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

Ronny Bergmann[a], Roland Herzog[a],
José Vidal-Núñez[a], Daniel Tenbrinck[b]

[a]Technische Universität Chemnitz, Chemnitz, Germany
[b]Friedrich-Alexander-Universität, Erlangen, Germany.

Optimization and Numerical Analysis Seminar
School of Mathematics, University of Birmingham.

Birmingham, February 26, 2020.

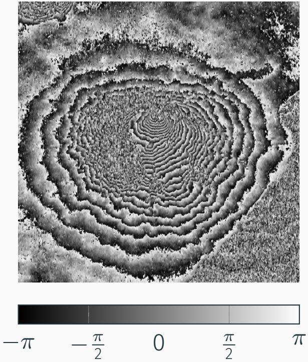
Contents

1. Introduction

Manifold-Valued Signals and Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



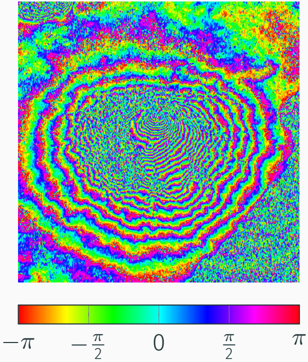
InSAR-Data of Mt. Vesuvius
[Rocca, Prati, Guarnieri, 1997]

phase-valued data, $\mathcal{M} = \mathbb{S}^1$

Manifold-Valued Signals and Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



InSAR-Data of Mt. Vesuvius

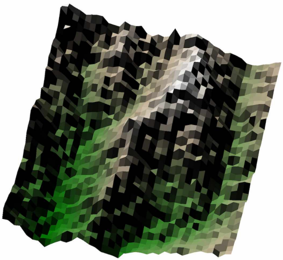
[Rocca, Prati, Guarnieri, 1997]

phase-valued data, $\mathcal{M} = \mathbb{S}^1$

Manifold-Valued Signals and Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



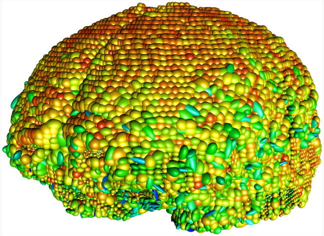
National elevation dataset
[Gesch et al., 2009]

directional data, $\mathcal{M} = \mathbb{S}^2$

Manifold-Valued Signals and Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,…
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



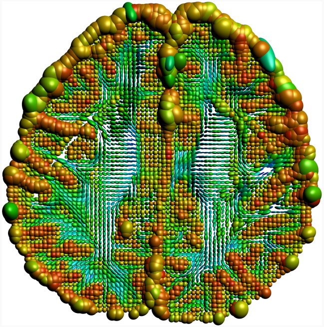
diffusion tensors in human brain from the Camino dataset
http://cmic.cs.ucl.ac.uk/camino

sym. pos. def. matrices, $\mathcal{M} = \mathrm{SPD}(3)$

Manifold-Valued Signals and Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



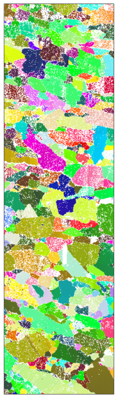
horizontal slice # 28
from the Camino dataset
http://cmic.cs.ucl.ac.uk/camino

sym. pos. def. matrices, $\mathcal{M} = \mathrm{SPD}(3)$

Manifold-Valued Signals and Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



EBSD example from the MTEX toolbox
[Bachmann, Hielscher, since 2005]

Rotations (mod. symmetry),
$\mathcal{M} = \mathrm{SO}(3)(/\mathcal{S})$.

Manifold-Valued Signals and Images

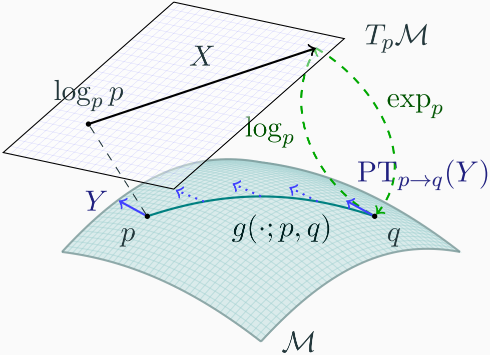
New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)

Common properties

- Range of values is a Riemannian manifold
- Tasks from “classical” image processing, e.g.
  - denoising
  - inpainting
  - interpolation
  - labeling
  - deblurring

