Nonsmooth Optimization on Riemannian Manifolds

Hajg Jasa

Kolloquium über Angewandte Mathematik
Institut für Numerische und Angewandte Mathematik, Göttingen

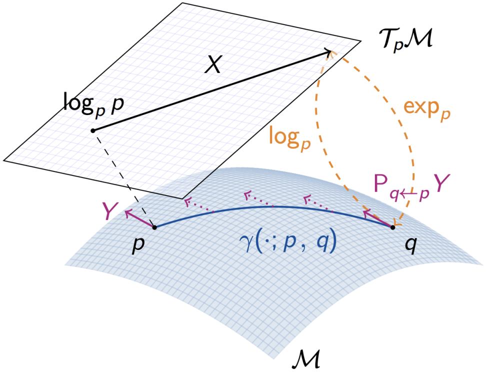
December 10th, 2024

Riemannian Geometry

A d-dimensional Riemannian manifold can be defined informally as a set $\mathcal{M}$ covered with a suitable collection of charts that identify subsets of $\mathcal{M}$ with open subsets of $\mathbb{R}^d$, paired with a continuously varying inner product on the tangent spaces $\mathcal{T}_p\mathcal{M}$. [Absil, Mahony, and Sepulchre 2008]

Riemannian Geometry

- Geodesic
  $\gamma_{pq}\colon [0,1] \to \mathcal{M}$ with initial velocity $\dot\gamma_{pq}(0) = X_p \in \mathcal{T}_p\mathcal{M}$
- Inner product
  $(\cdot,\cdot)_p : \mathcal{T}_p\mathcal{M} \times \mathcal{T}_p\mathcal{M} \to \mathbb{R}$
- Exponential map
  $\exp_p X_p = \gamma_{pq}(1) = q$
- Logarithmic map
  $\log_p q = \exp_p^{-1} q = X_p$
- Parallel transport
  $\mathrm{P}_{q\leftarrow p}\colon \mathcal{T}_p\mathcal{M} \to \mathcal{T}_q\mathcal{M}$
- Sectional curvature κ



The Problem

Goal Find numerically and fast

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} f(p),$$

where $\mathcal{M}$ is a Riemannian manifold.

The Problem

Goal Find numerically and fast

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} f(p),$$

where $\mathcal{M}$ is a Riemannian manifold.

Challenges $\mathcal{M}$ is nonlinear

- find a “good replacement” for the $+$ operation
- no “global space of directions”: $T_p\mathcal{M} \neq T_q\mathcal{M}$

The Problem

Goal Find numerically and fast

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} f(p),$$

where $\mathcal{M}$ is a Riemannian manifold.

Challenges $\mathcal{M}$ is nonlinear

- find a “good replacement” for the $+$ operation
- no “global space of directions”: $T_p\mathcal{M} \neq T_q\mathcal{M}$

Task How to handle the case when f is nonsmooth?

The Problem

Goal Find numerically and fast

$$\operatorname*{arg\,min}_{p\in\mathcal{M}} f(p),$$

where $\mathcal{M}$ is a Riemannian manifold.

Challenges $\mathcal{M}$ is nonlinear

- find a “good replacement” for the $+$ operation
- no “global space of directions”: $T_p\mathcal{M} \neq T_q\mathcal{M}$

Task How to handle the case when f is nonsmooth?

Example application. Given $A, B \in \mathbb{R}^{n\times d}$, $\mathcal{M} = \mathrm{SO}(d)$

- orthogonal Procrustes: $f(p) = \lVert A - Bp\rVert_{\mathrm{F}}$ has a closed form solution.
- spectral Procrustes: $f(p) = \lVert A - Bp\rVert_2$ does not

Geodesic Convexity

- A set $\mathcal{C} \subseteq \mathcal{M}$ is called strongly geodesically convex if for all $p, q \in \mathcal{C}$ the geodesic $\gamma_{pq} \colon [0,1] \to \mathcal{M}$ is unique and lies in $\mathcal{C}$.

Geodesic Convexity

- A set $\mathcal{C} \subseteq \mathcal{M}$ is called strongly geodesically convex if for all $p, q \in \mathcal{C}$ the geodesic $\gamma_{pq} \colon [0,1] \to \mathcal{M}$ is unique and lies in $\mathcal{C}$.

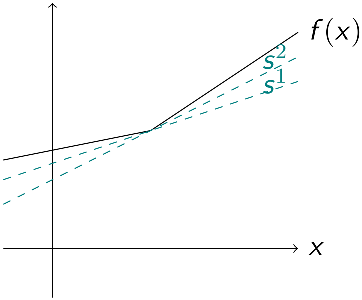
- A function $f \colon \mathcal{C} \to \overline{\mathbb{R}}$ is called geodesically convex if for all $p, q \in \mathcal{C}$ the composition $f \circ \gamma_{pq}(t)$ is convex in the usual sense.

The Convex Subdifferential(s)

For a convex function, the subdifferential is defined as

$$\partial f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T(y - x) \text{ for all } y \in \mathbb{R}^n \right\}$$

and it is a non-empty, closed and convex subset.



The Convex Subdifferential(s)

For a convex function, the subdifferential is defined as

$$\partial f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T (y - x) \text{ for all } y \in \mathbb{R}^n \right\}$$

and it is a non-empty, closed and convex subset.
Let $\varepsilon > 0$. The $\varepsilon-$subdifferential is

$$\partial_\varepsilon f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T (y - x) - \varepsilon \text{ for all } y \in \mathbb{R}^n \right\}$$

and

$$\partial f(x) \subseteq \partial_\varepsilon f(x)$$

The Convex Subdifferential(s)

For a geodesically convex function, the subdifferential is defined as

$$\partial f(p) = \left\{ X_p \in \mathcal{T}_p\mathcal{M} \,\middle|\, f(q) \geq f(p) + \left(X_p, \log_p q\right) \text{ for all } q \in \operatorname{dom} f \right\}$$

and it is a non-empty, closed and convex subset.
Let $\varepsilon > 0$. The $\varepsilon-$subdifferential is

$$\partial_\varepsilon f(p) = \left\{ X_p \in \mathcal{T}_p\mathcal{M} \,\middle|\, f(q) \geq f(p) + \left(X_p, \log_p q\right) - \varepsilon \text{ for all } q \in \operatorname{dom} f \right\}$$

and

$$\partial f(p) \subseteq \partial_\varepsilon f(p)$$

The Riemannian Convex Bundle Method

joint work with

Ronny Bergmann and Roland Herzog

The Problem

Consider the following minimization problem

$$\text{minimize } f(p), \quad p \in \mathcal{M},$$

The Problem

Consider the following minimization problem

$$\text{minimize } f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold

The Problem

Consider the following minimization problem

$$\text{minimize } f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$ on $\operatorname{dom} f$

