Intrinsic Riemannian Proximal Gradient Methods

Hajg Jasa

joint work with

Ronny Bergmann, Paula John, and Max Pfeffer

NTNU

REMODEL 2026 - Trondheim, Norway

May 12th, 2026

Riemannian Geometry: Notation

- Geodesic
  $\gamma_{pq} : [0,1] \to \mathcal{M}$ 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
  $\log_p q = \exp_p^{-1} q = X_p$
- Sectional curvature
  $\kappa \in [\kappa_\mathrm{l}, \kappa_\mathrm{u}]$

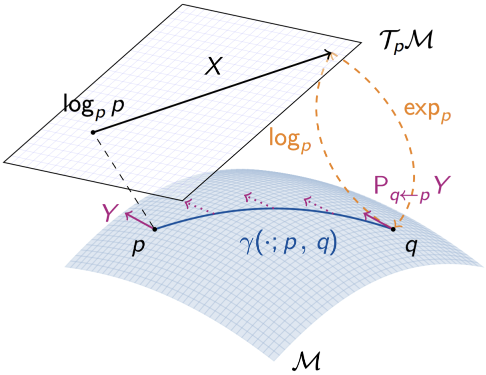

Figure: Courtesy of Ronny Bergmann.

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 (g-convex) if for all $p, q \in \mathcal{C}$ the composition $f \circ \gamma_{pq}(t)$ is convex in the usual sense.

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

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

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.

The Problem

Consider the following minimization problem

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

The Problem

Consider the following minimization problem

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

where

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

The Problem

Consider the following minimization problem

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

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvatures in $[\kappa_l, \kappa_u]$

The Problem

Consider the following minimization problem

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

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvatures in $[\kappa_l, \kappa_u]$
- and

The Problem

Consider the following minimization problem

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

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvatures in $[\kappa_l, \kappa_u]$
- and
  - $g\colon \mathcal{M} \to \overline{\mathbb{R}}$ is L_g-smooth
  - $h\colon \mathcal{M} \to \overline{\mathbb{R}}$ is (locally) Lipschitz continuous
  - f has compact level sets

The Problem

Consider the following minimization problem

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

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvatures in $[\kappa_l, \kappa_u]$
- and
  - $g\colon \mathcal{M} \to \overline{\mathbb{R}}$ is L_g-smooth
  - $h\colon \mathcal{M} \to \overline{\mathbb{R}}$ is (locally) Lipschitz continuous
  - f has compact level sets

Goal. Solve this problem with the proximal gradient method.

The Problem

Consider the following minimization problem

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

where

- $\mathcal{M}$ is a Riemannian manifold
  - complete
  - bounded sectional curvatures in $[\kappa_l, \kappa_u]$
- and
  - $g\colon \mathcal{M} \to \overline{\mathbb{R}}$ is proper, closed, L_g-smooth, and g-convex
  - $h\colon \mathcal{M} \to \overline{\mathbb{R}}$ is proper, closed, and g-convex
  - f has compact level sets

Goal. Solve this problem with the proximal gradient method.

Some Definitions

The objective f is continuous and its level set is compact, hence it is lower bounded by $f_{opt} := \min_{q \in \mathcal{L}_{p^{(0)}}} f(q)$.

Since h and $\operatorname{grad} g$ are continuous we obtain bounds $\alpha_1, \alpha_2, \alpha_g \in \mathbb{R}$ such that

$$\alpha_1 \leq h(q) \leq \alpha_2 \text{ for all } q \in \mathcal{L}_{p^{(0)}}, \text{ and } \|\operatorname{grad} g(q)\| \leq \alpha_g \text{ for all } q \in \mathcal{L}_{p^{(0)}}.$$

Some Definitions

The objective f is continuous and its level set is compact, hence it is lower bounded by $f_{opt} := \min_{q \in \mathcal{L}_{p(0)}} f(q)$.