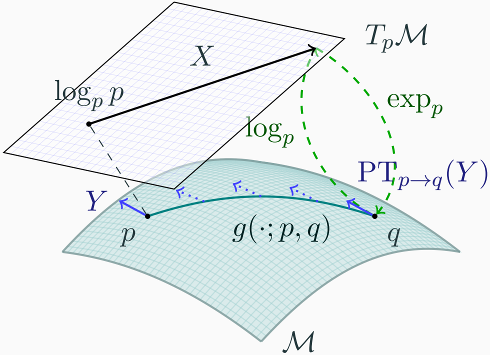
A d-dimensional Riemannian Manifold $\mathcal{M}$



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 tangential spaces.

[Absil, Mahony, Sepulchre, 2008]

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



Geodesic $g(\cdot; p, q)$ shortest path (on $\mathcal{M}$) between $p, q \in \mathcal{M}$
Tangent space $\mathrm{T}_p\mathcal{M}$ at p, with inner product $(\cdot, \cdot)_p$
Logarithmic map $\log_p q = \dot{g}(0; p, q)$ "speed towards q"
Exponential map $\exp_p X = g(1)$, where $g(0) = p$, $\dot{g}(0) = X$
Parallel transport $\mathrm{PT}_{p\to q}(Y)$ of $Y \in \mathrm{T}_p\mathcal{M}$ along $g(\cdot; p, q)$

The Model

We consider the minimization problem

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

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

Splitting Methods & Algorithms

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

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

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

- Primal-Dual Hybrid Gradient Algorithm (PDHGA) [Esser, Zhang, Chan, 2010]
- Chambolle-Pock Algorithm (CPA) [Chambolle, Pock, 2011; Pock et al., 2009]

Goals of this talk.

Formulate Duality on a Manifold

Derive a Riemannian Chambolle–Pock Algorithm (RCPA)

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

- $\flat\colon \mathcal{T}_p\mathcal{M} \ni X \mapsto X^\flat \in \mathcal{T}_p^*\mathcal{M}$ via $\langle X^\flat, Y\rangle = (X,Y)_p$ for all $Y \in \mathcal{T}_p\mathcal{M}$
- $\sharp\colon \mathcal{T}_p^*\mathcal{M} \ni \xi \mapsto \xi^\sharp \in \mathcal{T}_p\mathcal{M}$ 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 $g(\cdot; p, q)$ is unique and lies in $\mathcal{C}$.

A function $F : \mathcal{C} \to \overline{\mathbb{R}}$ is called convex if for all $p, q \in \mathcal{C}$ the composition $F(g(t; p, q)), t \in [0, 1]$, is convex.

The Subdifferential

[Lee, 2003; Udrişte, 1994]

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

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

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 $\mathcal{T}^*_p\mathcal{M} \times \mathcal{T}_p\mathcal{M}$

2. Fenchel Duality

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

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

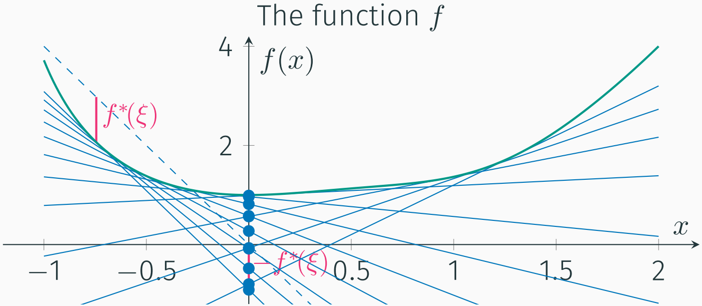
- interpretation: maximize the distance of $\xi^{\mathrm{T}}x$ to f

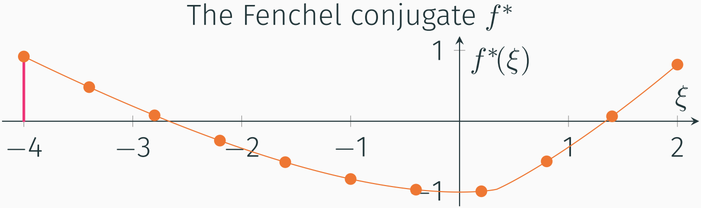
$\Rightarrow$ extremum seeking problem on the epigraph

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 $g^*(\xi) = f^*(\xi) - \xi^{\mathrm{T}} b$ 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.,