The Problem

Consider the following minimization problem

$$\operatorname{minimize} f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$ on $\operatorname{dom} f$
- $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ is such that

The Problem

Consider the following minimization problem

$$\text{minimize } f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$ on $\operatorname{dom} f$
- $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ is such that
  - $\operatorname{dom} f \neq \emptyset$ strongly geodesically convex
  - $\operatorname{diam}(\operatorname{dom} f) < \infty$ if $\kappa \neq 0$
  - $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$ in $\mathcal{M}$
  - geodesically convex
  - lower semi-continuous: $\liminf_{q \to p} f(q) \geq f(p)$

The Problem

Consider the following minimization problem

$$\operatorname{minimize} f(p), \quad p \in \mathcal{M},$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$ on $\operatorname{dom} f$
- $f\colon \mathcal{M} \to \overline{\mathbb{R}}$ is such that
  - $\operatorname{dom} f \neq \emptyset$ strongly geodesically convex
  - $\operatorname{diam}(\operatorname{dom} f) < \infty$ if $\kappa \neq 0$
  - $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$ in $\mathcal{M}$
  - geodesically convex
  - lower semi-continuous: $\liminf_{q \to p} f(q) \geq f(p)$

Goal Solve this optimization problem with a convex bundle method.

The Bundle Method

Bundle methods are about descent as well as stability.

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of general bundle algorithms: [Bonnans et al. 2006]

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of general bundle algorithms: [Bonnans et al. 2006]

- keep track of the “best” points

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of general bundle algorithms: [Bonnans et al. 2006]

- keep track of the "best" points
- solve a stabilization subproblem

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of general bundle algorithms: [Bonnans et al. 2006]

- keep track of the "best" points
- solve a stabilization subproblem

  dual approach [Geiger and Kanzow 2002]

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of general bundle algorithms: [Bonnans et al. 2006]

- keep track of the "best" points
- solve a stabilization subproblem

  dual approach [Geiger and Kanzow 2002]
- generate sequences of *candidate* points and *stability centers*

The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of general bundle algorithms: [Bonnans et al. 2006]

- keep track of the "best" points
- solve a stabilization subproblem

  dual approach [Geiger and Kanzow 2002]
- generate sequences of *candidate* points and *stability centers*

Goal Approximate ∂f with $\partial_\varepsilon f$ on $\operatorname{int}(\operatorname{dom} f)$.

Approximating the $\varepsilon-$subdifferential

Given $x^{(0)}, \ldots, x^{(k)} \in \mathbb{R}^n$, and $s^{(j)} \in \partial f(x^{(j)})$ for $j = 0, \ldots, k$, and

Approximating the $\varepsilon-$subdifferential

Given $x^{(0)}, \ldots, x^{(k)} \in \mathbb{R}^n$, and $s^{(j)} \in \partial f(x^{(j)})$ for $j = 0, \ldots, k$, and

$$e_j^{(k)} := f(x^{(k)}) - f(x^{(j)}) - (s^{(j)})^T(x^{(k)} - x^{(j)}),$$

Approximating the $\varepsilon-$subdifferential

Given $x^{(0)}, \dots, x^{(k)} \in \mathbb{R}^n$, and $s^{(j)} \in \partial f(x^{(j)})$ for $j = 0, \dots, k$, and

$$e_j^{(k)} := f(x^{(k)}) - f(x^{(j)}) - (s^{(j)})^T (x^{(k)} - x^{(j)}),$$

then

$$s^{(j)} \in \partial_{e_j^{(k)}} f(x^{(k)}).$$

Approximating the $\varepsilon-$subdifferential

Given $x^{(0)}, \dots, x^{(k)} \in \mathbb{R}^n$, and $s^{(j)} \in \partial f(x^{(j)})$ for $j = 0, \dots, k$, and

$$e_j^{(k)} := f(x^{(k)}) - f(x^{(j)}) - (s^{(j)})^T(x^{(k)} - x^{(j)}),$$

then

$$s^{(j)} \in \partial_{e_j^{(k)}} f(x^{(k)}).$$

This gives an inner approximation of $\partial_\varepsilon f(x^{(k)})$ as:

$$G_\varepsilon^{(k)} := \left\{ \sum_{j=0}^{k} \lambda_j s^{(j)} \,\middle|\, \sum_{j=0}^{k} \lambda_j \, e_j^{(k)} \leq \varepsilon, \ \lambda \in \Delta_k \right\} \subseteq \partial_\varepsilon f(x^{(k)})$$

with $\Delta_k := \{\lambda \in \mathbb{R}^{k+1} \mid \sum_{j=0}^{k} \lambda_j = 1, \ \lambda_j \geq 0, \ j = 0, \dots, k\}$.

Approximating the $\varepsilon-$subdifferential

Given $x^{(0)}, \dots, x^{(k)} \in \mathbb{R}^n$, and $s^{(j)} \in \partial f(x^{(j)})$ for $j = 0, \dots, k$, and

$$e_j^{(k)} := f(x^{(k)}) - f(x^{(j)}) - (s^{(j)})^T (x^{(k)} - x^{(j)}),$$

then

$$s^{(j)} \in \partial_{e_j^{(k)}} f(x^{(k)}).$$

This gives an inner approximation of $\partial_\varepsilon f(x^{(k)})$ as:

$$G_\varepsilon^{(k)} := \left\{ \sum_{j=0}^{k} \lambda_j s^{(j)} \,\middle|\, \sum_{j=0}^{k} \lambda_j \, e_j^{(k)} \leq \varepsilon,\ \lambda \in \Delta_k \right\} \subseteq \partial_\varepsilon f(x^{(k)})$$

with $\Delta_k := \{\lambda \in \mathbb{R}^{k+1} \mid \sum_{j=0}^{k} \lambda_j = 1,\ \lambda_j \geq 0,\ j = 0, \dots, k\}$.

