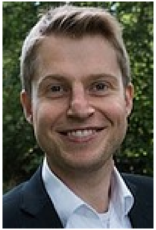
Higher-Order Non-Smooth Optimization on Manifolds





Willem Diepeveen, University of Lübeck/TU Berlin, now Cambridge University
Caterina Rust and Jan Lellmann, University of Lübeck

Oberwolfach Mini-Workshop on Computational Optimization on Manifolds
18 November 2020

MIC Institute of Mathematics and Image Computing

UNIVERSITÄT ZU LÜBECK

Setting

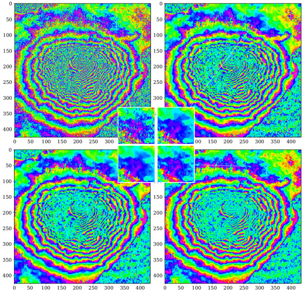
Problem

Find the minimizer of

$$\inf_{x \in \mathcal{M}} F(x)$$

- $\mathcal{M}$ Riemannian manifold
- objective $F : \mathcal{M} \to \mathbb{R}$ is non-smooth

Why?

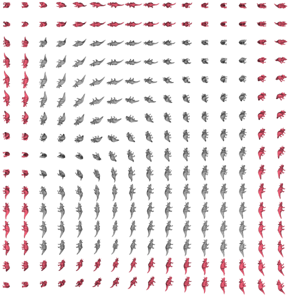


InSAR

$\mathcal{S}(1)$

DW-MRI

$\mathcal{P}(3)$



Orientations

$SO(3)$

- Typical in function space: Given $b : \Omega \to \mathcal{M}$, find minimizer $u^* : \Omega \to \mathcal{M}$ of

$$\inf_{u:\Omega\to\mathcal{M}} \left\{ \int_\Omega d_{\mathcal{M}}(u(x), b(x))^p + \lambda \mathrm{TV}_{\mathcal{M}}(u) \right\}$$

InSAR: Rocca, Prati, Ferreti'97; DWI: Human Connectome Project; SO(3) interpolation: Vogt, Strekalovskiy, Cremers, Lellmann'20; Absil, Gousenbourger, Striewski, Wirth'16; Gousenbourger, Massart, Absil'19; Denoising overview: Bergmann, Laus, Persch, Steidl'18

Higher-order Non-smooth Optimization on Manifolds

Higher-order **Non-smooth Optimization** on Manifolds

The Euclidean Convex Case

Euclidean model

$$\inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax)\}$$

- $A : \mathbb{R}^n \to \mathbb{R}^m$ linear operator, often discretizes derivatives
- $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, $g : \mathbb{R}^m \to \bar{\mathbb{R}}$ convex but non-smooth

The Euclidean Convex Case

Euclidean model

$$\inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax)\}$$

- $A : \mathbb{R}^n \to \mathbb{R}^m$ linear operator, often discretizes derivatives
- $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, $g : \mathbb{R}^m \to \bar{\mathbb{R}}$ convex but non-smooth

Saddle-point formulation

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^m} \{f(x) + \langle y, Ax \rangle - g^*(y)\}$$

g^* Legendre-Fenchel conjugate: $g^*(y) := \sup_{y' \in \mathbb{R}^m} \{\langle y', y \rangle - g(y)\}$

The Euclidean Convex Case

Euclidean model

$$\inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax)\}$$

- $A : \mathbb{R}^n \to \mathbb{R}^m$ linear operator, often discretizes derivatives
- $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, $g : \mathbb{R}^m \to \bar{\mathbb{R}}$ convex but non-smooth

Saddle-point formulation

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^m} \{f(x) + \langle y, Ax \rangle - g^*(y)\}$$

g^* Legendre-Fenchel conjugate: $g^*(y) := \sup_{y' \in \mathbb{R}^m} \{\langle y', y \rangle - g(y)\}$

Dual problem

$$\sup_{y \in \mathbb{R}^m} \{-g^*(y) - f^*(-A^\top y)\}$$

Optimality Conditions

Saddle-point formulation

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^m} \{f(x) + \langle y, Ax \rangle - g^*(y)\}$$

Optimality Conditions

Saddle-point formulation

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^m} \{ f(x) + \langle y, Ax \rangle - g^*(y) \}$$

Primal-dual optimality conditions

$$\begin{aligned} 0 &\in A^\top y^* + \partial f(x^*) \\ 0 &\in -Ax^* + \partial g^*(y^*) \end{aligned}$$

Optimality Conditions

Saddle-point formulation

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^m} \{f(x) + \langle y, Ax \rangle - g^*(y)\}$$

Primal-dual optimality conditions

$$\begin{aligned} 0 &\in A^\top y^* + \partial f(x^*) \\ 0 &\in -Ax^* + \partial g^*(y^*) \end{aligned}$$

- For non-smooth f, g^* this is a system of inclusions – non-unique, hard to manage

Proximal map

Proximal map

$$\operatorname{prox}_{\sigma f}(x) := \arg\min_{x' \in \mathbb{R}^n} \left\{ \frac{1}{2}\|x' - x\|_2^2 + \sigma f(x') \right\}, \quad \sigma > 0.$$

Important property:

$$0 \in z + \partial f(x) \iff 0 = x - \operatorname{prox}_{\sigma f}(x - \sigma z)$$