Since h and $\operatorname{grad} g$ are continuous we obtain bounds $\alpha_1, \alpha_2, \alpha_g \in \mathbb{R}$ such that

$$\alpha_1 \leq h(q) \leq \alpha_2 \text{ for all } q \in \mathcal{L}_{p(0)}, \text{ and } \|\operatorname{grad} g(q)\| \leq \alpha_g \text{ for all } q \in \mathcal{L}_{p(0)}.$$

For real numbers κ_1, κ_2, s, define the quantities

$$\zeta_{1,\kappa_1}(s) := \begin{cases} 1 & \text{if } \kappa_1 \geq 0, \\ \sqrt{-\kappa_1}\, s \coth(\sqrt{-\kappa_1}\, s) & \text{if } \kappa_1 < 0, \end{cases}$$

$$\zeta_{2,\kappa_2}(s) := \begin{cases} 1 & \text{if } \kappa_2 \leq 0, \\ \sqrt{\kappa_2}\, s \cot(\sqrt{\kappa_2}\, s) & \text{if } \kappa_2 > 0, \end{cases}$$

$$\pi_\kappa := \begin{cases} \infty & \text{if } \kappa \leq 0, \\ \frac{\pi}{\sqrt{\kappa}} & \text{if } \kappa > 0, \end{cases}$$

important for taking curvature into account.

Curvature Functions

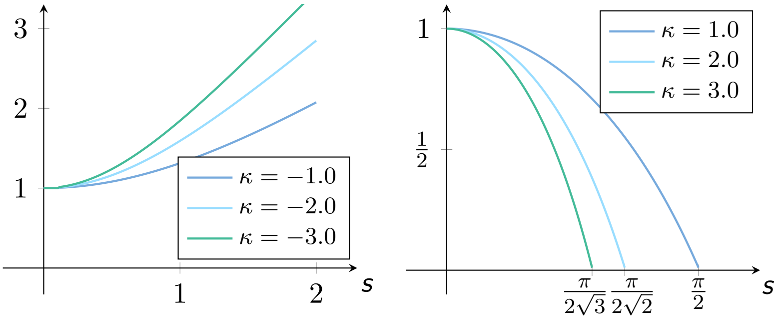

Figure: Illustration of the functions $\zeta_{1,\kappa}$ (left) and $\zeta_{2,\kappa}$ (right).

Some Definitions II

For $p \in \mathcal{M}$ and $\lambda > 0$, the *iteration map* $\mathrm{T}_\lambda \colon \mathcal{M} \to \mathcal{M}$ that to each point p assigns a stationary point $\mathrm{T}_\lambda(p)$ of the function

$$H(\cdot, p, \lambda) \coloneqq h(\cdot) + \frac{1}{2\lambda} \operatorname{dist}^2(\cdot, \exp_p(-\lambda \operatorname{grad} g(p))),$$

and $\frac{1}{\lambda} \log_{\mathrm{T}_\lambda(q)} z(q) \in \partial h(\mathrm{T}_\lambda(q))$, with $z(p) = \exp_p(-\lambda \operatorname{grad} g(p))$.

Some Definitions II

For $p \in \mathcal{M}$ and $\lambda > 0$, the *iteration map* $\mathrm{T}_\lambda \colon \mathcal{M} \to \mathcal{M}$ that to each point p assigns a stationary point $\mathrm{T}_\lambda(p)$ of the function

$$H(\cdot, p, \lambda) := h(\cdot) + \frac{1}{2\lambda} \operatorname{dist}^2(\cdot, \exp_p(-\lambda \operatorname{grad} g(p))),$$

and $\frac{1}{\lambda} \log_{\mathrm{T}_\lambda(q)} z(q) \in \partial h(\mathrm{T}_\lambda(q))$, with $z(p) = \exp_p(-\lambda \operatorname{grad} g(p))$. If h is e.g. geodesically convex,

$$\mathrm{T}_\lambda(p) = \operatorname{prox}_{\lambda h}\Big(\exp_p(-\lambda \operatorname{grad} g(p))\Big) \in \mathcal{M}.$$