Main challenge on manifolds: given $p^{(0)}, \dots, p^{(k)} \in \mathcal{M}$, and $X_{p^{(j)}} \in \partial f(p^{(j)})$, then

$$\mathrm{P}_{p^{(k)} \leftarrow p^{(j)}} X_{p^{(j)}} \in \partial_c f(p^{(k)}) \quad \text{for some} \quad c > 0?$$

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f)$ and

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f)$ and

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

and $\varrho := \max\bigl\{\zeta_{1,\omega}(\delta) - 1,\ 1 - \zeta_{2,\Omega}(\delta)\bigr\}$.

[Alimisis et al. 2021, Appendix C]

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f)$ and

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

and $\varrho := \max\{\zeta_{1,\omega}(\delta) - 1,\ 1 - \zeta_{2,\Omega}(\delta)\}$. [Alimisis et al. 2021, Appendix C]

With

$$e_j^{(k)} := f(p^{(k)}) - f(p^{(j)}) - \left(X_{p^{(j)}}, \log_{p^{(j)}} p^{(k)} \right),$$

$$r_j^{(k)} := \varrho \|\log_{p^{(j)}} p^{(k)}\| \|X_{p^{(j)}}\|,$$

for $X_{p_j} \in \partial f(p^{(j)})$,

Curvature Correction

Let $\delta := \operatorname{diam}(\operatorname{dom} f)$ and

$$\zeta_{1,\omega}(\delta) := \begin{cases} 1 & \text{if } \omega \geq 0, \\ \sqrt{-\omega}\,\delta \coth(\sqrt{-\omega}\,\delta) & \text{if } \omega < 0, \end{cases}$$

$$\zeta_{2,\Omega}(\delta) := \begin{cases} 1 & \text{if } \Omega \leq 0, \\ \sqrt{\Omega}\,\delta \cot(\sqrt{\Omega}\,\delta) & \text{if } \Omega > 0, \end{cases}$$

and $\varrho := \max\bigl\{\zeta_{1,\omega}(\delta) - 1,\ 1 - \zeta_{2,\Omega}(\delta)\bigr\}$. [Alimisis et al. 2021, Appendix C]

With

$$e_j^{(k)} := f(p^{(k)}) - f(p^{(j)}) - \left(X_{p^{(j)}}, \log_{p^{(j)}} p^{(k)}\right),$$

$$r_j^{(k)} := \varrho \|\log_{p^{(j)}} p^{(k)}\| \|X_{p^{(j)}}\|,$$

for $X_{p_j} \in \partial f(p^{(j)})$, we get

$$G_\varepsilon^k := \left\{ \sum_{j=1}^{k} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow p^{(j)}} X_{p^{(j)}} \;\middle|\; \sum_{j=1}^{k} \lambda_j \left(e_j^{(k)} + r_j^{(k)}\right) \leq \varepsilon,\ \lambda \in \Delta_k \right\} \subseteq \partial_\varepsilon f(p^{(k)})$$

[Bergmann, Herzog, and HJ 2024]

NTNU

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{0, \ldots, k\} =: J^{(k)}$ and let λ_j be convex coefficients.

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{0, \ldots, k\} =: J^{(k)}$ and let λ_j be convex coefficients.

$$s^{(j)} \in \partial f(x^{(j)}) \implies \sum_{j=0}^{k} \lambda_j s^{(j)} \in \partial_\varepsilon f(x^{(k)})$$

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{0, \ldots, k\} =: J^{(k)}$ and let λ_j be convex coefficients.

$$s^{(j)} \in \partial f(x^{(j)}) \implies \sum_{j=0}^{k} \lambda_j s^{(j)} \in \partial_\varepsilon f(x^{(k)}) \quad \text{iff} \quad \varepsilon \geq \sum_{j=0}^{k} \lambda_j e_j^{(k)}$$

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{0, \ldots, k\} =: J^{(k)}$ and let λ_j be convex coefficients.

$$s^{(j)} \in \partial f(x^{(j)}) \implies \sum_{j=0}^{k} \lambda_j s^{(j)} \in \partial_\varepsilon f(x^{(k)}) \quad \text{iff} \quad \varepsilon \geq \sum_{j=0}^{k} \lambda_j e_j^{(k)}$$

The *search direction* is

$$d^{(k)} := -\sum_{j=0}^{k} \lambda_j s^{(j)}$$

The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{0, \ldots, k\} =: J^{(k)}$ and let λ_j be convex coefficients.

$$s^{(j)} \in \partial f(x^{(j)}) \implies \sum_{j=0}^{k} \lambda_j s^{(j)} \in \partial_\varepsilon f(x^{(k)}) \quad \text{iff} \quad \varepsilon \geq \sum_{j=0}^{k} \lambda_j e_j^{(k)}$$

The *search direction* is

$$d^{(k)} := -\sum_{j=0}^{k} \lambda_j s^{(j)}$$

where λ_j are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J^{(k)}|}} \quad & \frac{1}{2} \left\| \sum_{j \in J^{(k)}} \lambda_j s^{(j)} \right\|^2 + \sum_{j \in J^{(k)}} \lambda_j e_j^{(k)} \\ \text{s. t.} \quad & \sum_{j \in J^{(k)}} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J^{(k)} \end{aligned}$$

NTNU

The Riemannian Subproblem

Let $k \in \mathbb{N}$ and $j \in \{0, \dots, k\} =: J^{(k)}$ and let λ_j be convex coefficients.

$$X_{q^{(j)}} \in \partial f(q^{(j)}) \implies \sum_{j=0}^{k} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}} \in \partial_\varepsilon f(p^{(k)}) \quad \text{iff} \quad \varepsilon \geq \sum_{j=0}^{k} \lambda_j \left(e_j^{(k)} + r_j^{(k)} \right)$$

The *search direction* is

$$d^{(k)} := -\sum_{j=0}^{k} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}}$$

where λ_j are solution to

$$\operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J^{(k)}|}} \quad \frac{1}{2} \left\| \sum_{j \in J^{(k)}} \lambda_j \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}} \right\|^2 + \sum_{j \in J^{(k)}} \lambda_j \left(e_j^{(k)} + r_j^{(k)} \right)$$

$$\text{s. t.} \quad \sum_{j \in J^{(k)}} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J^{(k)}$$

The Euclidean Convex Bundle Method

Data: $x^{(0)} = y^{(0)} \in \mathbb{R}^n$, $g^{(0)} = s^{(0)} \in \partial f(x^{(0)})$, $m \in (0,1)$,
$\varepsilon^{(0)} = e_0^{(0)} = 0$, $k = 0$, $J^{(k)} = \{0\}$

while *not converged* **do**

Compute a solution $\lambda \in \mathbb{R}^{|J^{(k)}|}$ of the subproblem and set

$$g^{(k)} := \sum_{j \in J^{(k)}} \lambda_j^{(k)} s^{(j)}, \quad \varepsilon^{(k)} := \sum_{j \in J^{(k)}} \lambda_j^{(k)} e_j^{(k)},$$

$$d^{(k)} := -g^{(k)}, \quad \xi^{(k)} := -\|g^{(k)}\|^2 - \varepsilon^{(k)}$$

Set $y^{(k+1)} := x^{(k)} + d^{(k)}$.

