The Intrinsic Riemannian Proximal Gradient Method for Convex Optimization

Ronny Bergmann*, Hajg Jasa*, Paula John‡, Max Pfeffer‡
hajg.jasa@ntnu.no
*Department of Mathematical Sciences,
NTNU, Trondheim, Norway
‡Institute for Numerical and Applied Mathematics,
Georg-August-Universität Göttingen, Göttingen, Germany

Motivation

Optimization problems of the form

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

arise, e. g. when solving

- convex constrained smooth optimization
- regularized minimization

in Euclidean space. We generalize this splitting with

- $\mathcal{M}$ a Riemannian manifold with curvature in $[\kappa_{\mathrm{l}}, \kappa_{\mathrm{u}}]$
- g an L_g-smooth and g-convex function
- h a g-convex function

Algorithm: CRPG (available in Manopt.jl[1])

Require: g, $\operatorname{grad} g$, h, $\operatorname{prox}_{\lambda h}$, λ_k, $p_0 \in \mathcal{M}$.

1: **while** not converged **do**
2: $\quad p_{k+1} = \operatorname{prox}_{\lambda_k h}\big(\exp_{p_k}(-\lambda_k \operatorname{grad} g(p_k))\big)$
3: $\quad$ Set $k \leftarrow k+1$
4: **end while**

with

$$\operatorname{prox}_{\lambda h}(p) = \operatorname*{arg\,min}_{q\in\mathcal{M}} \left\{ h(q) + \frac{1}{2\lambda} d^2(p,q) \right\}$$

and d the Riemannian distance

Convergence Rates

On Hadamard manifolds ($\kappa_{\mathrm{u}} \leq 0$) with f

- **g-convex:** sublinear rate $\mathcal{O}\left(\frac{\omega}{k}\right)$
- **strongly g-convex:** linear rate $\mathcal{O}\left(\Omega^k\right)$

where ω and Ω include dependence on *curvature*, *smoothness*, and *geodesic convexity*

[1] manoptjl.org/stable/solvers/proximal_gradient_method/

Optimize Geodesically Convex Problems with an Intrinsic Riemannian Proximal Gradient Method

Prox-Grad Inequality (PGI)

- $z(q) = \exp_q(-\lambda \operatorname{grad} g(q))$
- $\mathrm{T}_\lambda(q) = \operatorname{prox}_{\lambda h}(z(q))$
- $l_g(p,q) = g(p) - g(q) - \big(\operatorname{grad} g(q), \log_q p\big)$

$$\begin{aligned} 4\lambda[f(p) - f(\mathrm{T}_\lambda(q))] \geq{}& 4\lambda l_g(p,q) + \mathbf{A}\, d^2(p, \mathrm{T}_\lambda(q)) \\ &- \mathbf{B}\, d^2(p,q) + \mathbf{C}\, d^2(q, z(q)) \\ &+ \mathbf{D}\, d^2(q, \mathrm{T}_\lambda(q)) + \mathbf{E}\, d^2(z(q), \mathrm{T}_\lambda(q)) \end{aligned}$$

where

$\mathbf{A} = 1 + \zeta_{2,\kappa_{\mathrm{u}}}(D_3)$
$\mathbf{B} = \zeta_{1,\kappa_{\mathrm{l}}}(D_2) + 1$ $\qquad \mathbf{D} = \zeta_{2,\kappa_{\mathrm{u}}}(D_1) - 1$
$\mathbf{C} = \zeta_{2,\kappa_{\mathrm{u}}}(D_1) - \zeta_{1,\kappa_{\mathrm{l}}}(D_2)$ $\qquad \mathbf{E} = \zeta_{2,\kappa_{\mathrm{u}}}(D_3) - 1$

with $D_1 = D(q, z(q), \mathrm{T}_\lambda(q))$, $D_2 = D(q, z(q), p)$, and $D_3 = D(\mathrm{T}_\lambda(q), z(q), p)$, where $D(p,q,r)$ is the diameter of the uniquely geodesic triangle with vertices $p, q, r \in \mathcal{M}$, and