NTNU

Some Definitions II

For $p \in \mathcal{M}$ and $\lambda > 0$, the *iteration map* $\mathrm{T}_\lambda \colon \mathcal{M} \to \mathcal{M}$ that to each point p assigns a stationary point $\mathrm{T}_\lambda(p)$ of the function

$$H(\cdot, p, \lambda) \coloneqq h(\cdot) + \frac{1}{2\lambda}\operatorname{dist}^2(\cdot, \exp_p(-\lambda \operatorname{grad} g(p))),$$

and $\frac{1}{\lambda}\log_{\mathrm{T}_\lambda(q)} z(q) \in \partial h(\mathrm{T}_\lambda(q))$, with $z(p) = \exp_p(-\lambda \operatorname{grad} g(p))$. If h is e. g. geodesically convex,

$$\mathrm{T}_\lambda(p) = \operatorname{prox}_{\lambda h}\Bigl(\exp_p(-\lambda \operatorname{grad} g(p))\Bigr) \in \mathcal{M}.$$

The nontrivial task is analyzing stationarity/optimality due to the nonlinearity of $\log_p \mathrm{T}_\lambda(p)$.

NTNU

Some Definitions II

For $p \in \mathcal{M}$ and $\lambda > 0$, the *iteration map* $\mathrm{T}_\lambda \colon \mathcal{M} \to \mathcal{M}$ that to each point p assigns a stationary point $\mathrm{T}_\lambda(p)$ of the function

$$H(\cdot, p, \lambda) \coloneqq h(\cdot) + \frac{1}{2\lambda}\,\mathrm{dist}^2(\cdot, \exp_p(-\lambda \operatorname{grad} g(p))),$$

and $\frac{1}{\lambda}\log_{\mathrm{T}_\lambda(q)} z(q) \in \partial h(\mathrm{T}_\lambda(q))$, with $z(p) = \exp_p(-\lambda \operatorname{grad} g(p))$. If h is e. g. geodesically convex,

$$\mathrm{T}_\lambda(p) = \operatorname{prox}_{\lambda h}\Bigl(\exp_p(-\lambda \operatorname{grad} g(p))\Bigr) \in \mathcal{M}.$$

The nontrivial task is analyzing stationarity/optimality due to the nonlinearity of $\log_p \mathrm{T}_\lambda(p)$.

This is overcome via Riemannian cosine inequalities:

$$\begin{aligned}
2(\log_q y\,,\ \log_q p) &\leq \zeta_{1,\kappa_\mathrm{l}}(s)\,\mathrm{dist}^2(q,y) + \mathrm{dist}^2(p,q) - \mathrm{dist}^2(p,y),\\
2(\log_q y\,,\ \log_q p) &\geq \zeta_{2,\kappa_\mathrm{u}}(s)\,\mathrm{dist}^2(q,y) + \mathrm{dist}^2(p,q) - \mathrm{dist}^2(p,y).
\end{aligned}$$

Technical Lemma

Let $p \in \mathcal{L}_{p(0)}$ and $\delta > 0$,

$$\lambda_\delta := \frac{\sqrt{4\,(\alpha_2 - \alpha_1)^2 + \frac{{\pi_{\kappa_\mathrm{u}}}^2}{(2+\delta)^2}\,{\alpha_g}^2} - 2\,(\alpha_2 - \alpha_1)}{2\,{\alpha_g}^2},$$

and

$$\zeta_\delta := \zeta_{2,\kappa_\mathrm{u}}\left(\frac{\pi_{\kappa_\mathrm{u}}}{2+\delta}\right) = \begin{cases} 1 & \text{if } \kappa_\mathrm{u} \leq 0, \\ \frac{\pi}{(2+\delta)} \cot\left(\frac{\pi}{(2+\delta)}\right) & \text{if } \kappa_\mathrm{u} > 0. \end{cases}$$

Then, for all $\lambda \in (0, \lambda_\delta]$ and $z(p) = \exp_p(-\lambda \operatorname{grad} g(p))$,

$$\operatorname{dist}(p, z(p)) + \operatorname{dist}(\mathrm{T}_\lambda(p), z(p)) \leq \frac{\pi_{\kappa_\mathrm{u}}}{2+\delta},$$

and

$$\zeta_{2,\kappa_\mathrm{u}}(\operatorname{dist}(p, z(p)) + \operatorname{dist}(\mathrm{T}_\lambda(p), z(p))) \geq \zeta_\delta > 0.$$