If $f(y^{(k+1)}) \leq f(x^{(k)}) + m\xi^{(k)}$, $x^{(k+1)} := y^{(k+1)}$, else $x^{(k+1)} := x^{(k)}$.

Compute $s^{(k+1)} \in \partial f(y^{(k+1)})$, update $J^{(k+1)}$, and $e_j^{(k+1)}$

end

Result: $x^{(k_*)}$, for some $k_* \in \mathbb{N}$.

The Riemannian Convex Bundle Method

Data: $p^{(0)} = q^{(0)} \in \operatorname{int}(\operatorname{dom} f)$, $g^{(0)} = X_1 \in \partial f(p^{(0)})$, $m \in (0,1)$,
$\varepsilon^{(0)} = c_0^{(0)} = 0$, $k = 0$, $J^{(k)} = \{0\}$, $\beta > 0$, $\varrho \in \mathbb{R}$,

while *not converged* **do**

Compute a solution $\lambda \in \mathbb{R}^{|J^{(k)}|}$ of the subproblem and set

$$g^{(k)} := \sum_{j \in J^{(k)}} \lambda_j^{(k)} \mathrm{P}_{p^{(k)} \leftarrow q^{(j)}} X_{q^{(j)}}, \quad \varepsilon^{(k)} := \sum_{j \in J^{(k)}} \lambda_j^{(k)} \left(e_j^{(k)} + r_j^{(k)} \right),$$

$$d^{(k)} := -g^{(k)}, \quad \xi^{(k)} := -\|g^{(k)}\|^2 - \varepsilon^{(k)}$$

Set $q^{(k+1)} := \exp_{p^{(k)}}(t^{(k)} d^{(k)})$ [1]

If $f(q^{(k+1)}) \le f(p^{(k)}) + m\xi^{(k)}$, $p^{(k+1)} := q^{(k+1)}$, else $p^{(k+1)} := p^{(k)}$ [2] .

Compute $X_{q^{(k+1)}} \in \partial f(q^{(k+1)})$, update $J^{(k+1)}$, and $e_j^{(k+1)}$ and $r_j^{(k+1)}$

end

Result: $p^{(k_*)}$, for some $k_* \in \mathbb{N}$.

---

[1] Backtrack $t^{(k)} \leftarrow \beta t^{(k)}$ if $q^{(k+1)} \notin \operatorname{int}(\operatorname{dom} f)$ or equal to $p^{(k)}$

[2] Look for a $\bar{t}^{(k)} \le t^{(k)}$ that satisfies a geometric condition

Convergence

- In the Euclidean case [Geiger and Kanzow 2002, Theorem 6.80] holds.

Theorem

Let the solution set $S = \{x^* \in \mathbb{R}^n \mid f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x^{(k)}\}$ *generated by the bundle method algorithm converges to a minimizer of* f*.*

Convergence

- In the Euclidean case [Geiger and Kanzow 2002, Theorem 6.80] holds.

Theorem
Let the solution set $S = \{x^* \in \mathbb{R}^n \mid f(x^*) = \inf_{x\in\mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x^{(k)}\}$ *generated by the bundle method algorithm converges to a minimizer of* f*.*

- In the Riemannian case, assuming no accumulation point of $p^{(k)}$ is allowed to lie on $\partial \operatorname{dom} f$, we have an analogous result.

[Bergmann, Herzog, and HJ 2024]

Need to backtrack until

$$\left(\mathrm{P}_{p^{(k)}\leftarrow\bar{q}^{(k+1)}} X_{\bar{q}^{(k+1)}}\,,\ \bar{t}^{(k)}\, g^{(k)}\right) < -m\, \bar{t}^{(k)}\, \xi^{(k)} - e^{(k)}_{k+1} - r^{(k)}_{k+1}, \qquad (1)$$

as inspired by [Bagirov et al. 2020]

Implementation

The algorithm is implemented in Julia using `Manopt.jl` ([Bergmann 2022]) and `Manifolds.jl` ([Axen et al. 2023]). A solver call looks like [1]

```
p* = convex_bundle_method(M, f, ∂f, p0; diameter = δ,
     domain = dom f, k_max = Ω, k_min = ω, m = 10^−3)
```

where

- `M` is a Riemannian manifold
- `f` is the objective function
- `∂f` is a subgradient of the objective function
- `p0` is an initial point on the manifold

The default stopping criterion for the algorithm is set to

$$-\xi^{(k)} \leq 10^{-8}.$$

[1] https://manoptjl.org/stable/solvers/convex_bundle_method/

The Proximal Gradient Method

joint work with

Ronny Bergmann, Paula J. John, Lukas Klingbiel, and Max Pfeffer

(preliminary results)

The Problem

Consider the following minimization problem

$$\text{minimize } f(p) = g(p) + h(p), \quad p \in \mathcal{M}, \tag{2}$$

The Problem

Consider the following minimization problem

$$\text{minimize } f(p) = g(p) + h(p), \quad p \in \mathcal{M}, \tag{2}$$

where

- $\mathcal{M}$ is a Riemannian manifold

The Problem

Consider the following minimization problem

$$\text{minimize } f(p) = g(p) + h(p), \quad p \in \mathcal{M}, \tag{2}$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$

The Problem

Consider the following minimization problem

$$\operatorname{minimize} f(p) = g(p) + h(p), \quad p \in \mathcal{M}, \tag{2}$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$
- and

The Problem

Consider the following minimization problem

$$\text{minimize } f(p) = g(p) + h(p), \quad p \in \mathcal{M}, \tag{2}$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$
- and
  - $g\colon \mathcal{M} \to \mathbb{R}$ is continuously differentiable and L_g-smooth
  - $h\colon \mathcal{M} \to \mathbb{R}$ is (locally) Lipschitz continuous
  - f has compact level sets
  - optimal set of Equation (2) is nonempty

The Problem

Consider the following minimization problem

$$\text{minimize } f(p) = g(p) + h(p), \quad p \in \mathcal{M}, \tag{2}$$

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvature $\omega \leq \kappa \leq \Omega$
- and
  - $g : \mathcal{M} \to \mathbb{R}$ is continuously differentiable and L_g-smooth
  - $h : \mathcal{M} \to \mathbb{R}$ is (locally) Lipschitz continuous
  - f has compact level sets
  - optimal set of Equation (2) is nonempty

Goal Solve this problem with the proximal gradient method.

