Berlin Mathematics Research Center MATH+

Accelerated Riemannian Optimization: Handling Constraints to Bound Geometric Penalties

David Martínez-Rubio (TUB, ZIB), Sebastian Pokutta (TUB, ZIB)

Problem

Design accelerated first-order methods for smooth and (strongly or not) geodesically-convex problems.

Taking into account that:

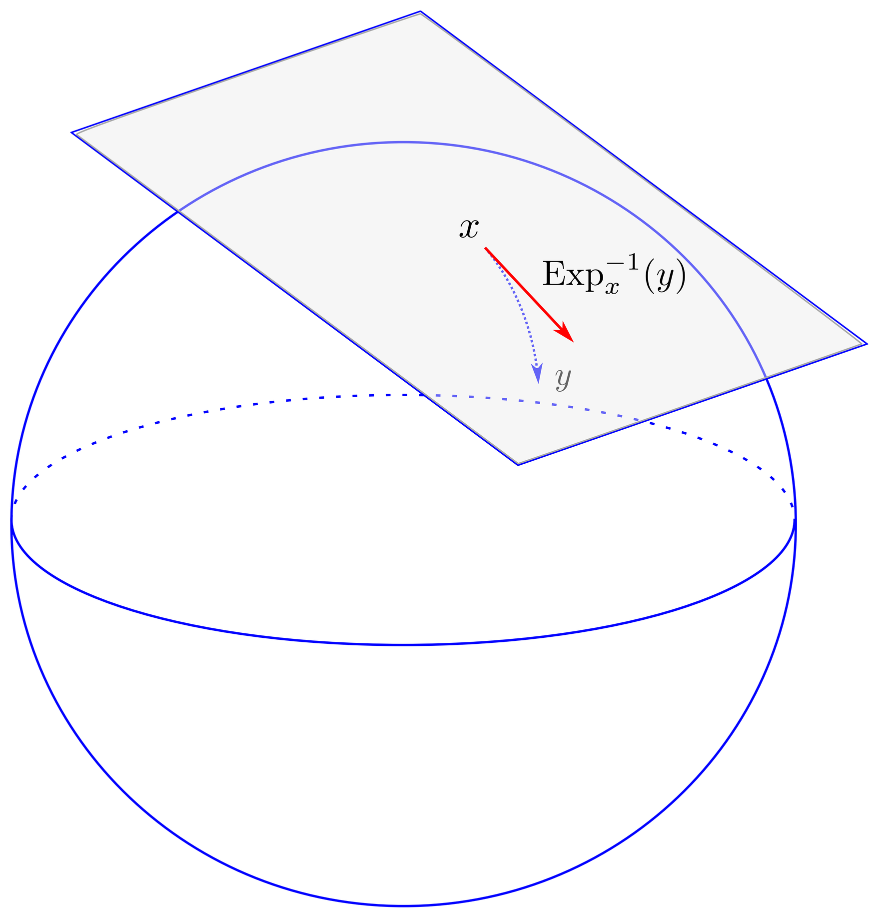
- All accelerated methods **require the iterates to stay in some pre-specified set** to bound errors caused by the interplay of estimations and the geometry and its curvature: **geometric penalties**.
- Most previous works **just assume** that the iterates of their algorithms are going to stay in this pre-specified set **without any mechanism for enforcing this condition**.
- Two works do not make this assumption, but they **only work in limited settings**: in a small neighborhood of the minimizer or in manifolds of constant curvature, respectively.

Riemannian Optimization

This kind of optimization concerns the following problem:

minimize $f(x)$ subject to $x \in \mathcal{M}$, for a Riemannian manifold $\mathcal{M}$.

- It turns **constrained** problems into **unconstrained** ones by working inside of the manifold and exploiting its structure.
- A function can be **Euclidean non-convex but g-convex** on a manifold with the right metric.
- The g-convex case is a useful tool to understand the **non g-convex case**, similarly to what happens for Euclidean problems.
- **Some applications of Riemannian optimization:**
  - **Hadamard and g-convex**: Gaussian mixture models, robust covariance estimation in Gaussians, operator scaling, Wasserstein Barycenters, Karcher means, computing Brascamp-Lieb constants.
  - **Others**: PCA, low-rank matrix completion, dictionary learning, optimization under orthogonality constraints (with applications to RNNs [LM19]).