NCRPG: Nonconvex Case

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

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

end

NCRPG: Nonconvex Case

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

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

  find a stationary point $p^{(k+1)}$ of $H(\cdot, p^{(k)}, \lambda^{(k)})$

end

NCRPG: Nonconvex Case

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

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

  find a stationary point $p^{(k+1)}$ of $H(\cdot, p^{(k)}, \lambda^{(k)})$

  such that $H(p^{(k+1)}, p^{(k)}, \lambda^{(k)}) \leq H(p^{(k)}, p^{(k)}, \lambda^{(k)})$

end

NCRPG: Nonconvex Case

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

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

    find a stationary point $p^{(k+1)}$ of $H(\cdot, p^{(k)}, \lambda^{(k)})$

    such that $H(p^{(k+1)}, p^{(k)}, \lambda^{(k)}) \leq H(p^{(k)}, p^{(k)}, \lambda^{(k)})$

    Set $k := k + 1$

end

NTNU

NCRPG: Nonconvex Case

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

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

  find a stationary point $p^{(k+1)}$ of $H(\cdot, p^{(k)}, \lambda^{(k)})$

  such that $H(p^{(k+1)}, p^{(k)}, \lambda^{(k)}) \leq H(p^{(k)}, p^{(k)}, \lambda^{(k)})$

  Set $k := k + 1$

end

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

NCRPG: Nonconvex Case

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

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

find a stationary point $p^{(k+1)}$ of $H(\cdot, p^{(k)}, \lambda^{(k)})$
such that $H(p^{(k+1)}, p^{(k)}, \lambda^{(k)}) \leq H(p^{(k)}, p^{(k)}, \lambda^{(k)})$

Set $k := k + 1$

end

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

Backtracking strategy

Data: f, $p^{(k)}$, initial guess $s > 0$, $\eta \in (0, 1)$ and $\beta \in (0, \frac{\zeta_\delta}{2})$.

Set $\lambda^{(k)} = s$

NCRPG: Nonconvex Case

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

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

  find a stationary point $p^{(k+1)}$ of $H(\cdot, p^{(k)}, \lambda^{(k)})$

  such that $H(p^{(k+1)}, p^{(k)}, \lambda^{(k)}) \leq H(p^{(k)}, p^{(k)}, \lambda^{(k)})$

  Set $k := k + 1$

end

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

Backtracking strategy

Data: f, $p^{(k)}$, initial guess $s > 0$, $\eta \in (0, 1)$ and $\beta \in (0, \frac{\zeta_\delta}{2})$.

Set $\lambda^{(k)} = s$

while $f(p^{(k)}) - f\left(\mathrm{T}_{\lambda^{(k)}}\left(p^{(k)}\right)\right) < \frac{\beta}{\lambda^{(k)}} \operatorname{dist}^2(p^{(k)}, \mathrm{T}_{\lambda^{(k)}}\left(p^{(k)}\right))$ **do**

  Set $\lambda^{(k)} = \eta\lambda^{(k)}$

end

Convergence: Nonconvex Case

Given either a constant step-size $\lambda^{(k)} := \lambda \in \left(0, \min\left\{\lambda_\delta, \frac{\zeta_\delta}{L_g}\right\}\right)$ or one that is chosen by the backtracking strategy, then

Convergence: Nonconvex Case

Given either a constant step-size $\lambda^{(k)} := \lambda \in \left(0, \min\left\{\lambda_\delta, \frac{\zeta_\delta}{L_g}\right\}\right)$ or one that is chosen by the backtracking strategy, then