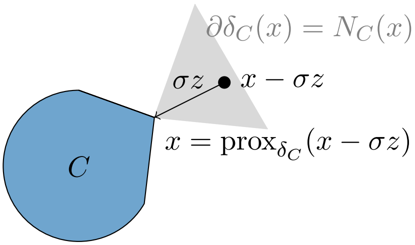
Proximal map

Proximal map

$$\text{prox}_{\sigma f}(x) := \arg\min_{x' \in \mathbb{R}^n} \left\{ \frac{1}{2}\|x' - x\|_2^2 + \sigma f(x') \right\}, \quad \sigma > 0.$$

Important property:

$$0 \in z + \partial f(x) \iff 0 = x - \text{prox}_{\sigma f}(x - \sigma z)$$



Proximal map

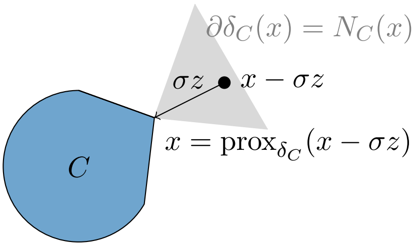
Proximal map

$$\mathrm{prox}_{\sigma f}(x) := \arg\min_{x' \in \mathbb{R}^n} \left\{ \frac{1}{2}\|x' - x\|_2^2 + \sigma f(x') \right\}, \quad \sigma > 0.$$

Important property:

$$0 \in z + \partial f(x) \iff 0 = x - \mathrm{prox}_{\sigma f}(x - \sigma z)$$

This turns an inclusion into an equality!



Optimality conditions II

Primal-dual optimality conditions, inclusion form

$$\begin{aligned} 0 &\in A^\top y^* + \partial f(x^*) \\ 0 &\in -Ax^* + \partial g^*(y^*) \end{aligned}$$

Optimality conditions II

Primal-dual optimality conditions, inclusion form

$$
\begin{aligned}
0 &\in A^\top y^* + \partial f(x^*) \\
0 &\in -Ax^* + \partial g^*(y^*)
\end{aligned}
$$

Primal-dual optimality conditions, equality form

$$
\begin{aligned}
0 &= x - \mathrm{prox}_{\sigma f}(x - \sigma A^\top y) \\
0 &= y - \mathrm{prox}_{\tau g^*}(y + \tau A x)
\end{aligned}
$$

First-Order Methods

Primal-dual optimality conditions, equality form

$$
\begin{aligned}
0 &= x - \operatorname{prox}_{\sigma f}(x - \sigma A^\top y) \\
0 &= y - \operatorname{prox}_{\tau g^*}(y + \tau A x)
\end{aligned}
$$

Chambolle, Pock'11; non-linear A: Valkonen'14; Clason, Mazurenko, Valkonen'18; Clason, Mazurenko, Valkonen'20

First-Order Methods

Primal-dual optimality conditions, equality form

$$
\begin{aligned}
0 &= x - \operatorname{prox}_{\sigma f}(x - \sigma A^\top y) \\
0 &= y - \operatorname{prox}_{\tau g^*}(y + \tau A x)
\end{aligned}
$$

Fixed-point iteration

$$
\begin{aligned}
x^{k+1} &:= \operatorname{prox}_{\sigma f}(x^k - \sigma A^\top y^k) \\
y^{k+1} &:= \operatorname{prox}_{\tau g^*}(y^k + \tau A x^k)
\end{aligned}
$$

Chambolle, Pock'11; non-linear A: Valkonen'14; Clason, Mazurenko, Valkonen'18; Clason, Mazurenko, Valkonen'20

First-Order Methods

Primal-dual optimality conditions, equality form

$$\begin{aligned} 0 &= x - \mathrm{prox}_{\sigma f}(x - \sigma A^\top y) \\ 0 &= y - \mathrm{prox}_{\tau g^*}(y + \tau A x) \end{aligned}$$

Fixed-point iteration

$$\begin{aligned} y^{k+1} &:= \mathrm{prox}_{\tau g^*}(y^k + \tau A x^k) \\ x^{k+1} &:= \mathrm{prox}_{\sigma f}(x^k - \sigma A^\top y^k) \end{aligned}$$

Chambolle, Pock'11; non-linear A: Valkonen'14; Clason, Mazurenko, Valkonen'18; Clason, Mazurenko, Valkonen'20

First-Order Methods

Primal-dual optimality conditions, equality form

$$
\begin{aligned}
0 &= x - \operatorname{prox}_{\sigma f}(x - \sigma A^\top y) \\
0 &= y - \operatorname{prox}_{\tau g^*}(y + \tau A x)
\end{aligned}
$$

Fixed-point iteration + Gauss-Seidel + overrelaxation = Primal-Dual-Hybrid-Gradient

$$
\begin{aligned}
y^{k+1} &:= \operatorname{prox}_{\tau^k g^*}(y^k + \tau^k A(x^k + \theta^k(x^k - x^{k-1}))) \\
x^{k+1} &:= \operatorname{prox}_{\sigma^k f}(x^k - \sigma^k A^\top y^{k+1})
\end{aligned}
$$

Extensions: non-linear A, acceleration schemes for $\theta^k, \tau^k, \sigma^k$ for strong convexity, preconditioning (non-scalar τ, σ), …

PDHG: Pock, Cremers, Bischof, Chambolle'09; Esser, Zhang, Chan'10; Chambolle, Pock'11; non-linear A: Valkonen'14; Clason, Mazurenko, Valkonen'18; Clason, Mazurenko, Valkonen'20