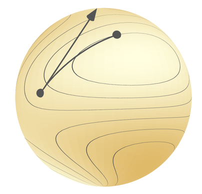
$$\zeta_{1,\kappa_1}(s) = \begin{cases} 1 & \text{if } \kappa_1 \geq 0, \\ \sqrt{-\kappa_1}\, s \coth(\sqrt{-\kappa_1}\, s) & \text{if } \kappa_1 < 0, \end{cases}$$

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

- if $g = 0$, PGI $\Longrightarrow$ proximal point convergence rate
- if $h = 0$, PGI $\Longrightarrow$ gradient descent convergence rate

(!) otherwise, PGI $\not\Longrightarrow$ CRPG convergence rate

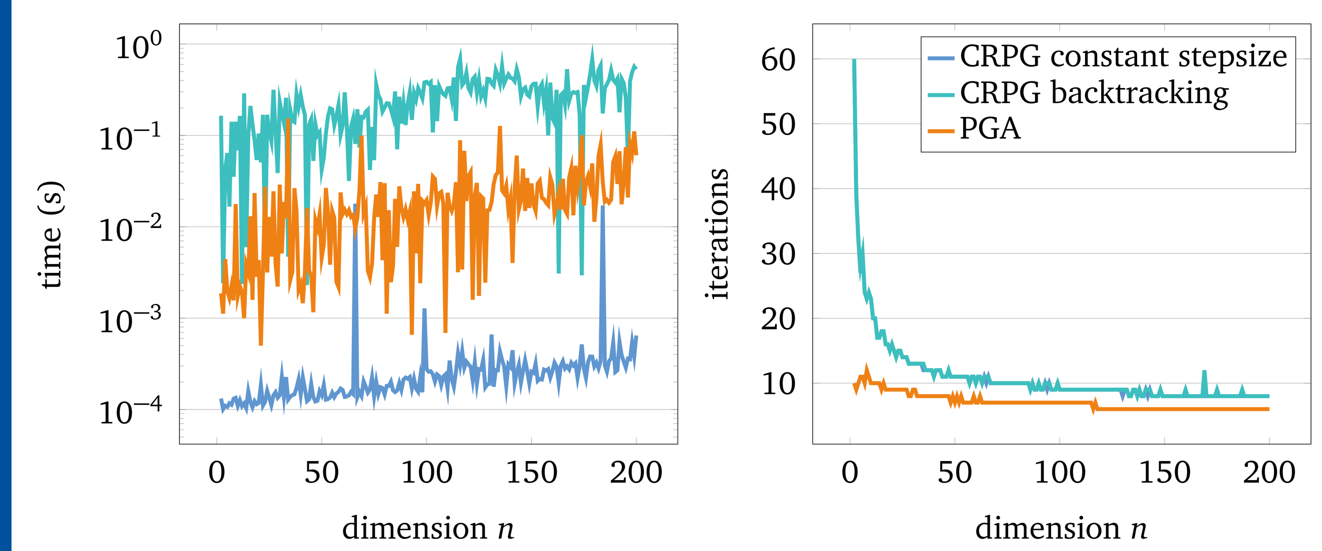
Example



Constrained Mean on $\mathcal{H}^n$ On the hyperbolic space $\mathcal{M} = \mathcal{H}^n$ with the Minkowski metric, let $\{q_1, \ldots, q_N\} \subset \mathcal{M}$ be $N = 400$ randomly generated data points, $\mathcal{B}(q_0, r)$ a geodesic ball of radius r around a point $q_0 \in \mathcal{M}$ and

$$g(p) = \frac{1}{2N} \sum_{j=1}^{N} d^2(p, q_j), \quad h(p) = \chi_{\mathcal{B}(q_0,r)}(p),$$

where $\chi_{\mathcal{B}(q_0,r)}(p)$ is the characteristic function of $\mathcal{B}(q_0, r)$. Then $\operatorname{prox}_{\lambda h}$ is the projection onto $\mathcal{B}(q_0, r)$.



CRPG and the Projected Gradient Algorithm (PGA), varying dimension $n \in \{2, 3, \ldots, 200\}$.

NTNU | Norwegian University of Science and Technology