The Proximal Operator

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a real number $\lambda > 0$, the proximal operator of f is

$$\operatorname{prox}_{\lambda f}(x) := \operatorname*{arg\,min}_{y \in \mathbb{R}^n} f(y) + \frac{1}{2\lambda} \operatorname{dist}^2(y, x),$$

The Proximal Operator

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a real number $\lambda > 0$, the proximal operator of f is

$$\operatorname{prox}_{\lambda f}(x) := \operatorname*{arg\,min}_{y \in \mathbb{R}^n} f(y) + \frac{1}{2\lambda} \operatorname{dist}^2(y, x),$$

If f is proper, lower semi-continuous, and convex, then $\operatorname{prox}_{\lambda f}$ is single-valued.

The Proximal Operator

Given a function $f\colon \mathcal{M} \to \mathbb{R}$ and a real number $\lambda > 0$, the proximal operator of f is

$$\operatorname{prox}_{\lambda f}(p) := \operatorname*{arg\,min}_{q\in\mathcal{M}} f(q) + \frac{1}{2\lambda}\operatorname{dist}^2(q, p),$$

If f is proper, lower semi-continuous, and geodesically convex, then $\operatorname{prox}_{\lambda f}$ is single-valued.

Derivation of the Method

Main idea Replace g by a proximal regularization of its linearized function.

Derivation of the Method

Main idea Replace g by a proximal regularization of its linearized function. For $\lambda > 0$ and a point x, compute a new candidate as

$$\operatorname*{arg\,min}_{y \in \mathbb{R}^n} g(x) + (\operatorname{grad} g(x)\,,\ y - x) + \frac{1}{2\lambda}\operatorname{dist}^2(x, y) + h(y).$$

Derivation of the Method

Main idea Replace g by a proximal regularization of its linearized function. For $\lambda > 0$ and a point x, compute a new candidate as

$$\operatorname*{arg\,min}_{y \in \mathbb{R}^n} g(x) + (\operatorname{grad} g(x)\,,\, y - x) + \frac{1}{2\lambda} \operatorname{dist}^2(x, y) + h(y).$$

Taylor expand the distance term to second order and obtain

$$\operatorname*{arg\,min}_{y \in \mathbb{R}^n} g(x) + \frac{1}{2\lambda} \operatorname{dist}^2 (x - \lambda \operatorname{grad} g(x), y) + h(y).$$

Derivation of the Method

Main idea Replace g by a proximal regularization of its linearized function. For $\lambda > 0$ and a point x, compute a new candidate as

$$\operatorname*{arg\,min}_{y\in\mathbb{R}^n} g(x) + (\operatorname{grad} g(x)\,,\; y - x) + \frac{1}{2\lambda}\operatorname{dist}^2(x, y) + h(y).$$

Taylor expand the distance term to second order and obtain

$$\operatorname*{arg\,min}_{y\in\mathbb{R}^n} g(x) + \frac{1}{2\lambda}\operatorname{dist}^2\left(x - \lambda \operatorname{grad} g(x), y\right) + h(y).$$

Since $g(x)$ does not change the minimizer, we get

$$\operatorname{prox}_{\lambda h}(x - \lambda \operatorname{grad} g(x)) = \operatorname*{arg\,min}_{y\in\mathbb{R}^n} \frac{1}{2\lambda}\operatorname{dist}^2(x - \lambda \operatorname{grad} g(x), y) + h(y)$$

Derivation of the Method

Main idea Replace g by a proximal regularization of its linearized function. For $\lambda > 0$ and a point p, compute a new candidate as

$$\operatorname*{arg\,min}_{q \in \mathcal{M}} g(p) + \left(\operatorname{grad} g(p)\,,\ \log_p q \right) + \frac{1}{2\lambda} \operatorname{dist}^2(p, q) + h(q).$$

Taylor expand the distance term to second order and obtain

$$\operatorname*{arg\,min}_{q \in \mathcal{M}} g(p) + \frac{1}{2\lambda} \operatorname{dist}^2 \left(\exp_p(-\lambda \operatorname{grad} g(p)), q \right) + h(q).$$

Since $g(p)$ does not change the minimizer, we get

$$\operatorname{prox}_{\lambda h}(\exp_p(-\lambda \operatorname{grad} g(p))) = \operatorname*{arg\,min}_{q \in \mathcal{M}} \frac{1}{2\lambda} \operatorname{dist}^2(\exp_p(-\lambda \operatorname{grad} g(p)), q) + h(q)$$

Proximal Gradient Method

Data: g, $\operatorname{grad} g$, h, $\operatorname{prox}_{\lambda h}$, a sequence $\lambda^{(k)}$, an initial point $x^{(0)} \in \mathbb{R}^n$.

while *convergence criterion is not fulfilled* **do**

$\quad x^{(k+1)} = \operatorname{prox}_{\lambda^{(k)} h}\left(x^{(k)} - \lambda^{(k)} \operatorname{grad} g(x^{(k)})\right)$

$\quad$ Set $k := k + 1$

end

Result: $x^{(k_*)}$, for some $k_* \in \mathbb{N}$.

Proximal Gradient Method

Data: g, $\operatorname{grad} g$, h, $\operatorname{prox}_{\lambda h}$, a sequence $\lambda^{(k)}$, an initial point $p^{(0)} \in \mathcal{M}$.

while *convergence criterion is not fulfilled* **do**

$$p^{(k+1)} = \operatorname{prox}_{\lambda^{(k)} h}\Bigl(\exp_{p^{(k)}}\bigl(-\lambda^{(k)} \operatorname{grad} g(p^{(k)})\bigr)\Bigr)$$

Set $k := k + 1$

end

Result: $p^{(k_*)}$, for some $k_* \in \mathbb{N}$.