Higher-order **Non-smooth Optimization** on Manifolds

Higher-order **Non-smooth Optimization on Manifolds**

Manifold Setting

Problem

$$\inf_{p\in\mathcal{M}} \{f(p) + g(\Lambda(p))\},$$

- $\mathcal{M}, \mathcal{N}$ smooth Riemannian manifolds
- $\Lambda : \mathcal{M} \to \mathcal{N}$ (non-linear!) mapping
- $f : \mathcal{M} \to \overline{\mathbb{R}}$, $g : \mathcal{N} \to \overline{\mathbb{R}}$ functions, non-smooth

Duality & saddle-point formulation: Bergmann, Herzog, Tenbrinck, Vidal-Núñez'19

Manifold Setting

Problem

$$\inf_{p\in\mathcal{M}} \{f(p) + g(\Lambda(p))\},$$

- $\mathcal{M}, \mathcal{N}$ smooth Riemannian manifolds
- $\Lambda : \mathcal{M} \to \mathcal{N}$ (non-linear!) mapping
- $f : \mathcal{M} \to \overline{\mathbb{R}}$, $g : \mathcal{N} \to \overline{\mathbb{R}}$ functions, non-smooth

Saddle-Point Problem [Bergmann et al.'19]

For a fixed base point $n \in \mathcal{N}$ and cotangent space $T_n^*\mathcal{N}$:

$$\inf_{p\in\mathcal{M}} \sup_{\xi_n \in T_n^*} \{f(p) + \langle \log_n \Lambda(p), \xi_n \rangle - g_n^*(\xi_n)\},$$

where $g_n^*(\xi_n) := \sup_{X\in\mathcal{L}_{\mathcal{C},n}} \{\langle \xi_n, X\rangle - g(\exp_n X)\}$ is the n-conjugate.

- For proper lsc convex g, we get $g_{nn}^{**} = g$ and Fenchel: $f(p) + f_m^*(\xi_m) \geqslant \langle \xi_m, \log_m p\rangle$.

Duality & saddle-point formulation: Bergmann, Herzog, Tenbrinck, Vidal-Núñez'19

Linearized optimality system

Saddle-Point

$$\inf_{p\in\mathcal{M}} \sup_{\xi_n \in T_n^*} \{f(p) + \langle \log_n \Lambda(p), \xi_n\rangle - g_n^*(\xi_n)\},$$

- Disadvantage: No good concept of adjoint for Λ.

Bergmann, Herzog, Tenbrinck, Vidal-Núñez'19; Linearization in Euclidean case: Valkonen'14; Clason, Mazurenko, Valkonen'18; Clason, Mazurenko, Valkonen'20

Linearized optimality system

Saddle-Point

$$\inf_{p\in\mathcal{M}} \sup_{\xi_n\in T_n^*} \{f(p) + \langle \log_n \Lambda(p), \xi_n\rangle - g_n^*(\xi_n)\},$$

- Disadvantage: No good concept of adjoint for Λ.

Linearized Saddle-Point

Around base point m, approximate $\Lambda(p) \approx \exp_{\Lambda(m)} D_m\Lambda[\log_m p]$, where $D_m F[v]$ differential of F at $m \in \mathcal{M}$ applied to $v \in T_p\mathcal{M}$. Then for $n := \Lambda(m)$,

$$\inf_{p\in\mathcal{M}} \sup_{\xi_n\in T_n^*} \{f(p) + \langle D_m\Lambda[\log_m p], \xi_n\rangle - g_n^*(\xi_n)\}.$$

Bergmann, Herzog, Tenbrinck, Vidal-Núñez'19; Linearization in Euclidean case: Valkonen'14; Clason, Mazurenko, Valkonen'18; Clason, Mazurenko, Valkonen'20

Linearized optimality system

Saddle-Point

$$\inf_{p\in\mathcal{M}} \sup_{\xi_n\in T_n^*} \{f(p) + \langle \log_n \Lambda(p), \xi_n\rangle - g_n^*(\xi_n)\},$$

- Disadvantage: No good concept of adjoint for Λ.

Linearized Saddle-Point

Around base point m, approximate $\Lambda(p) \approx \exp_{\Lambda(m)} D_m\Lambda[\log_m p]$, where $D_m F[v]$ differential of F at $m \in \mathcal{M}$ applied to $v \in T_p\mathcal{M}$. Then for $n := \Lambda(m)$,

$$\inf_{p\in\mathcal{M}} \sup_{\xi_n\in T_n^*} \{f(p) + \langle D_m\Lambda[\log_m p], \xi_n\rangle - g_n^*(\xi_n)\}.$$

- Can show weak duality on Hadamard manifolds for proper, lsc, convex problems
- Saddle point solves linearized problem

Bergmann, Herzog, Tenbrinck, Vidal-Núñez'19; Linearization in Euclidean case: Valkonen'14; Clason, Mazurenko, Valkonen'18; Clason, Mazurenko, Valkonen'20

Optimality by Equality on Manifolds

Primal-dual optimality conditions on manifolds [Bergmann et al.'19]

Manifold case: $p \in \mathcal{M}$, $\xi_n \in T_n^*$, fixed base point $n \in \mathcal{N}$, linearized version:

$$\begin{aligned} p &= \operatorname{prox}_{\sigma f}\Bigl(\exp_p\Bigl(\mathcal{P}_{m\to p}\bigl(-\sigma\,(D_m\Lambda)^*\,[\mathcal{P}_{n\to\Lambda(m)}\xi_n]\bigr)^\sharp\Bigr)\Bigr)\\ \xi_n &= \operatorname{prox}_{\tau g_n^*}\Bigl(\xi_n + \tau\,\bigl(\mathcal{P}_{\Lambda(m)\to n} D_m\Lambda\,[\log_m p]\bigr)^\flat\Bigr) \end{aligned}$$

Compare Euclidean case: $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$:

$$\begin{aligned} 0 &= x - \operatorname{prox}_{\sigma f}(x - \sigma A^\top y)\\ 0 &= y - \operatorname{prox}_{\tau g^*}(y + \tau A x) \end{aligned}$$

Requires linearization of primal problem around m, but...