- When optimizing in Riemannian manifolds, we make use of the **tangent space** $T_x\mathcal{M}$ of points $x \in \mathcal{M}$, depicted in the figure.
- Given two points $x, y \in \mathcal{M}$, the **inverse exponential map** $\mathrm{Exp}_x^{-1}(y)$ returns a vector in $T_x\mathcal{M}$ such that the geodesic segment starting from x with that vector's direction and length ends at y.
- We work on **geodesically convex and uniquely geodesic sets**, for which this vector is well defined and unique.

(note that we work on Hadamard manifolds in this work, which does not include the sphere)

Our Setting

For a Riemannian manifold of bounded sectional curvature in $[\kappa_{\min}, \kappa_{\max}]$, we define:

$$\zeta \stackrel{\text{def}}{=} R\sqrt{|\kappa_{\min}|}\coth(R\sqrt{|\kappa_{\min}|}) \text{ if } \kappa_{\min} \leq 0 \text{ and } \stackrel{\text{def}}{=} 1 \text{ otherwise.}$$

It is $\zeta \in [R\sqrt{|\kappa_{\min}|}, R\sqrt{|\kappa_{\min}|} + 1]$.

We work with a wide class of Hadamard manifolds $\mathcal{H}$, thus $\kappa_{\min} \leq \kappa_{\max} \leq 0$.

We have a differentiable function f with a global minimizer at x^*. Let $x_0 \in \mathcal{H}$ be an initial point and $R > d(x_0, x^*)$ a bound. For any two points x, y in $\bar{B}(x_0, 3R)$, we have smoothness and (possibly μ-strongly) g-convexity:

$$f(y) \leq f(x) + \langle \nabla f(x), \mathrm{Exp}_x^{-1}(y)\rangle + \frac{L}{2}d(x,y)^2,$$
$$f(y) \geq f(x) + \langle \nabla f(x), \mathrm{Exp}_x^{-1}(y)\rangle + \frac{\mu}{2}d(x,y)^2,$$

where $d(x, y)$ is the Riemannian distance. A function is geodesically convex, if it is 0-strongly geodesically convex, that is, the function is convex when restricted to every geodesic segment in the set.

Goal: Accelerated optimization of f with first-order methods under these assumptions.

- We **ensure** iterates stay in $\bar{B}(x_0, 3R)$.
- We develop an accelerated Riemannian **inexact proximal point method**.
- We instantiate the method and boost convergence implementing **ball optimization oracles**.

Comparison with Previous Works

Legend

- **K?**: sectional curvature values?
- **G?**: is the algorithm global? L and L′ mean they are local algorithms. They require initial distance $O((L/\mu)^{-3/4})$ and $O((L/\mu)^{-1/2})$, respectively.
- **F?**: Full acceleration? That is, dependence on L, μ, and ε like AGD, up to log factors.
- **C?**: can some constraints be enforced? All methods require their iterates to be in some pre-specified compact set. '✗' means: iterates will stay inside of the set by assumption only.
- $W \stackrel{\text{def}}{=} \sqrt{L/\mu}\log(LR^2/\varepsilon)$.

Method	g-convex	μ-st. g-convex	K?	G?	F?	C?
[Nes05, AGD]	$O(\sqrt{\frac{LR^2}{\varepsilon}})$	$O(W)$	0	✓	✓	✓
[ZS18]	-	$O(W)$	bounded	L	✓	✗
[AS20]	-	$\widetilde{O}(\frac{L}{\mu} + W)$	bounded	✓	✗	✗
[Mar22]	$\widetilde{O}(\zeta^2\sqrt{\zeta + \frac{LR^2}{\varepsilon}})$	$\widetilde{O}(\zeta^2 \cdot W)$	ctant.$\neq 0$	✓	✓	✓
[CB21]	-	$O(W)$	bounded*	L′	✓	✓
[KY22]	$O(\zeta\sqrt{\frac{LR^2}{\varepsilon}})$	$O(\zeta \cdot W)$	bounded	✓	✓	✗
This work	$\widetilde{O}(\zeta^2\sqrt{\zeta + \frac{LR^2}{\varepsilon}})$	$\widetilde{O}(\zeta^2 \cdot W)$	Hadamard*	✓	✓	✓
This work**	$\widetilde{O}(\zeta\sqrt{\zeta + \frac{LR^2}{\varepsilon}})$	$\widetilde{O}(\zeta \cdot W)$	Hadamard*	✓	✓	✓

* Covariant derivative of the metric tensor is assumed to be 0 (bounded also works). It is 0 for all applications known to us.
** With access to a convex projection oracle (see below).