Convergence

In the Euclidean case we need a backtracking strategy given by three parameters $s > 0$, $\gamma \in (0,1)$, $\eta > 1$ such that

$$f(x^{(k)}) - f(\mathrm{prox}_{s\eta^{i_k} h}(x^{(k)})) \geq \gamma s \eta^{i_k} \,\mathrm{dist}^2(x^{(k)}, \mathrm{prox}_{s\eta^{i_k} h}(x^{(k)})), \tag{3}$$

$i_k \geq 0$ is the smallest integer for which the condition is satisfied.

Convergence

In the Euclidean case we need a backtracking strategy given by three parameters $s > 0$, $\gamma \in (0,1)$, $\eta > 1$ such that

$$f(x^{(k)}) - f(\operatorname{prox}_{s\eta^{i_k} h}(x^{(k)})) \geq \gamma s \eta^{i_k} \operatorname{dist}^2(x^{(k)}, \operatorname{prox}_{s\eta^{i_k} h}(x^{(k)})), \tag{3}$$

$i_k \geq 0$ is the smallest integer for which the condition is satisfied.

Theorem

Given either a constant step-size $\lambda \in \left(0, \frac{2}{L_g}\right)$ or one that is chosen by the backtracking procedure Equation (3), then

1. *the sequence $\left(f(p^{(k)})\right)$ is nonincreasing, and $f(x^{(k+1)}) < f(x^{(k)})$ if and only if $x^{(k)}$ is not a stationary point of the original problem;*
2. $\lambda \operatorname{dist}(x^{(k)}, x^{(k+1)}) \to 0$ *as* $k \to \infty$*;*
3. $\min_{n=0,1,\dots,k} \lambda^{(n)} \operatorname{dist}(x^{(n)}, x^{(n+1)}) = O\left(\frac{1}{\sqrt{k+1}}\right)$*;*
4. *all limit points of the sequence $(x^{(k)})$ are stationary points of the original problem.*

[Beck 2017, Theorem 10.15]

Convergence

In the Riemannian case we have similar preliminary results for Item 1 and Item 2.

Convergence

In the Riemannian case we have similar preliminary results for Item 1 and Item 2.

Key differences:

- if $\Omega > 0$, no global law of cosines
  - need to ensure $\lambda \in \left(0, \frac{\zeta_{2,\Omega}(r)}{L_g}\right)$ and $\zeta_{2,\Omega}(r) > 0$, where

$$r \leq \operatorname{dist}(p^{(k)}, q^{(k)}) + \operatorname{dist}(q^{(k)}, p^{(k+1)}),$$

and $q^{(k)} := \exp_{p^{(k)}}\left(-\lambda \operatorname{grad} g(p^{(k)})\right)$;

Convergence

In the Riemannian case we have similar preliminary results for Item 1 and Item 2.

Key differences:

- if $\Omega > 0$, no global law of cosines
  - need to ensure $\lambda \in \left(0, \frac{\zeta_{2,\Omega}(r)}{L_g}\right)$ and $\zeta_{2,\Omega}(r) > 0$, where

    $$r \leq \operatorname{dist}(p^{(k)}, q^{(k)}) + \operatorname{dist}(q^{(k)}, p^{(k+1)}),$$

    and $q^{(k)} := \exp_{p^{(k)}}\left(-\lambda \operatorname{grad} g(p^{(k)})\right)$;
- working with locally Lipschitz functions (not necessarily geodesically convex)

Convergence

In the Riemannian case we have similar preliminary results for Item 1 and Item 2.

Key differences:

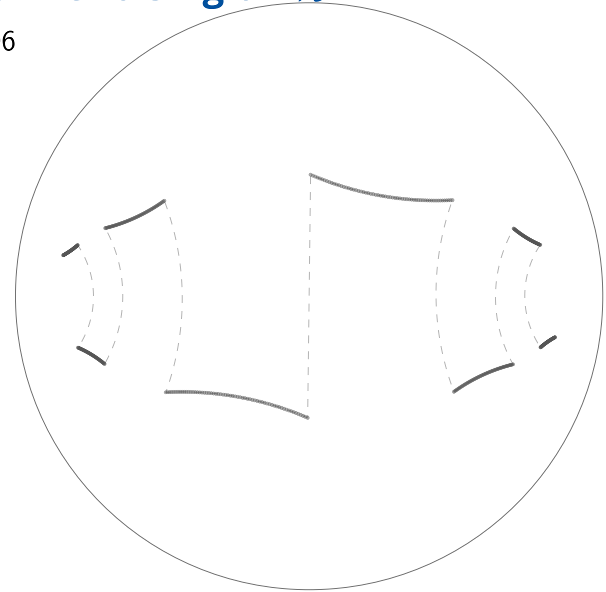
- if $\Omega > 0$, no global law of cosines
  - need to ensure $\lambda \in \left(0, \frac{\zeta_{2,\Omega}(r)}{L_g}\right)$ and $\zeta_{2,\Omega}(r) > 0$, where

$$r \leq \operatorname{dist}(p^{(k)}, q^{(k)}) + \operatorname{dist}(q^{(k)}, p^{(k+1)}),$$

    and $q^{(k)} := \exp_{p^{(k)}}\left(-\lambda \operatorname{grad} g(p^{(k)})\right);$
- working with locally Lipschitz functions (not necessarily geodesically convex)
  - we can achieve ε-statioinarity.

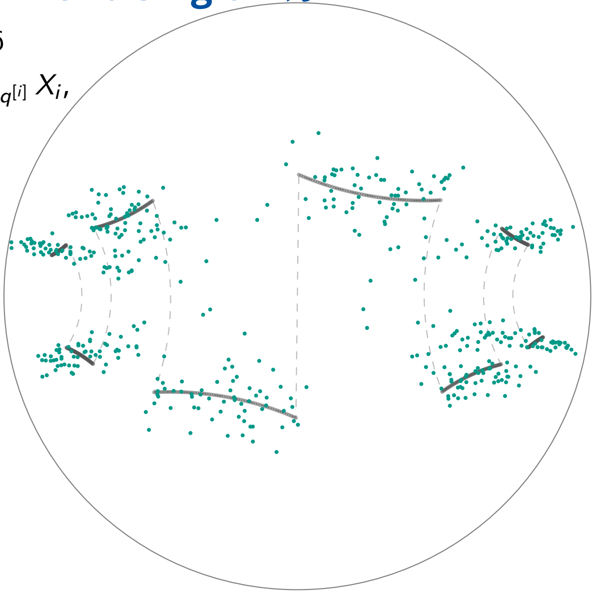
Numerical Example: Signal Denoising on $\mathcal{H}^2$