- Can construct first-order Linearized Riemannian Chambolle-Pock method (lRCPA)
- lRCPA converges to saddle point for proper, convex, lsc problems on Hadamard manifolds

Other Non-Smooth First-Order Methods on Manifolds

- Subgradient descent, gradient sampling
- Primal proximal steps:
  - Cyclic Proximal Point Algorithm (CPPA)
  - Iteratively Reweighted Least Squares (IRLS)
  - Parallel Douglas-Rachford (PDR)
- Specialized/exact algorithms/dynamic programming

Convergence generally for Hadamard manifolds, but also seem to work on manifolds with positive sectional curvature.

Subgradient: Ferreira, Oliveira'98; Sampling: Hosseini, Uschmajew'17; CPPA: Bacák'14; Bacák, Bergmann, Steidl, Weinmann'16; IRLS: Bergmann, Chan, Hielscher, Persch, Steidl'15; PDR: Bergmann, Persch, Steidl'16; Proximal method: de Carvalho Bento, da Cruz Neto, Oliveira'16; Specialized: Weinmann, Demaret, Storath'16; Storath, Weinmann, Unser'19

Higher-order **Non-smooth Optimization on Manifolds**

Higher-order Non-smooth Optimization on Manifolds

Higher-Order Methods

- First-order methods converge sublinearly or linearly (strongly convex primal & dual problems) – very slow tail convergence
- Goal: apply higher-order methods for superlinear convergence. Choices:
  - Extrinsic approach & use Euclidean solver
  - Smooth energy & use Riemannian (Quasi-)Newton-type method/Trust region/CG
  - Here: Semismooth Newton

Smooth problems: Udriste'93; CG: Ring, Wirth'12; Sato, Iwai'15; Boumal, Absil, Cartis'18; Books: Absil, Mahony, Sepulchre'08; Boumal'20

Euclidean Semismooth Newton

Problem

Solve non-linear system equation

$$X(x) = 0,$$

where $X : \mathbb{R}^n \to \mathbb{R}^m$ is (Lipschitz-)continuous but not C^1.

Typical application:

$$x^* \in \arg\min_{x \in \mathbb{R}^n} F(x)$$
$$\text{s.t. } x \leqslant 0$$

Overview: Hintermüller'10 MFO Seminar; Clason'17 Lecture Notes

Euclidean Semismooth Newton

Problem

Solve non-linear system equation

$$X(x) = 0,$$

where $X : \mathbb{R}^n \to \mathbb{R}^m$ is (Lipschitz-)continuous but not C^1.

Typical application:

$$\begin{aligned} x^* \in \arg\min_{x \in \mathbb{R}^n} F(x) \\ \text{s.t. } x \leqslant 0 \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} 0 = \nabla F(x^*) + \lambda \\ x \leqslant 0,\ \lambda \geqslant 0,\ \lambda x = 0 \end{aligned}$$

Overview: Hintermüller'10 MFO Seminar; Clason'17 Lecture Notes

Euclidean Semismooth Newton

Problem

Solve non-linear system equation

$$X(x) = 0,$$

where $X : \mathbb{R}^n \to \mathbb{R}^m$ is (Lipschitz-)continuous but not C^1.

Typical application:

$$\begin{array}{c} x^* \in \arg\min_{x \in \mathbb{R}^n} F(x) \\ \text{s.t. } x \leqslant 0 \end{array} \iff \begin{array}{c} 0 = \nabla F(x^*) + \lambda \\ x \leqslant 0,\ \lambda \geqslant 0,\ \lambda x = 0 \end{array} \iff \begin{array}{c} 0 = \nabla F(x^*) + \lambda \\ 0 = \lambda - \max\{0, \lambda + \sigma x\} \end{array}$$

Overview: Hintermüller'10 MFO Seminar; Clason'17 Lecture Notes

Euclidean Semismooth Newton

Algorithm 1 Semismooth Newton

Initialization: $x^0 \in \mathbb{R}^n$
for $k = 0, 1, \ldots$ **do**
  Choose any $V(x^k) \in \partial_C X(x^k)$
  Solve $V(x^k)d^k = -X(x^k)$
  $x^{k+1} := x^k + d^k$
end for

Theorem (consequence of [Qi, Sun'93])

Assume that x^* satisfies $X(x^*) = 0$, X is locally Lipschitz and semismooth at x^* and all $V \in \partial_C X(x^*)$ are nonsingular. Then the iteration in Alg. 1 is well-defined and converges superlinearly to x^* in a neighborhood of x^*. If in addition X is μ-order semismooth at x^*, then the convergence is of order $1 + \mu$.