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

We can characterize subdifferentials

- For a proper, convex function f

$$\xi \in \partial f(x) \Leftrightarrow f(x) + f^*(\xi) = \xi^\mathrm{T} x$$

- For a proper, convex, lsc function f, then

$$\xi \in \partial f(x) \Leftrightarrow x \in \partial f^*(\xi)$$

The Riemannian $m-$Fenchel Conjugate

[Bergmann, Herzog, et al., 2019]
alternative approach: [Ahmadi Kakavandi, 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

$\mathcal{L}_{\mathcal{C},m} \coloneqq \{X \in \mathcal{T}_m\mathcal{M} \mid q = \exp_m X \in \mathcal{C}$ 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

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

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 $F_m^*(\xi) \geq G_m^*(\xi_m)$ holds for all $\xi_m \in \mathcal{T}_m^*\mathcal{M}$
- If $G(x) = F(x) + a$ 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 convex, 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

$$\xi_p \in \partial_{\mathcal{M}} F(p) \Leftrightarrow F(p) + F_m^*(\mathcal{P}_{p \to m}\, \xi_p) = \langle \mathcal{P}_{p \to m}\, \xi_p, \log_m p \rangle.$$

- For a proper, convex, lsc function F

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

3. The Chambolle–Pock Algorithm

The Euclidean Chambolle–Pock Algorithm

[Chambolle, 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

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

The Euclidean Chambolle–Pock Algorithm

[Chambolle, 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

Chambolle–Pock Algorithm. with $\sigma > 0$, $\tau > 0$, $\theta \in \mathbb{R}$

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

Proximal Map

For $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ and $\lambda > 0$ we define the Proximal Map as

[Moreau, 1965; Rockafellar, 1976; Ferreira, Oliveira, 2002]

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

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

- For F proper, convex, lsc:
  - the proximal map is unique.
  - PPA $x_k = \operatorname{prox}_{\lambda F}(x_{k-1})$ converges to $\operatorname{arg\,min} F$
- $q = \operatorname{prox}_{\lambda F}(p)$ is equivalent to

$$\frac{1}{\lambda}\bigl(\log_q p\bigr)^\flat \in \partial_{\mathcal{M}} F(q)$$

Saddle Point Formulation

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

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

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

Approach. Linearization: on $\mathbb{R}^n$: [Valkonen, 2014]

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

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

$$\begin{aligned} \mathcal{P}_{m \to \widehat{p}} - (D\Lambda)^*[\widehat{\xi}_n] &\in \partial_{\mathcal{M}} F(\widehat{p}) \\ \log_n \Lambda(\hat{p}) &\in \partial G_n^*(\widehat{\xi}_n) \end{aligned}$$

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

Linearization & the Dual Problem

Linearizing the primal problem obtain e.g. for $n = \Lambda(m)$

Primal Problem.

$$\min_{p\in\mathcal{C}} F(p) + G(\exp_{\Lambda(m)} D\Lambda(m)[\log_m p])$$

Saddle Point Problem.

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

Dual Problem.

$$\max_{\xi_n\in\mathcal{T}_n^*\mathcal{N}} -F_m^*(-D\Lambda(m)^*[\xi_n]) - G_n^*(\xi_n).$$

and a classical duality theory including weak duality.

Optimality Conditions for the Saddle Point Problem

The first order optimality conditions for a saddle point of the linearized problem

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

can be formally derived as

$$\mathcal{P}_{m \to \widehat{p}} - (D\Lambda)^*[\widehat{\xi_n}] \in \partial_{\mathcal{M}} F(\widehat{p})$$
$$D\Lambda(m)[\log_m \hat{p}] \in \partial G_n^*(\widehat{\xi_n})$$

Advantage. A complete duality theory and a certain symmetry in the optimality conditions.

For $\mathcal{M} = \mathbb{R}^d$ and $K = \Lambda$ linear both approaches yield the classical conditions

$$-K^* \widehat{\xi} \in \partial F(\widehat{p})$$
$$K\widehat{p} \in \partial G^*(\widehat{\xi})$$