1. the sequence $\left(f(p^{(k)})\right)$ is nonincreasing, and $f(p^{(k+1)}) < f(p^{(k)})$ if and only if $p^{(k)}$ is not a stationary point of the original problem;
2. $\lambda^{(k)} \operatorname{dist}(p^{(k)}, p^{(k+1)}) \to 0$ as $k \to \infty$;
3. $\min_{n=0,1,\dots,k} \lambda^{(n)} \operatorname{dist}(p^{(n)}, p^{(n+1)}) = O\left(\frac{1}{\sqrt{k+1}}\right)$;
4. all limit points of the sequence $(p^{(k)})$ are stationary points of f;
5. the algorithm returns an ε-stationary point in $\mathcal{O}\left(\frac{\omega}{\varepsilon^2}\right)$, where ω includes terms that depend on the curvature bounds of $\mathcal{M}$.

[Bergmann et al. 2025b, Lemma 4.7, Theorem 4.9, Theorem 4.10]

CRPG: Convex Case

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**

end

CRPG: Convex Case

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**

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

end

CRPG: Convex Case

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**

$$p^{(k+1)} = \operatorname{prox}_{\lambda^{(k)} h}\Big(\exp_{p^{(k)}}\big(-\lambda^{(k)} \operatorname{grad} g(p^{(k)})\big)\Big)$$

Set $k := k + 1$

end

CRPG: Convex Case

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**

$$p^{(k+1)} = \operatorname{prox}_{\lambda^{(k)} h}\Big(\exp_{p^{(k)}}\big(-\lambda^{(k)} \operatorname{grad} g(p^{(k)})\big)\Big)$$

Set $k := k+1$

end

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

CRPG: Convex Case

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**

$$p^{(k+1)} = \operatorname{prox}_{\lambda^{(k)} h}\Big(\exp_{p^{(k)}}\big(-\lambda^{(k)} \operatorname{grad} g(p^{(k)})\big)\Big)$$

Set $k := k + 1$

end

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

Backtracking strategy

Data: f, $p^{(k)}$, initial guess $s \in (0, \lambda_\delta)$, $\eta \in (0, 1)$ and $\theta \geq 1$.

Set $\lambda^{(-1)} = s$, and $\lambda^{(k)} = \min\{s, \theta\lambda^{(k-1)}\}$

CRPG: Convex Case

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**

$p^{(k+1)} = \operatorname{prox}_{\lambda^{(k)} h}\Big(\exp_{p^{(k)}}\big(-\lambda^{(k)} \operatorname{grad} g(p^{(k)})\big)\Big)$

Set $k := k + 1$

end

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

Backtracking strategy

Data: f, $p^{(k)}$, initial guess $s \in (0, \lambda_\delta)$, $\eta \in (0, 1)$ and $\theta \geq 1$.

Set $\lambda^{(-1)} = s$, and $\lambda^{(k)} = \min\{s, \theta\lambda^{(k-1)}\}$

while $g\left(\mathrm{T}_{\lambda^{(k)}}(p^{(k)})\right) >$
$g(p^{(k)}) + \left(\operatorname{grad} g(p^{(k)}), \mathrm{T}_{\lambda^{(k)}}(p^{(k)})\right) + \frac{\zeta_\delta}{\lambda^{(k)}} \operatorname{dist}^2(p^{(k)}, \mathrm{T}_{\lambda^{(k)}}(p^{(k)}))$ **do**

Set $\lambda^{(k)} = \eta\lambda^{(k)}$

end

Convergence: Convex Case

Given either a constant step-size $\lambda^{(k)} := \lambda \in \left(0, \min\{\lambda_\delta, \frac{\zeta_\delta}{L_g}\}\right)$ or one that is chosen by the backtracking strategy, then, CRPG returns an ε-stationary point in $\mathcal{O}\left(\frac{\Omega}{\varepsilon}\right)$, where Ω includes terms that depend on the curvature bounds of $\mathcal{M}$,

[Bergmann et al. 2025a, Theorem 4.6, Theorem 4.9]