Euclidean Semismoothness

Definition

Let $X : U \to \mathbb{R}^n$ be defined on the open set $U \subset \mathbb{R}^n$. Then, for $0 < \mu \leq 1$, X is called μ-order semismooth at $x \in U$ if X is locally Lipschitz at x, all one-sided directional derivatives $X'(x, \cdot)$ exist, and, for every $V \in \partial_C X(x+d)$ in the generalized Clarke differential,

$$\left\|Vd - X'(x,d)\right\| = \mathcal{O}\left(\|d\|^{1+\mu}\right) \qquad \text{as } d \to 0.$$

If X is μ-order semismooth at all $x \in U$, then we call X μ-order semismooth (on U).

Semismoothness

Equivalent: X is locally Lipschitz continuous at x, $X'(x, \cdot)$ exist, and, for every $V \in \partial_C X(x+d)$ it holds

$$\|X(x+d) - (X(x) + Vd)\| = \mathcal{O}\left(\|d\|^{1+\mu}\right) \qquad \text{as } d \to 0.$$

Compare definition of differentiability in smooth case, where $V = \nabla X(x)$.

- Often fulfilled: norms, convex functions, piecewise C^1, piecewise affine

Semismoothness: Mifflin'77; Properties: Sun'01; Xiao, Li, Wen, Zhang'18; Overview: Hintermüller'10 MFO Seminar

Non-Smooth Euclidean SSN

Optimality Conditions

$$x^* \in \arg\min_{x \in \mathbb{R}^n} F(x) \quad \text{s.t. } x \leqslant 0 \quad \Leftrightarrow \quad \begin{array}{c} 0 = \nabla F(x^*) + \lambda \\ x \leqslant 0,\ \lambda \geqslant 0,\ \lambda x = 0 \end{array} \quad \Leftrightarrow \quad \begin{array}{c} 0 = \nabla F(x^*) + \lambda \\ 0 = \lambda - \max\{0, \lambda + \sigma x\} \end{array}$$

Euclidean case: Rust thesis'17; Xiao, Li, Wen, Zhang'16/18

Non-Smooth Euclidean SSN

Optimality Conditions

$$0 \overset{!}{=} X(x,y) := \begin{pmatrix} x - \operatorname{prox}_{\sigma f}(x - \sigma A^\top y) \\ y - \operatorname{prox}_{\tau g^*}(y + \tau A x) \end{pmatrix}$$

Euclidean case: Rust thesis'17; Xiao, Li, Wen, Zhang'16/18

Non-Smooth Euclidean SSN

Optimality Conditions

$$0 \stackrel{!}{=} X(x,y) := \begin{pmatrix} x - \operatorname{prox}_{\sigma f}(x - \sigma A^\top y) \\ y - \operatorname{prox}_{\tau g^*}(y + \tau A x) \end{pmatrix}$$

Newton system

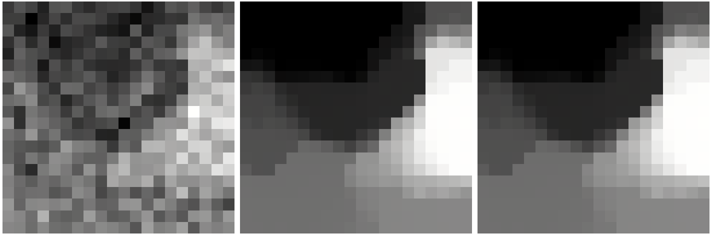
$$\begin{pmatrix} I - K & \sigma K A^\top \\ -\tau H A & I - H \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -X(x,y)$$

where $K \in \partial_C \operatorname{prox}_{\sigma f}(x - \sigma A^\top y)$, $H \in \partial_C \operatorname{prox}_{\tau g^*}(y + \tau A x)$.

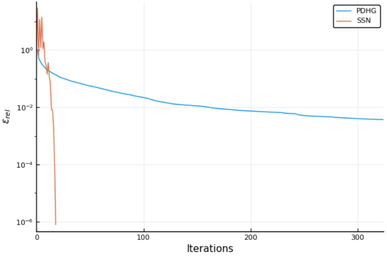
- Lipschitz continuous, positive semidefinite at solution [Rust'17]
- Semismoothness not clear in general, but relatively straightforward to check
- Invertibility can be an issue – for TV, dual regularization helps [Diepeveen'20]
- Globalization not finally solved [XLWZ18]

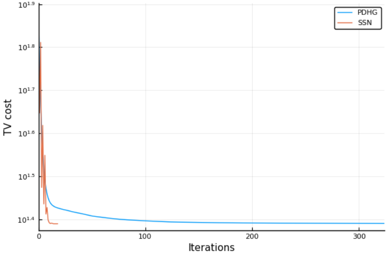
Euclidean case: Rust thesis'17; Xiao, Li, Wen, Zhang'16/18

