A Primal Dual Algorithm for Convex Nonsmooth Optimization 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.

Geometric Methods in Optimization of Variational Problems (Part I)
International Conference on Continuous Optimization 2019 (ICCOPT'19),

Berlin,
August 8, 2019.

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

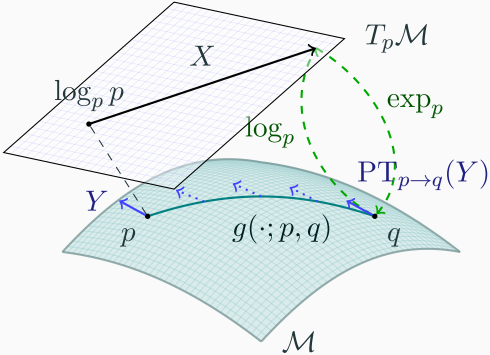
- Cyclic Proximal Point Algorithm (CPPA) [Bačák, 2014]
- (parallel) Douglas–Rachford Algorithm (PDRA) [RB, Persch, Steidl, 2016]

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]

Goal.
Formulate Duality on a Manifold
Derive a Riemannian Chambolle–Pock Algorithm (RCPA)

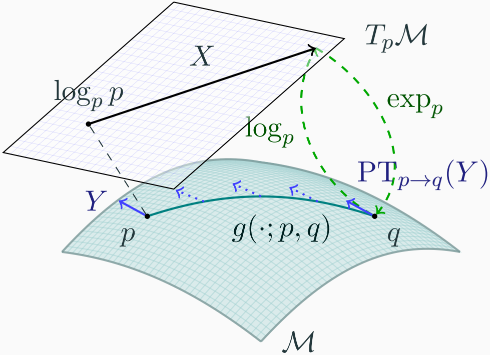
A d-dimensional Riemannian Manifold $\mathcal{M}$



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 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 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

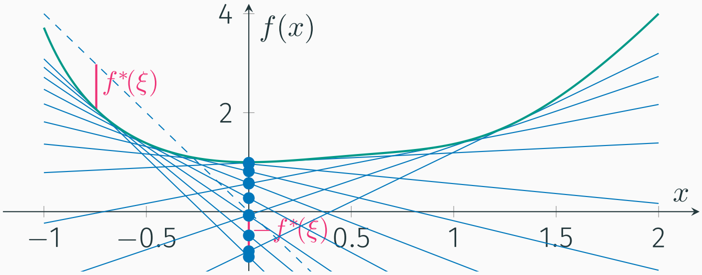
$$f^*(\xi) \coloneqq \sup_{x\in\mathbb{R}^n} \langle \xi, x\rangle - f(x) = \sup_{x\in\mathbb{R}^n} \begin{pmatrix} \xi \\ -1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} x \\ F(x) \end{pmatrix}$$

Prominent Properties. [Rockafellar, 1970]

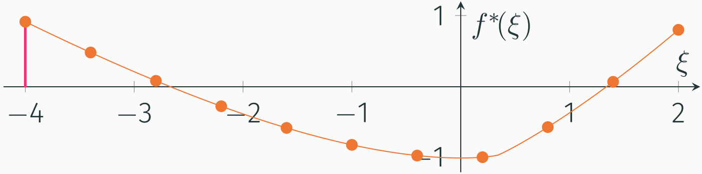
- f^* is a convex lsc function
- Fenchel–Young inequality: $f(x) + f^*(\xi) \geq \langle \xi, x\rangle$
- If f is proper, convex, lsc
  - $x \in \partial f^*(\xi) \Leftrightarrow \xi \in \partial f(x)$
  - $f^{**} = f$

Illustration of the Fenchel Conjugate

The function f



The Fenchel conjugate f^*



Musical Isomorphisms

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 $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ 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

$$\partial_{\mathcal{M}} F(p) := \{\xi \in \flat \exp_p^{-1}\mathcal{C} \,|\, F(q) \geq F(p) + \langle \xi, \log_p q \rangle \ \text{ for } q \in \mathcal{C}\},$$

where

- $\exp_p^{-1}\mathcal{C} \subset \mathcal{T}_p\mathcal{M}$ is the subset of tangent vectors such that $\exp_p X \in \mathcal{C}$.
- $\mathcal{T}_p^*\mathcal{M}$ 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 Riemannian $m-$Fenchel Conjugate

[RB, 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

$$F_m^*(\xi_m) := \sup_{X \in \exp_m^{-1}\mathcal{C}} \bigl\{\langle \xi_m, X\rangle - F(\exp_m X)\bigr\}.$$

Prominent Properties.