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 σ, τ, $\theta > 0$

$k \leftarrow 0$
$\bar{p}^{(0)} \leftarrow p^{(0)}$
while not converged **do**
  $\xi_n^{(k+1)} \leftarrow \operatorname{prox}_{\tau G_n^*}\bigl(\xi_n^{(k)} + \tau\bigl(\log_n \Lambda(\bar{p}^{(k)})\bigr)^\flat\bigr)$
  $p^{(k+1)} \leftarrow \operatorname{prox}_{\sigma F}\Bigl(\exp_{p^{(k)}}\Bigl(\mathcal{P}_{m\to p^{(k)}}\bigl(-\sigma D\Lambda(m)^*[\xi_n^{(k+1)}]\bigr)^\sharp\Bigr)\Bigr)$
  $\bar{p}^{(k+1)} \leftarrow \exp_{p^{(k+1)}}\bigl(-\theta \log_{p^{(k+1)}} p^{(k)}\bigr)$
  $k \leftarrow k + 1$
end while

Output: $p^{(k)}$

Generalizations & Variants of the RCPA

Classically [Chambolle, 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 [Bergmann, Herzog, et al., 2019]

- introduce the lRCPA: linearize Λ, too, i.e.

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

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

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

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

The Linearized Riemannian Chambolle–Pock Algorithm (lRCPA)

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, $\theta > 0$

$k \leftarrow 0$

$\bar{\xi}_n^{(0)} \leftarrow \xi^{(0)}$

while not converged **do**

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

$\quad \xi_n^{(k+1)} \leftarrow \operatorname{prox}_{\tau G_n^*}\Bigl(\xi_n^{(k)} + \tau\bigl(D\Lambda(m)[\log_m p^{(k+1)}])\bigr)^\flat\Bigr)$

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

$\quad k \leftarrow k+1$

end while

Output: $p^{(k)}$

The Linearized RCPA with Dual Relaxation

We introduce for ease of notation

$$\widetilde{p}^{(k)} = \exp_{p^{(k)}}\Big(\mathcal{P}_{m\to p^{(k)}} - \big(\sigma(D\Lambda(m))^*[\bar{\xi}_n^{(k)}]\big)^\sharp\Big)$$

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

$$\zeta_k = \mathcal{P}_{p^{(k)} \to m} \left(\log_{p^{(k)}} p^{(k+1)} - \mathcal{P}_{\tilde{p}^{(k)} \to p^{(k)}} \log_{\tilde{p}^{(k)}} \widehat{p} \right) - \log_m p^{(k+1)} + \log_m \widehat{p},$$

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 $k > K$, $K \in \mathbb{N}$.

Theorem. [Bergmann, Herzog, et al., 2019]

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 $(\widehat{p}, \widehat{\xi_n})$. Choose σ, τ 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', ξ_n') such that $p^{(k)} \to p'$ and $\xi_n^{(k)} \to \xi_n'$.

4. Numerical Examples

The ℓ^2-TV Model

[Rudin, Osher, Fatemi, 1992; Lellmann et al., 2013; Weinmann, Demaret, 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" $\Lambda\colon \mathcal{M} \to (T\mathcal{N})^{d_1 - 1,\, d_2 - 1,\, 2}$,

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

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

[Duran et al., 2016]

The $d \times d$ Symmetric Positive Definite Matrices $\mathcal{P}(d)$.

$$\mathcal{P}(d) = \bigl\{p \in \mathbb{R}^{d\times d} \bigm| x^{\mathrm{T}}px > 0 \quad \text{for all } 0 \neq x \in \mathbb{R}^d\bigr\}$$

Tangent Space. $\mathcal{T}_p\mathcal{P}(d) = \bigl\{p^{\frac{1}{2}}Xp^{\frac{1}{2}}\bigm|X \in \mathbb{R}^{d\times d} \text{ with } X = X^{\mathrm{T}}\}\bigr\}$

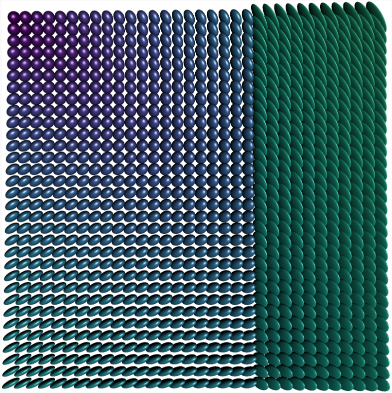
Riemannian Metric. $(X, Y)_p = \operatorname{tr}(p^{-1}Xp^{-1}Y)$,

Exponential Map. $\exp_p X = p^{\frac{1}{2}} \operatorname{Exp}(p^{-\frac{1}{2}}Xp^{-\frac{1}{2}})p^{\frac{1}{2}}$,
where Exp is the matrix exponential.