Euclidean Case $\ell^2 - \mathrm{TV}$ Runtimes

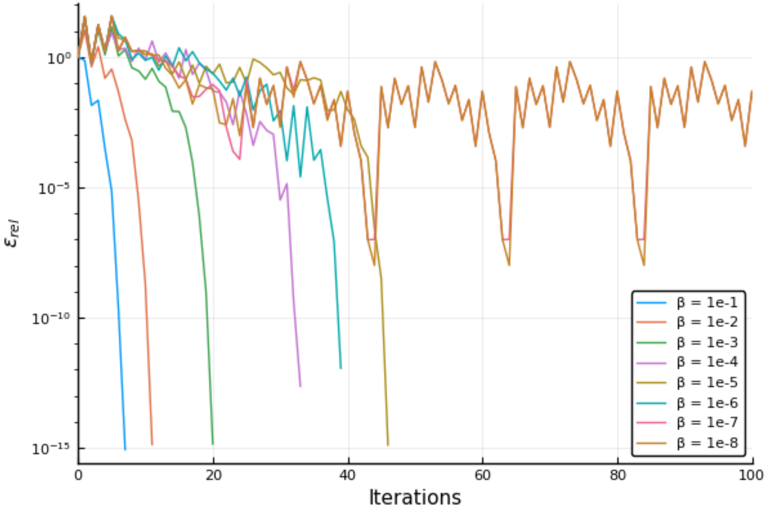


(a) Noisy image **(b)** Result PDHG **(c)** Result SSN





Dual regularization



Higher-order Non-smooth Optimization on Manifolds

Higher-order Non-smooth Optimization on Manifolds

Semismooth Newton on Manifolds

Goal

Find $p \in \mathcal{M}$ that solves

$$X(p) = 0,$$

where X is a locally Lipschitz continuous vector field on $\mathcal{M}$.

Algorithm 2 Riemannian Semismooth Newton

Initialization: $p^0 \in \mathcal{M}$
for $k = 0, 1, \ldots$ **do**
  Choose any $V(p^k) \in \partial_{\mathcal{M},C} X(p^k)$
  Solve $V(p^k)d^k = -X(p^k)$ in the vector space $\mathcal{T}_{p^k}\mathcal{M}$
  $p^{k+1} := \exp_{p^k}(d^k)$
end for

Convergence of Semismooth Newton on Manifolds for general X

Theorem (Oliveira, Ferreira'20)

Let X be a locally Lipschitz continuous vector field on $\mathcal{M}$ and $p^* \in \mathcal{M}$ be a solution of problem

$$X(p) = 0,$$

Assume that X is semismooth at p^* and all $V_{p^*} \in \partial_{\mathcal{M},C} X(p^*)$ are invertible. Then, there exists a $\delta > 0$ such that for each $p^0 \in B_\delta(p^*) \setminus \{p^*\}$, $(p^k)_{k\geq 0}$ generated by RSSN is well-defined, belongs to $B_\delta(p^*)$ and converges superlinearly to p^* Additionally, if X is μ-order semismooth at p^*, then the convergence of $(p^k)_{k\geq 0}$ to p^* is of order $1+\mu$.

General formulation for vector field equations.

Semismooth Newton on Manifolds – Nonsmooth Case

Riemannian Semismooth Newton [Diepeveen'20]

$$
\begin{aligned}
p &= \operatorname{prox}_{\sigma f}\left(\exp_p\left(\mathcal{P}_{m\to p}\left(-\sigma\left(D_m\Lambda\right)^*\left[\mathcal{P}_{n\to\Lambda(m)}\xi_n\right]\right)^\sharp\right)\right)\\
\xi_n &= \operatorname{prox}_{\tau g_n^*}\left(\xi_n+\tau\left(\mathcal{P}_{\Lambda(m)\to n}D_m\Lambda\left[\log_m p\right]\right)^\flat\right)
\end{aligned}
$$

Diepeveen'20

Semismooth Newton on Manifolds – Nonsmooth Case

Riemannian Semismooth Newton [Diepeveen'20]

$$
\begin{aligned}
p &= \operatorname{prox}_{\sigma f}\left(\exp_p\left(\mathcal{P}_{m\to p}\left(-\sigma\left(D_m\Lambda\right)^*\left[\mathcal{P}_{n\to\Lambda(m)}\xi_n\right]\right)^\sharp\right)\right)\\
\xi_n &= \operatorname{prox}_{\tau g_n^*}\left(\xi_n + \tau\left(\mathcal{P}_{\Lambda(m)\to n} D_m\Lambda\left[\log_m p\right]\right)^\flat\right)
\end{aligned}
$$

Unfortunately not a vector field...

Semismooth Newton on Manifolds – Nonsmooth Case

Riemannian Semismooth Newton [Diepeveen'20]

$$
\begin{aligned}
0 &= -\log_p \operatorname{prox}_{\sigma f}\left(\exp_p\left(\mathcal{P}_{m\to p}\left(-\sigma\left(D_m\Lambda\right)^*\left[\mathcal{P}_{n\to\Lambda(m)}\xi_n\right]\right)^\sharp\right)\right)\\
0 &= \xi_n - \operatorname{prox}_{\tau g_n^*}\left(\xi_n + \tau\left(\mathcal{P}_{\Lambda(m)\to n} D_m\Lambda\left[\log_m p\right]\right)^\flat\right)
\end{aligned}
$$

Semismooth Newton on Manifolds – Nonsmooth Case

Riemannian Semismooth Newton [Diepeveen'20]

$$0 = X(p, \xi_n) := \begin{pmatrix} -\log_p \operatorname{prox}_{\sigma f}\left(\exp_p\left(\mathcal{P}_{m\to p}\left(-\sigma\left(D_m\Lambda\right)^*\left[\mathcal{P}_{n\to\Lambda(m)}\xi_n\right]\right)^\sharp\right)\right) \\ \xi_n - \operatorname{prox}_{\tau g_n^*}\left(\xi_n + \tau\left(\mathcal{P}_{\Lambda(m)\to n} D_m\Lambda\left[\log_m p\right]\right)^\flat\right) \end{pmatrix}$$

- find zero of vector field $X : \mathcal{M} \times T_n^*\mathcal{M} \to T\mathcal{M} \times T_n^*\mathcal{M}$
- prox steps are still non-smooth