- F_m^* is convex lsc on $\mathcal{T}_m^*\mathcal{M}$
- $F(p) + F_m^*(\xi_m) \geq \langle \xi_m, \log_m p\rangle$
- If F is proper, convex, lsc
  - $\xi_p \in \partial_M F(p) \Leftrightarrow \log_m p \in \partial_M F_m^*(\mathcal{P}_{p\to m}\,\xi_p)$
  - $F = F_{mm}^{**}$ on $\mathcal{C}$

Saddle Point Formulation

From

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

we derive the saddle point formulation for $\mathcal{D} \subset \mathcal{N}$ around n as

$$\min_{p \in \mathcal{C}} \max_{\xi_n \in \flat \exp_n^{-1} \mathcal{D}} \langle \xi_n, \log_n \Lambda(p) \rangle + F(p) - G_n^*(\xi_n).$$

For the Dual Prolem: What's Λ^*?

Linearization & the Dual Problem

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

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

We obtain for e.g. for $n = \Lambda(m)$ that

$$\min_{p\in\mathcal{C}} \max_{\xi_n \in \flat \exp_n^{-1} \mathcal{D}} \langle D\Lambda(m)^*[\xi_n], \log_m p\rangle + F(p) - G_n^*(\xi_n).$$

and hence the Dual Problem

$$\max_{\xi_n \in \flat \exp_n^{-1} \mathcal{D}} -F_m^*(D\Lambda(m)^*[\xi_n]) - G_n^*(\xi_n).$$

For more details see

Fenchel duality for convex optimization

on Riemannian manifolds,

today @11.35 in H 1029 by José Vidal–Núñez

The exact Riemannian Chambolle–Pock Algorithm (eRCPA)

Input: $m, p^{(0)} \in \mathcal{C} \subset \mathcal{M}$, $n = \Lambda(m)$, $\xi_n^{(k+1)} \in \mathcal{T}_n^*\mathcal{N}$,
and parameters σ, τ, $\theta > 0$

$k \leftarrow 0$

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

while not converged **do**

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

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

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

$\quad k \leftarrow k + 1$

end while

Output: $p^{(k)}$

Generalizations & Variants of the RCPA

Classically [Chambolle, 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, 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

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

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

The Linearized RCPA with Dual Relaxation

We introduce for ease of notation

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

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

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

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

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

Convergence of the lRCPA

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

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

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

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 (p^*, ξ_n^*) such that $p^{(k)} \to p^*$ and $\xi_n^{(k)} \to \xi_n^*$.

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

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

with

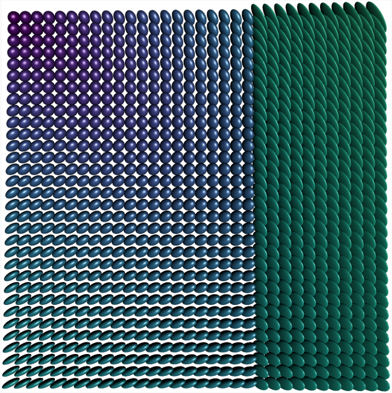
- data term $F(p) = \frac{1}{2} d^2_{\mathcal{M}}(p, f)$
- "forward differences" $\Lambda\colon \mathcal{M} \to (T\mathcal{N})^{d_1-1,\, d_2-1,\, 2}$,

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

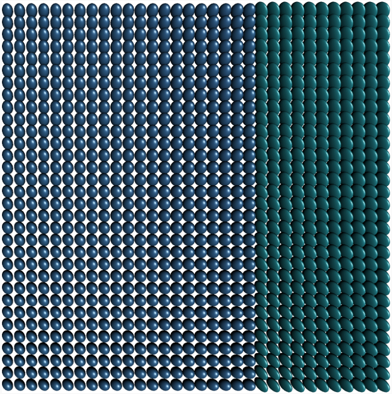
- prior $G(X) = \|X\|_{g,q,1}$ similar to a collaborative TV

[Duran et al., 2016]

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



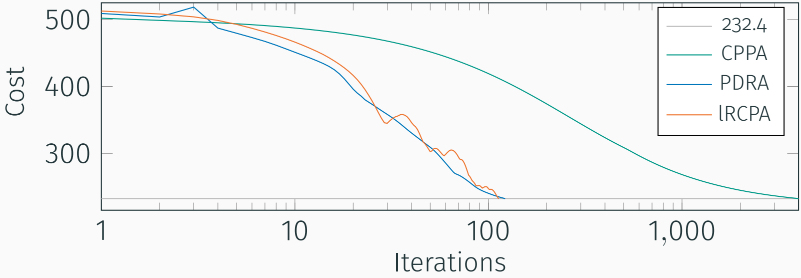
$\mathcal{P}(3)$-valued data.



anisotropic TV, $\alpha = 6$.