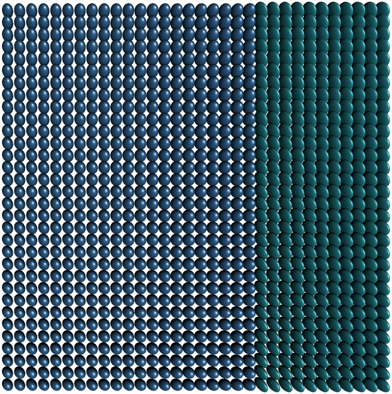
Parallel Transport. $P_{p\to q}(X) = p^{\frac{1}{2}}X'p^{-\frac{1}{2}}Xp^{-\frac{1}{2}}X'p^{\frac{1}{2}}$,
$X' = \operatorname{Exp}\bigl(\frac{1}{2}p^{-\frac{1}{2}}\log_p(q)p^{-\frac{1}{2}}\bigr)$,
where $\log$ is the logarithmic map.

The main tool to compute the matrix square root is the SVD.

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



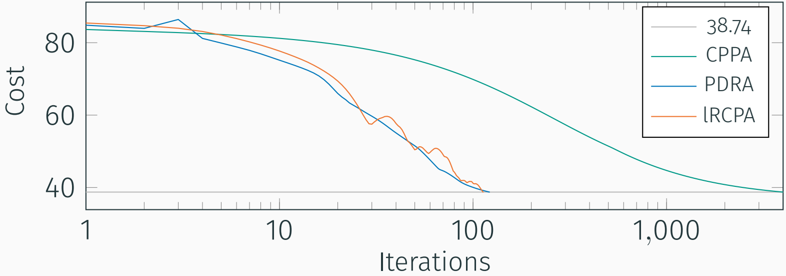
$\mathcal{P}(3)$-valued data.



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

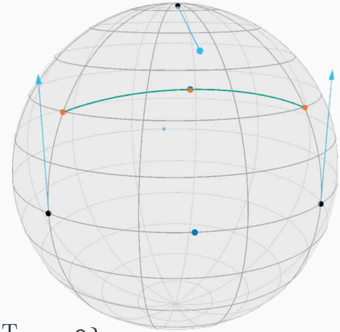


Approach. CPPA as benchmark

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

The Sphere $\mathbb{S}^d$ as a Manifold

$$\mathbb{S}^d = \left\{p \in \mathbb{R}^{d+1} \;\middle|\; \|p\| = 1\right\}$$



Tangent Space. $\mathcal{T}_p\mathbb{S}^2 = \left\{X \in \mathbb{R}^{d+1} \middle| X^{\mathrm{T}}p = 0\right\}$

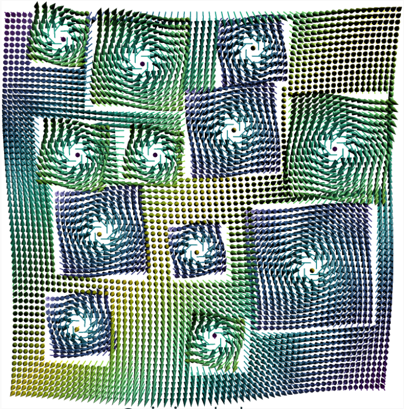
Riemannian Metric. $(X, Y)_p = \langle X, Y\rangle$ from the embedding

Distance. $d_{\mathbb{S}^d}(p, q) = \operatorname{acos}\bigl(\langle p, q\rangle\bigr)$

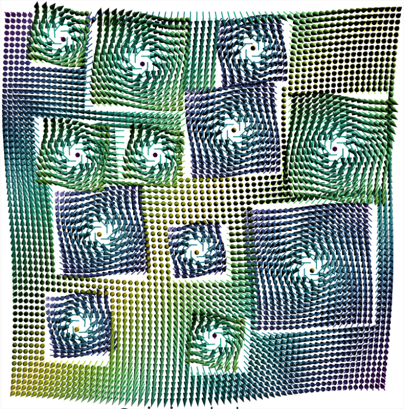
Exponential Map. $\exp_p X = \cos(\|X|_2)p + \sin(\|X\|_2)\frac{X}{\|X\|_2}$

Parallel Transport. $P_{p\to q}(X) = X - \frac{\langle \log_p q, X\rangle_x}{d^2_{\mathbb{S}^d}(p,q)}\bigl(\log_p q + \log_q p\bigr).$