- signal $q \in \mathcal{M} = (\mathcal{H}^2)^n$, $n = 496$



Numerical Example: Signal Denoising on $\mathcal{H}^2$

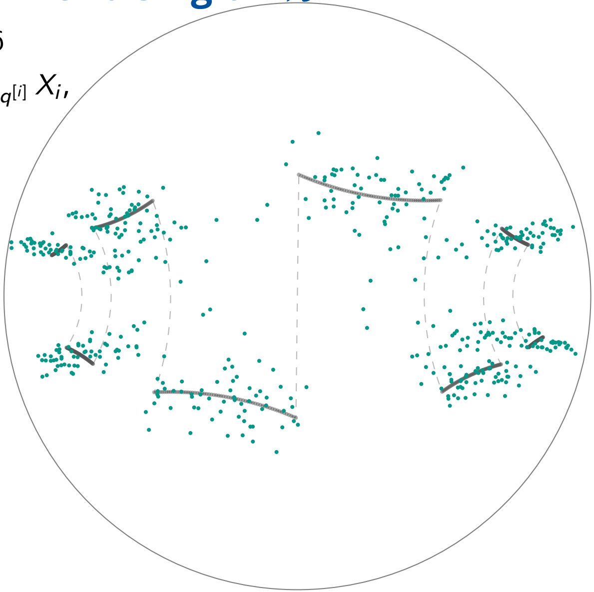
- signal $q \in \mathcal{M} = (\mathcal{H}^2)^n$, $n = 496$
- noisy signal $\bar{q} \in \mathcal{M}$, $\bar{q}^{[i]} = \exp_{q^{[i]}} X_i$, $\sigma = 0.1$



Numerical Example: Signal Denoising on $\mathcal{H}^2$

- signal $q \in \mathcal{M} = (\mathcal{H}^2)^n$, $n = 496$
- noisy signal $\bar{q} \in \mathcal{M}$, $\bar{q}^{[i]} = \exp_{q^{[i]}} X_i$, $\sigma = 0.1$
- ROF Model:

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} \frac{1}{n} d_{\mathcal{M}}(p, \bar{q})^2 + \alpha \sum_{i=1}^{n-1} d_{\mathcal{H}^2}(p^{[i]}, p^{[i+1]})$$

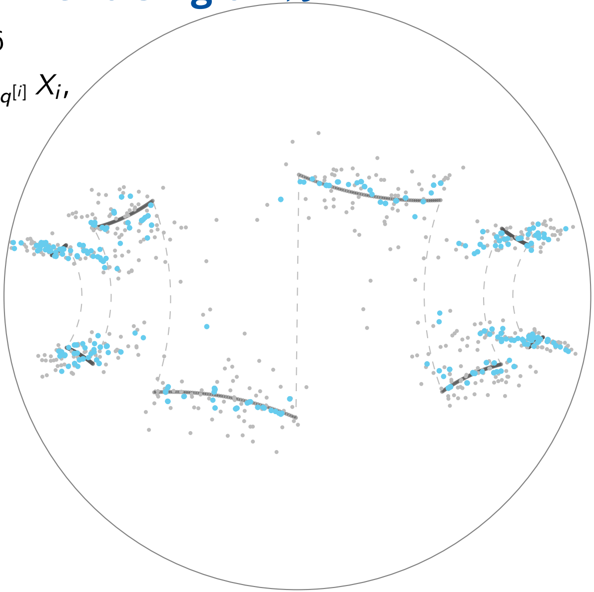


Numerical Example: Signal Denoising on $\mathcal{H}^2$

- signal $q \in \mathcal{M} = (\mathcal{H}^2)^n$, $n = 496$
- noisy signal $\bar{q} \in \mathcal{M}$, $\bar{q}^{[i]} = \exp_{q^{[i]}} X_i$, $\sigma = 0.1$
- ROF Model:

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} \frac{1}{n} d_{\mathcal{M}}(p, \bar{q})^2 + \alpha \sum_{i=1}^{n-1} d_{\mathcal{H}^2}(p^{[i]}, p^{[i+1]})$$

- in RCBM: set $\operatorname{diam}(\operatorname{dom} f) = b > 0$.
- Setting $\alpha = 0.5$ yields RCBM reconstruction p_1^*

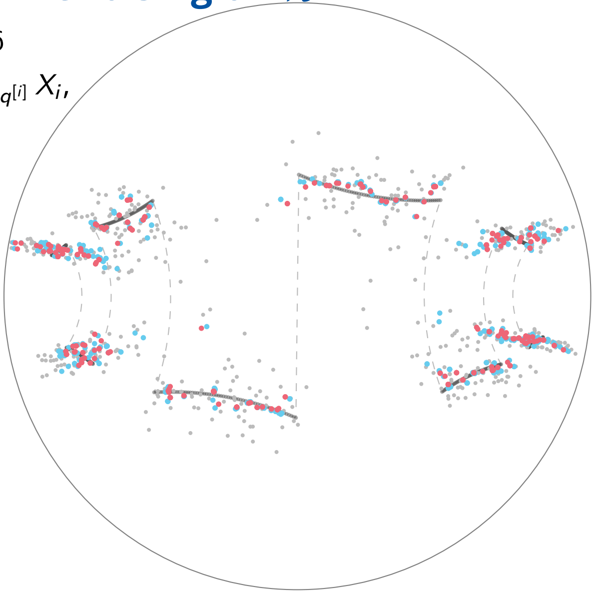


Numerical Example: Signal Denoising on $\mathcal{H}^2$

- signal $q \in \mathcal{M} = (\mathcal{H}^2)^n$, $n = 496$
- noisy signal $\bar{q} \in \mathcal{M}$, $\bar{q}^{[i]} = \exp_{q^{[i]}} X_i$, $\sigma = 0.1$
- ROF Model:

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} \frac{1}{n} d_{\mathcal{M}}(p, \bar{q})^2 + \alpha \sum_{i=1}^{n-1} d_{\mathcal{H}^2}(p^{[i]}, p^{[i+1]})$$

- in RCBM: set $\operatorname{diam}(\operatorname{dom} f) = b > 0$.
- Setting $\alpha = 0.5$ yields RCBM reconstruction p_1^*
- and PGM reconstruction p_2^*



Signal Denoising - Algorithms[2]