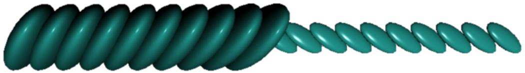
Implementation

- By W. Diepeveen'20, based on `manopt.jl`
- Chain rule for covariant derivative
- Jacobi fields for evaluating parts in chain rule/derivatives of $d_{\mathcal{M}}$, $\log_p$, $\exp_p$
- Pole ladder approximation for differentiating parallel transport, exact on symmetric manifolds

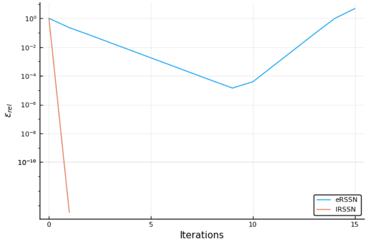
manopt.jl: Bergmann'19; Manopt: Boumal, Mishra, Absil, Sepulchre; Pole ladder: Pennec'18

Case 1b: $\mathcal{P}(3)$ Signals with Known Minimizer

(a) Original



(b) RSSN solution



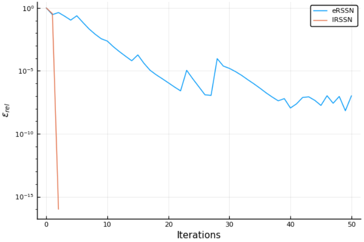
(c) Progression of relative error

Data: Bergmann, Herzog, Tenbrinck, Vidal-Núñez'19

Case 1a: S^2 Signals with Known Minimizer

(a) Original

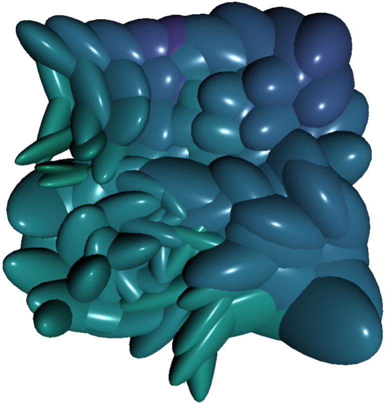
(b) RSSN solution



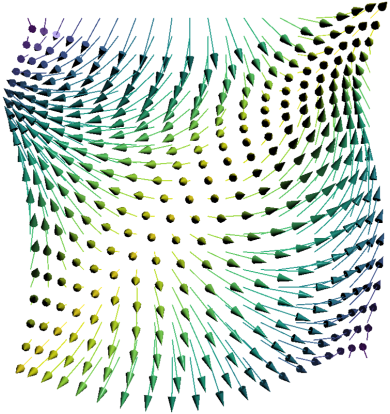
(c) Progression of relative error

Data: Bergmann, Herzog, Tenbrinck, Vidal-Núñez’19

Case 2: Manifold-valued Images

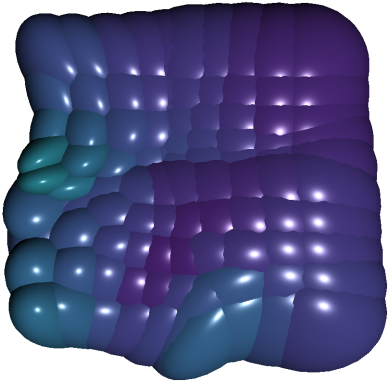


(a) Original $\mathcal{P}(3)$ data

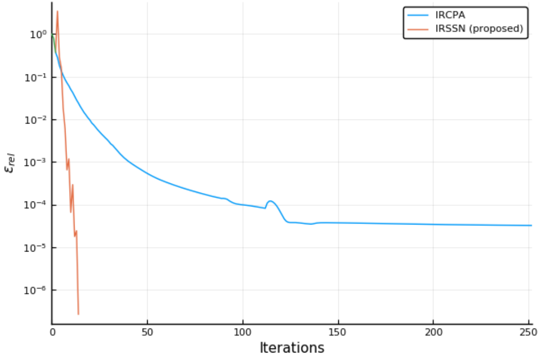


(b) Original S^2 data

Case 2a: $\mathcal{P}(3)$

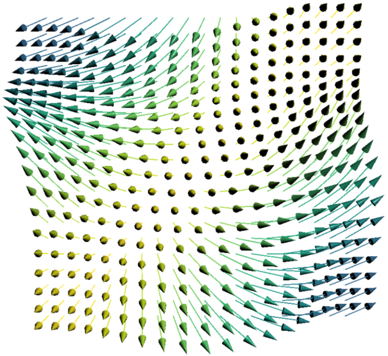


(a) Solution RSSN

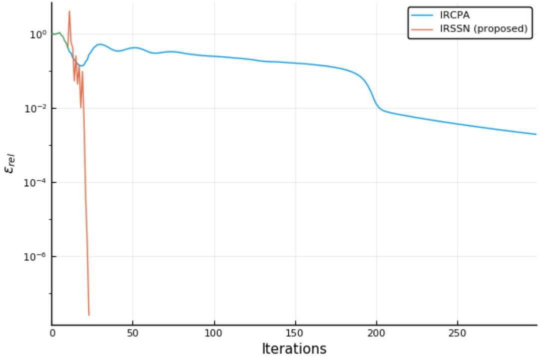


(b) Progression of relative error

Case 2b: S^2



(a) Solution RSSN



(b) Progression of relative error

Inexact Semismooth Newton on Manifolds

Algorithm 3 Inexact Semismooth Newton

Initialization: $p^0 \in \mathcal{M}, a^0 \geq 0$
for $k = 0, 1, \ldots$ **do**
  Choose $V_k(p^k) \in \partial_{\mathcal{M},C} X(p^k)$
  Solve $V_k(p^k)d^k = -X(p^k) + r^k$ in $\mathcal{T}_{p^k}\mathcal{M}$ where $\|r^k\|_{(p^k)} \leq a^k \|X(p^k)\|_{(p^k)}$
  $p^{k+1} := \exp_{p^k}(d^k)$
  Choose $a^{k+1} \geq 0$
end for