Convergence: Convex Case

Given either a constant step-size $\lambda^{(k)} := \lambda \in \left(0, \min\{\lambda_\delta, \frac{\zeta_\delta}{L_g}\}\right)$ or one that is chosen by the backtracking strategy, then, CRPG returns an ε-stationary point in $\mathcal{O}\left(\frac{\Omega}{\varepsilon}\right)$, where Ω includes terms that depend on the curvature bounds of $\mathcal{M}$,

[Bergmann et al. 2025a, Theorem 4.6, Theorem 4.9]

Convergence: Convex Case

Given either a constant step-size $\lambda^{(k)} := \lambda \in \left(0, \min\{\lambda_\delta, \frac{\zeta_\delta}{L_g}\}\right)$ or one that is chosen by the backtracking strategy, then, CRPG returns an ε-stationary point in $\mathcal{O}\left(\frac{\Omega}{\varepsilon}\right)$, where Ω includes terms that depend on the curvature bounds of $\mathcal{M}$,

- if at iteration $k-1$, it holds that $\frac{\lambda^{(k-1)}\Delta^{(k-1)}}{\zeta_{1,\kappa_1}(D^{(k-1)})\,\mathrm{dist}^2(p^{(k-1)},p^*)} \geq 1$, then

$$\Delta^{(k)} \leq \frac{1}{2}\Delta^{(k-1)};$$

[Bergmann et al. 2025a, Theorem 4.6, Theorem 4.9]

Convergence: Convex Case

Given either a constant step-size $\lambda^{(k)} := \lambda \in \left(0, \min\{\lambda_\delta, \frac{\zeta_\delta}{L_g}\}\right)$ or one that is chosen by the backtracking strategy, then, CRPG returns an ε-stationary point in $\mathcal{O}\left(\frac{\Omega}{\varepsilon}\right)$, where Ω includes terms that depend on the curvature bounds of $\mathcal{M}$,

- if at iteration $k-1$, it holds that $\frac{\lambda^{(k-1)}\Delta^{(k-1)}}{\zeta_{1,\kappa_1}(D^{(k-1)})\,\mathrm{dist}^2(p^{(k-1)},p^*)} \geq 1$, then

$$\Delta^{(k)} \leq \frac{1}{2}\Delta^{(k-1)};$$

- otherwise,

$$\Delta^{(k)} \leq \frac{\Delta^{(0)}}{1 + k\,\tau\Delta^{(0)}},$$

where τ depends on the step-size, $\mathcal{L}_{p^{(0)}}$, and $\Delta^{(0)}$.

[Bergmann et al. 2025a, Theorem 4.6, Theorem 4.9]

Convergence: Convex Case

Given either a constant step-size $\lambda^{(k)} := \lambda \in \left(0, \min\{\lambda_\delta, \frac{\zeta_\delta}{L_g}\}\right)$ or one that is chosen by the backtracking strategy, then, CRPG returns an ε-stationary point in $\mathcal{O}\left(\frac{\Omega}{\varepsilon}\right)$, where Ω includes terms that depend on the curvature bounds of $\mathcal{M}$,

- if at iteration $k-1$, it holds that $\frac{\lambda^{(k-1)}\Delta^{(k-1)}}{\zeta_{1,\kappa_1}(D^{(k-1)})\,\mathrm{dist}^2(p^{(k-1)},p^*)} \geq 1$, then

$$\Delta^{(k)} \leq \frac{1}{2}\Delta^{(k-1)};$$

- otherwise,

$$\Delta^{(k)} \leq \frac{\Delta^{(0)}}{1 + k\,\tau\Delta^{(0)}},$$

  where τ depends on the step-size, $\mathcal{L}_{p^{(0)}}$, and $\Delta^{(0)}$.

- if f is μ_f-strongly convex, CRPG convergences linearly to the unique minimizer of f.

