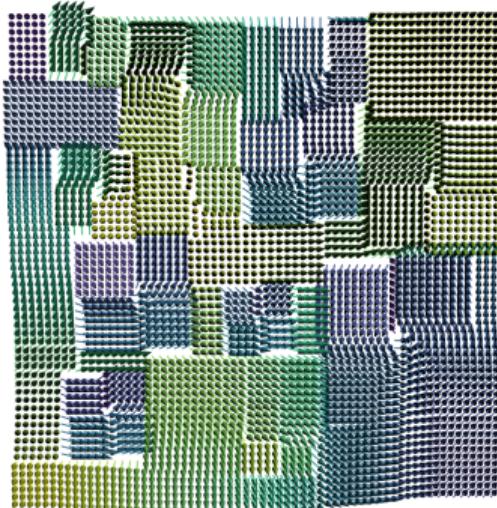


Result,  $m$  west (p. Px.)

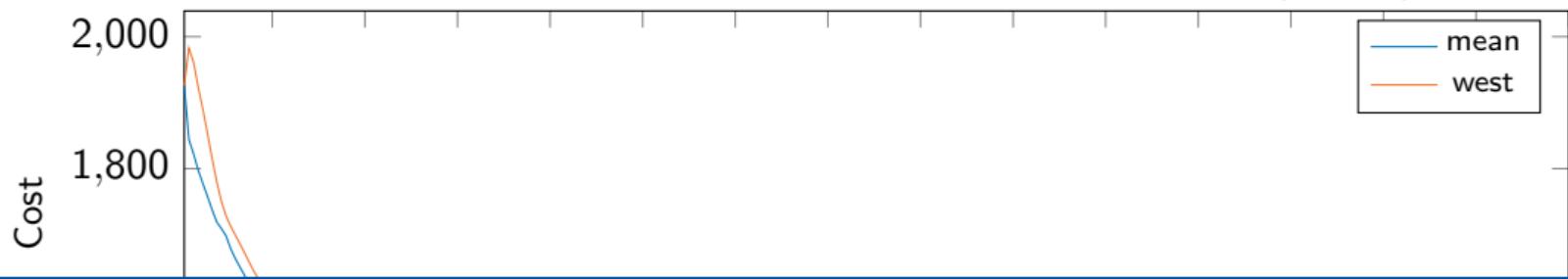
## Base point Effect on $\mathbb{S}^2$ -valued data



## Result, $m$ the mean (p. Px.)



### Result, $m$ west (p. Px.)



# Summary & Outlook

## Summary.

- ▶ We introduced a duality framework on Riemannian manifolds
- ▶ We derived a Riemannian Chambolle Pock Algorithm
- ▶ Numerical examples illustrate performance

## Outlook.

- ▶ investigate  $C(k)$  and the error of linearization
- ▶ strategies for choosing  $m, n$  (adaptively)
- ▶ alternative models of Fenchel duality (e.g. without  $m$ ) [RB, Herzog, and Silva Louzeiro 2021]
  - 👤 Thu @ 11:00 BST (16:00 CEST) in MS Optimization and Manifolds
- ▶ higher order methods non-smooth methods
  - 👤 W. Diepeveen, Thu @ 11:30 BST (16:30 CEST) in MS Optimization and Manifolds

# Reproducible Research

The algorithm is published in `Manopt.jl`, a **Julia** Package available at  
<http://manoptjl.org>.

It uses the interface from `ManifoldsBase.jl` and hence any manifold  
from `Manifolds.jl` can be used in the algorithms.

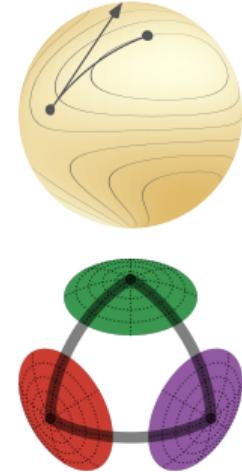
<https://juliamanifolds.github.io/Manifolds.jl/>  
[Axen, Baran, RB, and Rzecki 2021]

## Goal.

Being able to use an(y) algorithm for a(ny) model directly on a(ny)  
manifold easily and efficiently.

## Alternatives.

- ▶ `Manopt`, [manopt.org](http://manopt.org) (Matlab, by N. Boumal)
- ▶ `pymanopt`, [pymanopt.github.io](https://pymanopt.github.io) (Python, by S. Weichwald, J. Townsend, N. Koep)



# Selected References

-  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](https://doi.org/10.1016/j.na.2010.07.033).
-  Axen, S. D., M. Baran, RB, and K. Rzecki (2021). *Manifolds.jl: An Extensible Julia Framework for Data Analysis on Manifolds*. arXiv: [2106.08777](https://arxiv.org/abs/2106.08777).
-  RB, R. Herzog, and M. Silva Louzeiro (2021). *Fenchel duality and a separation theorem on Hadamard manifolds*. arXiv: [2102.11155](https://arxiv.org/abs/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](https://doi.org/10.1007/s10208-020-09486-5).
-  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](https://doi.org/10.1007/s10851-010-0251-1).
-  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](https://doi.org/10.1088/0266-5611/30/5/055012).



[ronnybergmann.net/talks/2021-IFIP-RCPA.pdf](https://ronnybergmann.net/talks/2021-IFIP-RCPA.pdf)