Convergence for Inexact RSSN

Theorem (Thesis [Diepeveen, L.'20])

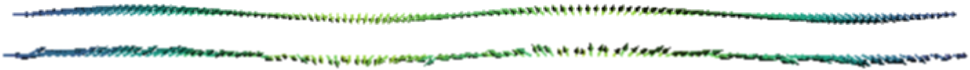
Let X be locally Lipschitz continuous vector field on $\mathcal{M}$ and $p^* \in \mathcal{M}$ be a solution of problem the $X(p) = 0$. Assume that X is semismooth at p^* and that all $V_{p^*} \in \partial_{\mathcal{M},C} X(p^*)$ are invertible. Then the following statements hold:

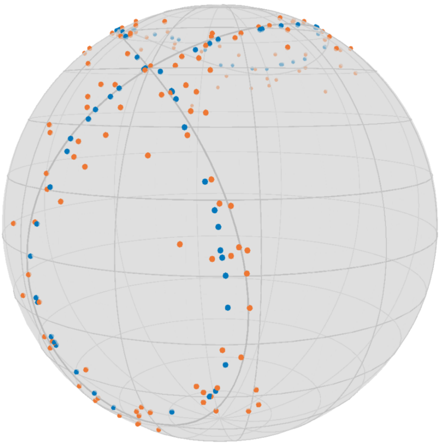
(i) *There exist $a > 0$ and $\delta > 0$ such that for every $p^0 \in B_\delta(p^*)$ and $a^k \leq a$, the sequence $(p^k)_{k\geq 0}$ generated by Alg. 3 is well-defined, is contained in $B_\delta(p^*)$ and converges Q-linearly to the solution p^*.*

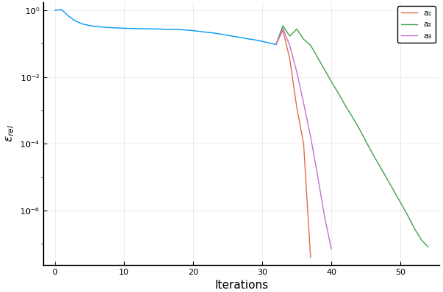
(ii) *If the sequence $(p^k)_{k\geq 0}$ generated by Alg. 3 converges to the solution p^* and further $\|r^k\|_{(p^k)} \in o\left(\|X(p^k)\|_{(p^k)}\right)$, then the rate of convergence is Q-superlinear.*

(iii) *If the sequence $(p^k)_{k\geq 0}$ generated by Alg. 3 converges to the solution p^*, X is μ-order semismooth at p^*, and $\|r^k\|_{(p^k)} \in O\left(\|X(p^k)\|_{(p^k)}^{1+\mu}\right)$, then the rate of convergence is Q-order $1+\mu$.*

Lemniscate using Inexact RSSN



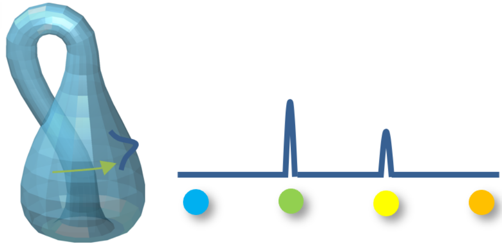




(a) blue: IRCPA pre-steps. **red:** exact I RSSN steps $a_1^k = 0$, **magenta:** increasing accuracy $a_3^k = \frac{1}{5k}$ for superlinear convergence, **green:** constant accuracy $a_2^k = \frac{1}{5}$ for linear convergence.

Lifting Approaches

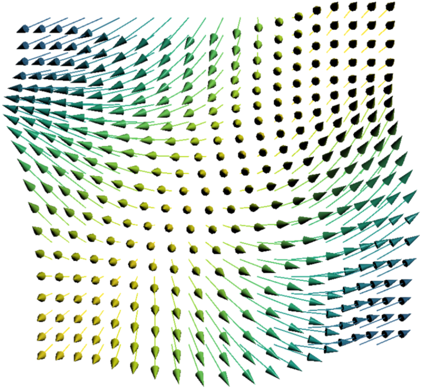
- "Lift" problem to space of probability measures over $\mathcal{M}$
- Higher-dimensional but convex – global optimality possible
- Works on discrete spaces
- Also in $\mathbb{R}^n$ to handle non-convexity
- Connections to dynamical optimal transport



Overview Lifting on Manifolds: Vogt, Lellmann'20; Connection to Dynamical OT: Vogt, Haase, Bednarski, Lellmann'20

Summary

- Non-smooth problem
  $\longrightarrow \mathrm{prox}$-optimality system
  $\longrightarrow$ Semismooth Newton
- Inexact Semismooth Newton with local convergence proof also for non-Hadamard case
- Implementation in `manopt.jl`
- Faster convergence confirmed in practice



https://www.mic.uni-luebeck.de/people/

https://www.mic.uni-luebeck.de/research-and-projects/student-projects.html

lellmann@mic.uni-luebeck.de