Accelerated Riemannian Inexact Proximal Point Method

- We design a **generic framework for accelerated optimization** that assumes **access to a linearly convergent subroutine** to approx. solve the constrained prox $\min_{x \in \mathcal{X}}\{f(x) + \frac{1}{\lambda}d(x_t, x)^2\}$ for some geodesically convex subset $\mathcal{X}$ of diameter D. For the right λ, the condition number of this problem **only depends on the geometry** and is $O(\zeta_D)$.
- We can use the g-convexity of the Moreau envelope to construct an **inexact version of an implicit subgradient descent step**, the exact one would be $y_k = \mathrm{Exp}_{x_k}(-\lambda\Gamma_{y_k}^{x_k}v_k)$, where $v_k \in \partial(f + I_{\mathcal{X}})(y_k)$ and $\Gamma_{y_k}^{x_k}v_k$ represents parallel transport to $T_{x_k}\mathcal{H}$.
- We design an accelerated algorithm using this inexact implicit descent **in combination with a mirror descent algorithm** whose *simple* (quadratic) regularized lower bound lives in $T_{x_k}\mathcal{H}$, is updated there and then it is "moved" to $T_{x_{k+1}}\mathcal{H}$ using a technique from [KY22] (i.e., another quadratic regularized lower bound is found in $T_{x_{k+1}}\mathcal{H}$).
- We minimize in $\widetilde{O}(\zeta_D\sqrt{LR^2/\varepsilon})$ iterations, each requiring the prox subroutine.

Even if the prox is computed exactly, there were no accelerated Riemannian proximal point methods before this work.

Inexact Ball Optimization Oracle Implementation and Convergence Boost

- For balls of diameter $D \stackrel{\text{def}}{=} \Theta(1/(R|\kappa_{\min}|))$ and center x_k, we can pull back the prox function to $T_{x_k}\mathcal{H}$ and the resulting Euclidean function is strongly convex smooth with condition number of the same order $O(\zeta_D)$: **only possible because the condition number is a geometric constant and is independent of the condition number of f**.
- We use an Euclidean algorithm on the pull back to **instantiate the subroutine in our algorithm**. This allows to obtain a fast implementation of an inexact ball optimization oracle.
- **Sequential application of the inexact ball optimization oracle** for $\widetilde{O}(R/D) = \widetilde{O}(\zeta^2)$ times leads to global accelerated convergence (distance to x^* can only grow by a factor of 3).
- Alternatively, for $D = O(|\kappa_{\min}|^{-1/2})$, we show that the following **projection operator** is a convex well-defined problem. And with access to it, we can implement inexact optimization oracles over balls $\mathcal{X}$ for bigger D, **shaving off a ζ in the convergence rates**. It can be easily solved for the hyperbolic space.

$$x_{t+1} = \operatorname*{arg\,min}_{y \in \mathcal{X}}\{\langle \nabla f(x_t), y - x_t\rangle_{x_t} + \frac{L}{2}d(x_t, y)^2\} = \mathrm{Exp}_{x_t}\Big(\operatorname*{arg\,min}_{y \in \mathrm{Exp}_{x_t}^{-1}(\mathcal{X})} \|-\frac{1}{L}\nabla f(x_t) - y\|_{x_t}^2\Big).$$

References

[AS20] Kwangjun Ahn and Suvrit Sra. "From Nesterov's Estimate Sequence to Riemannian Acceleration". In: *arXiv preprint arXiv:2001.08876* (2020).

[CB21] Christopher Criscitiello and Nicolas Boumal. "Negative curvature obstructs acceleration for geodesically convex optimization, even with exact first-order oracles". In: *CoRR* abs/2111.13263 (2021).

[KY22] Jungbin Kim and Insoon Yang. "Accelerated Gradient Methods for Geodesically Convex Optimization: Tractable Algorithms and Convergence Analysis". In: *International Conference on Machine Learning, ICML*. 2022.

[LM19] Mario Lezcano-Casado and David Martínez-Rubio. "Cheap Orthogonal Constraints in Neural Networks: A Simple Parametrization of the Orthogonal and Unitary Group". In: *Proceedings of the 36th International Conference on Machine Learning, ICML*. 2019.

[Mar22] David Martínez-Rubio. "Global Riemannian Acceleration in Hyperbolic and Spherical Spaces". In: *International Conference on Algorithmic Learning Theory (ALT)*. 2022.

[Nes05] Yurii Nesterov. "Smooth minimization of non-smooth functions". In: *Math. Program.* 103.1 (2005), pp. 127–152.

[ZS18] Hongyi Zhang and Suvrit Sra. "An Estimate Sequence for Geodesically Convex Optimization". In: *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*. 2018.