- in each pixel we have a symmetric positive definite matrix
- Applications: denoising/inpainting e.g. of DT-MRI data

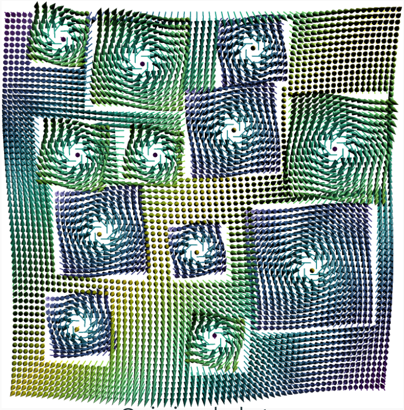
Numerical Example for a $\mathcal{P}(3)$-valued Image



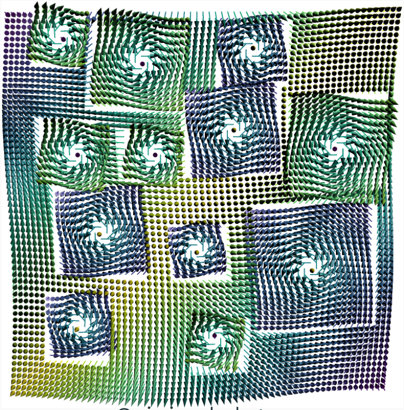
Approach. CPPA as benchmark

	CPPA	PDRA	lRCPA	CPPA c.	lRCPA c.
parameters	$\lambda_k = \frac{4}{k}$	$\eta = 0.58$ $\lambda = 0.93$	$\sigma = 0.37/\alpha$ $\tau = 0.37 * \alpha$ $\gamma = 0.2$, $m = I$		
iterations	4000	122	**113**	100 000	676
runtime	1484 s.	478 s.	**96.1 s.**	≈10.5 h	534 s.

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

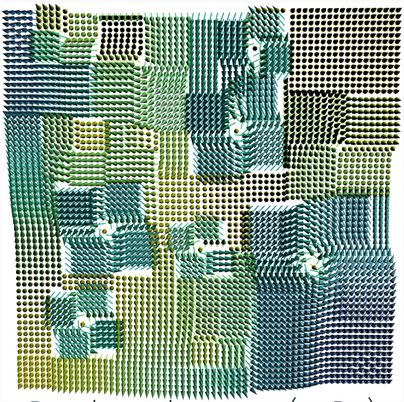


Original data

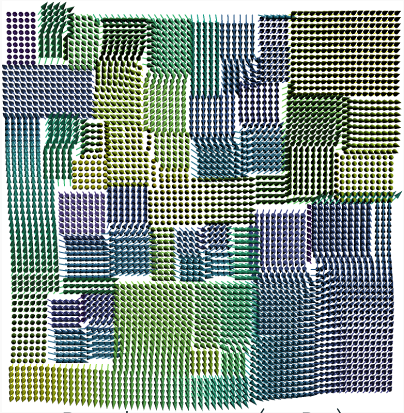


Original data

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



Result, m the mean (p. Px.)

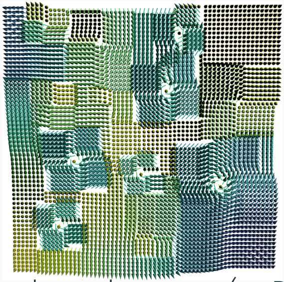


Result, m west (p. Px.)

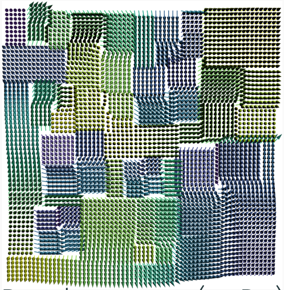
- pieceweise constant results for both

! different linearizations lead to different models

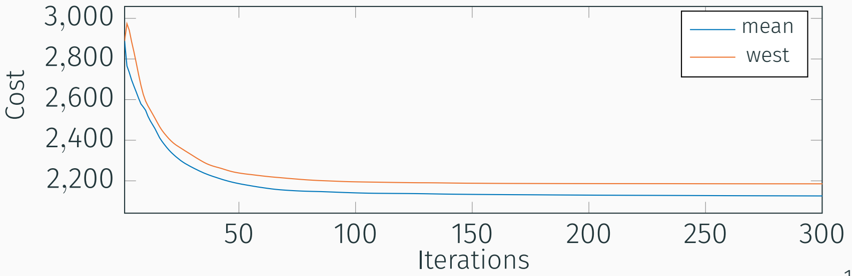
Base point Effect on $\mathbb{S}^2$-valued data



Result, m the mean (p. Px.)



Result, m west (p. Px.)



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.

Example.

```
pOpt = linearizedChambollePock(M, N, cost,
 p, ξ, m, n, DΛ, 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.
- RB; 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.
- RB; Herzog, R.; Tenbrinck, D.; Vidal-Núñez, J. (2019). *Fenchel-Duality for Convex Optimization 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., pp. xviii+451.

ronnybergmann.net/talks/2019-Berlin-RiemannianChambollePock.pdf