[Bergmann et al. 2025a, Theorem 4.6, Theorem 4.9]

NTNU

Numerical Example: Sparse PCA

For $n, r > 0$, we set up the problem on the *oblique manifold* $\mathrm{OB}(n, r) = \bigotimes_{i=1}^{r} \mathbb{S}^{n-1}$ of matrices with normalized columns as

$$\operatorname*{arg\,min}_{X \in \mathrm{OB}(n,r)} f(X),$$

Numerical Example: Sparse PCA

For $n, r > 0$, we set up the problem on the *oblique manifold* $\mathrm{OB}(n, r) = \bigotimes_{i=1}^{r} \mathbb{S}^{n-1}$ of matrices with normalized columns as

$$\operatorname*{arg\,min}_{X \in \mathrm{OB}(n,r)} f(X),$$

with

$$f(X) := \frac{1}{2}\|X^\top A^\top A X - D^2\|^2 + \mu\|X\|_1,$$

where $A \in \mathbb{R}^{m \times n}$ is a given data matrix, D contains the r leading singular values of A on its diagonal.

Numerical Example: Sparse PCA

NTNU

For $n, r > 0$, we set up the problem on the *oblique manifold* $\mathrm{OB}(n, r) = \bigotimes_{i=1}^{r} \mathbb{S}^{n-1}$ of matrices with normalized columns as

$$\operatorname*{arg\,min}_{X \in \mathrm{OB}(n,r)} f(X),$$

with

$$f(X) := \frac{1}{2}\|X^\top A^\top A X - D^2\|^2 + \mu\|X\|_1,$$

where $A \in \mathbb{R}^{m \times n}$ is a given data matrix, D contains the r leading singular values of A on its diagonal.

This allows for an efficient fixed-point iteration to solve the proximal subproblem.

Sparse PCA - Comparisons[1]

- Nonconvex Riemannian Proximal Gradient (NCRPG) [Bergmann et al. 2025b]
- Riemannian Proximal Gradient (RPG) [Huang and Wei 2021]
- Manifold Proximal Gradient (ManPG) [Chen et al. 2020]

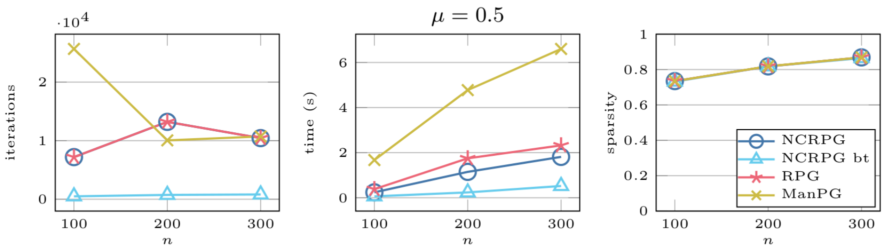

Figure: Results for Sparse PCA on $\mathrm{OB}(n, r)$ with $r = 5$, averaged over 10 runs.

[1] The code for the experiment is available at https://juliamanifolds.github.io/ManoptExamples.jl/stable/examples/NCRPG-Sparse-PCA/

Conclusion and Future Work

In summary:

- introduced an intrinsic Riemannian Proximal Gradient Method
- discussed convergence rates

To do:

- finish convex adaptation to manifolds with $\kappa_l \geq 0$

Conclusion and Future Work

In summary:

- introduced an intrinsic Riemannian Proximal Gradient Method
- discussed convergence rates

To do:

- finish convex adaptation to manifolds with $\kappa_l \geq 0$
- acceleration?

Thank you for your attention!

NTNU

Selected References

- Bergmann, Ronny, HJ, Paula John, and Max Pfeffer (2025a). *The Intrinsic Riemannian Proximal Gradient Method for Convex Optimization*. arXiv: 2507.16055.
- — (2025b). *The Intrinsic Riemannian Proximal Gradient Method for Nonconvex Optimization*. arXiv: 2506.09775.
- Chen, Shixiang, Shiqian Ma, Anthony Man-Cho So, and Tong Zhang (2020). "Proximal Gradient Method for Nonsmooth Optimization over the Stiefel Manifold". In: *SIAM Journal on Optimization* 30.1, pp. 210–239. DOI: 10.1137/18m122457x.
- Huang, Wen and Ke Wei (2021). "Riemannian proximal gradient methods". In: *Mathematical Programming* 194.1–2, pp. 371–413. DOI: 10.1007/s10107-021-01632-3.