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

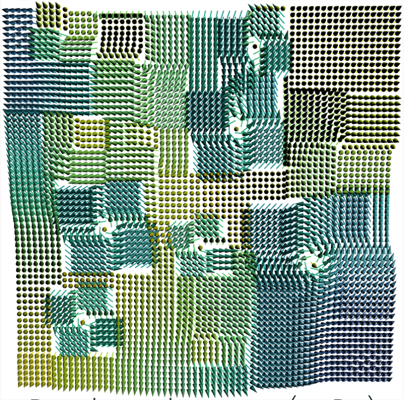


Original data

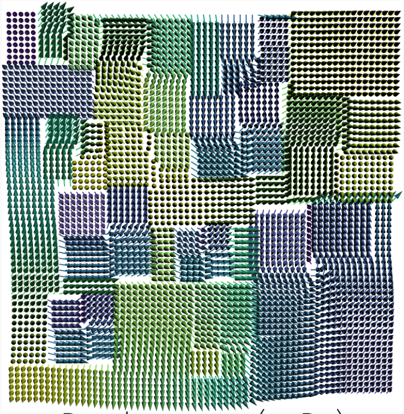


Original data

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



Result, m the mean (p. Px.)

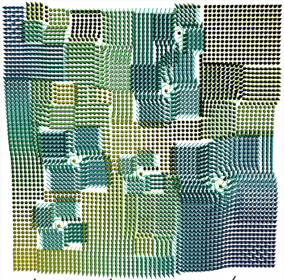


Result, m west (p. Px.)

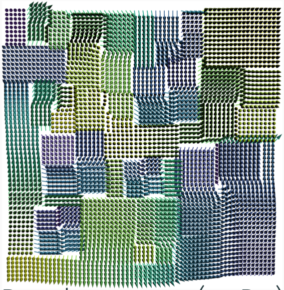
- pieceweise constant results for both

! different linearizations lead to different models

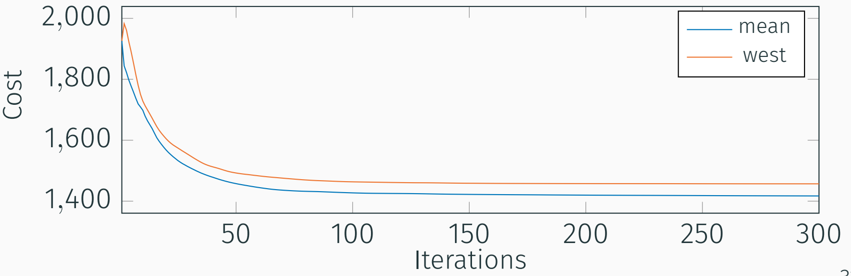
Base point Effect on $\mathbb{S}^2$-valued data



Result, m the mean (p. Px.)



Result, m west (p. Px.)



5. Summary & Conclusion

Summary & Outlook

Summary.

- We introduced a duality framework
  on Riemannian manifolds
- We derived a Riemannian Chambolle Pock Algorithm
- Numerical example illustrates performance

Outlook.

- investigate $C(k)$
- strategies for choosing m, n (adaptively)
- investigate linearization error
- extend algorithm to graph-structured data

Reproducible Research

The algorithm will be published in `Manopt.jl`, a Julia Package available at `http://manoptjl.org`.

Goal.

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

Alternatives.

Manopt – manopt.org
(Matlab, N. Boumal)

pymanopt – pymanopt.github.io
(Python, S. Weichwald et. al.)

Example.

```
pOpt = linearizedChambollePock(M, N, cost,
 p, ξ, m, n, DΛ, AdjDΛ, proxF, proxConjG, σ, τ)
```

The algorithm will be published in `Manopt.jl`, a Julia Package available at `http://manoptjl.org`.

Goal.

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

Alternatives.

Manopt – manopt.org (Matlab, N. Boumal)

pymanopt – pymanopt.github.io (Python, S. Weichwald et. al.)

Example.

```
pOpt = exactChambollePock(M, N, cost,
 p, ξ, m, n,  Λ, AdjDΛ, proxF, proxConjG, σ, τ)
```

Selected References

Absil, P.-A.; Mahony, R.; Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. DOI: 10.1515/9781400830244.

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

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

Bergmann, R.; Herzog, R.; Tenbrinck, D.; Vidal-Núñez, J. (2019). *Fenchel duality theory and a primal-dual algorithm on Riemannian manifolds*. arXiv: 1908.02022.

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

Rockafellar, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J.

ronnybergmann.net/talks/2020-Birmingham-RiemannianChambollePock.pdf