- Riemannian Convex Bundle Method (RCBM) [Bergmann, Herzog, and HJ 2024]
- Proximal Bundle Algorithm (PBA) [Hoseini Monjezi, Nobakhtian, and Pouryayevali 2021]
- Subgradient Method (SGM) [Ferreira and Oliveira 1998]
- Cyclic Proximal Point Algorithm (CPPA) [Bačák 2014]
- Proximal Gradient Method (PGM) - work in progress

Algorithm	Iter.	Time (sec.)	Objective	Error
RCBM	5000	13.892	1.4023×10^{-1}	1.3692×10^{-2}
PBA	5000	9.519	1.4289×10^{-1}	1.3032×10^{-2}
SGM	5000	7.897	1.4622×10^{-1}	1.2460×10^{-2}
CPPA	5000	3.739	1.3191×10^{-1}	1.7361×10^{-2}
PGM	5000	3.555	1.3191×10^{-1}	1.7351×10^{-2}

[2] The code for the experiment is available at `juliamanifolds.github.io/ManoptExamples.jl/stable/examples/H2-Signal-TV/`

Numerical Example: Riemannian Median on $\mathcal{S}^d$

- $\mathcal{S}^d$ $d-$dimensional sphere
- $\bar{p}$ north pole
- $q^{(1)}, \ldots, q^{(n)} \in \mathcal{S}^d$ are $n = 1000$ Gaussian random data points in $B_{\frac{\pi}{6}}(\bar{p})$
- $\mathcal{D} = \left\{ q^{(1)}, \ldots, q^{(n)} \,\middle|\, q^{(j)} \in \mathcal{S}^d \text{ for all } j = 1, \ldots, n \right\}$

Numerical Example: Riemannian Median on $\mathcal{S}^d$

- $\mathcal{S}^d$ $d-$dimensional sphere
- $\bar{p}$ north pole
- $q^{(1)}, \dots, q^{(n)} \in \mathcal{S}^d$ are $n = 1000$ Gaussian random data points in $B_{\frac{\pi}{6}}(\bar{p})$
- $\mathcal{D} = \{q^{(1)}, \dots, q^{(n)} \mid q^{(j)} \in \mathcal{S}^d \text{ for all } j = 1, \dots, n\}$
- $f(p) = \begin{cases} \frac{1}{n} \sum_{j=1}^{n} \operatorname{dist}(p, q^{(j)}) & \text{if } p \in B_{\frac{\pi}{6}}(\bar{p}), \\ +\infty & \text{otherwise.} \end{cases}$

Numerical Example: Riemannian Median on $\mathcal{S}^d$

- $\mathcal{S}^d$ $d-$dimensional sphere
- $\bar{p}$ north pole
- $q^{(1)}, \ldots, q^{(n)} \in \mathcal{S}^d$ are $n = 1000$ Gaussian random data points in $B_{\frac{\pi}{6}}(\bar{p})$
- $\mathcal{D} = \{q^{(1)}, \ldots, q^{(n)} \mid q^{(j)} \in \mathcal{S}^d \text{ for all } j = 1, \ldots, n\}$
- $f(p) = \begin{cases} \frac{1}{n} \sum_{j=1}^{n} \operatorname{dist}(p, q^{(j)}) & \text{if } p \in B_{\frac{\pi}{6}}(\bar{p}), \\ +\infty & \text{otherwise.} \end{cases}$

Solve

$$p^* := \arg\min_{p \in \mathcal{S}^d} f(p)$$

Riemannian Median on $\mathcal{S}^d$ - Algorithms[3]

	RCBM			PBA		
Dimension	Iter.	Time (sec.)	Objective	Iter.	Time (sec.)	Objective
2	43	1.58×10^{-2}	0.258 90	71	1.84×10^{-2}	0.258 90
4	74	2.30×10^{-2}	0.253 53	62	1.68×10^{-2}	0.253 53
32	102	4.30×10^{-2}	0.259 89	64	2.73×10^{-2}	0.259 89
1024	103	8.90×10^{-1}	0.266 99	68	6.23×10^{-1}	0.266 99
32 768	80	2.72×10^{1}	0.259 30	65	2.92×10^{1}	0.259 30

	SGM		
Dimension	Iter.	Time (sec.)	Objective
2	401	9.70×10^{-2}	0.258 90
4	5000	1.33	0.253 53
32	231	9.63×10^{-2}	0.259 89
1024	185	1.99	0.266 99
32 768	157	2.05×10^{2}	0.259 30

[3] The code for the experiment is available at `juliamanifolds.github.io/ManoptExamples.jl/stable/examples/RCBM-Median/`

NTNU

Conclusion and Future Work

In summary:

- introduced tools for nonsmooth optimization on Riemannian manifolds
- introduced the Riemannian Convex Bundle Method
- discussed convergence and related challenges
- introduced a preliminary version of a Riemannian Proximal Gradient Method
- showed two numerical examples

To do:

- work on the convergence analysis for the Riemannian Proximal Gradient Method
- acceleration?

Thank you!

Selected References

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

Alimisis, Foivos et al. (2021). “Momentum improves optimization on Riemannian manifolds”. In: Proceedings of Machine Learning Research. PMLR, pp. 1351–1359. URL: https://proceedings.mlr.press/v130/alimisis21a.html.

Axen, S. D. et al. (2023). “Manifolds.jl: An Extensible Julia Framework for Data Analysis on Manifolds”. In: *ACM Transactions on Mathematical Software* 49.4. DOI: 10.1145/3618296. 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.

Numerical Nonsmooth Optimization: State of the Art Algorithms (2020). Springer International Publishing. DOI: 10.1007/978-3-030-34910-3.

Beck, A. (2017). *First-Order Methods in Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9781611974997.

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

Bergmann, Ronny, Roland Herzog, and HJ (2024). “The Riemannian Convex Bundle Method”. URL: https://arxiv.org/abs/2402.13670.

Bonnans, J.-F. et al. (2006). *Numerical optimization: theoretical and practical aspects*. Springer-Verlag.

Ferreira, Orizon and Paulo Roberto Oliveira (1998). “Subgradient algorithm on Riemannian manifolds”. In: *Journal of Optimization Theory and Applications* 97.1, pp. 93–104. DOI: 10.1023/A:1022675100677.

Geiger, C. and C. Kanzow (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. New York: Springer. DOI: 10.1007/978-3-642-56004-0.

Hoseini Monjezi, Najmeh, Soghra Nobakhtian, and Mohamad Reza Pouryayevali (Dec. 2021). “A proximal bundle algorithm for nonsmooth optimization on Riemannian manifolds”. In: *IMA Journal of Numerical Analysis*.