Towards Characterizing the Complexity of Riemannian Online Convex Optimization

Hibiki Fukushima FUKUSHIMA-HIBIKI233@G.ECC.U-TOKYO.AC.JP
The University of Tokyo

Hiroshi Hirai HIRAI.HIROSHI@MATH.NAGOYA-U.AC.JP
Nagoya University

Shinji Ito SHINJI@MIST.I.U-TOKYO.AC.JP
The University of Tokyo and RIKEN

Abstract

Online Convex Optimization (OCO) over Riemannian manifolds raises fundamental questions about how geometry affects algorithmic performance. While Riemannian Online Gradient Descent (R-OGD) has been shown to achieve a regret upper bound of $O(DL\sqrt{\zeta T})$, where ζ depends on the manifold's curvature, the tightness of this bound remained unclear. We first establish a matching lower bound of $\Omega(DL\sqrt{\zeta T})$ for R-OGD, valid for any fixed step-size schedule. This shows that the worst-case regret of R-OGD is $\Theta(DL\sqrt{\zeta T})$, and that the effect of manifold curvature appears as a multiplicative factor of $\sqrt{\zeta}$ in the regret. In contrast to the Euclidean setting—where OGD is minimax optimal and regret bounds are independent of feedback models—this result reveals that geometry can substantially degrade the performance of first-order algorithms. Our second contribution addresses the full-information setting. In this setting, we propose a new algorithm, R-FTRL, based on a Riemannian extension of Follow-the-Regularized-Leader. R-FTRL achieves a regret bound of $O(DL\sqrt{T})$, independent of the curvature. This result confirms that curvature-independent regret is achievable with full-information feedback, consistent with prior results. Our findings further support the fundamental separation between first-order and full-information models in non-Euclidean settings, illuminating the interactions between feedback structure and geometry.

1. Introduction

In this paper, we consider the following Riemannian online convex optimization setting:

A player is given a compact g-convex set $\mathcal{K}$ on a manifold $\mathcal{M}$ and a number of steps T. At each step $t = 1, 2, \ldots, T$, the player selects $x_t \in \mathcal{K}$. After choosing x_t, a loss function $f_t : \mathcal{K} \to \mathbb{R}$, which is g-convex, is determined, and the player incurs a loss $f_t(x_t)$.The function f_t can be selected arbitrarily and may depend on previous decisions $x_1, \ldots, x_t$. Then, the player receives feedback about f_t. Common feedback types include *full information feedback*, where the player receives all information about the function f_t; *first order feedback*, where the player obtains the function value $f_t(x_t)$ and a subgradient $g_t \in \partial f_t(x_t)$; and *bandit feedback*, where only the function value $f_t(x_t)$ is observed. Algorithms that utilize first order feedback are referred to as *first order algorithms*. We measure the learning performance using regret, defined as:

$$\mathrm{Reg}_T(u) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(u), \quad \mathrm{Reg}_T = \sup_{u^* \in \mathcal{K}} \mathrm{Reg}_T(u^*) = \sum_{t=1}^{T} f_t(x_t) - \inf_{u^* \in \mathcal{K}} \sum_{t=1}^{T} f_t(u^*).$$

In this paper, as in the previous work by Wang et al. [27], we introduce the following assumptions:

Assumption 1 *The loss function $f_t : \mathcal{K} \to \mathbb{R}$ is g-convex and L-Lipschitz continuous.*

Assumption 2 *The feasible set $\mathcal{K}$ is a bounded compact g-convex set which has a diameter $D \in \mathbb{R}_{>0}$: $\max_{x,y\in\mathcal{K}} d(x, y) = D < +\infty$.*

Assumption 3 *$(\mathcal{M}, g)$ is a Hadamard manifold whose sectional curvature lies within the interval $[\kappa, 0]$, where $\kappa \leq 0$.*

The theory and methodology of Online Convex Optimization (OCO) [11, 18] have developed significantly, both through their diverse applications—such as sequential prediction [26], model ensemble (boosting) [7, 16], sequential portfolio selection [5, 10]—and through their role as a mathematical foundation in fields like online learning theory [3], algorithmic game theory [20], and reinforcement learning [14]. One most notable algorithms for this problem is Online Gradient Descent (OGD) [30], which achieves a regret bound of $O(DL\sqrt{T})$, where D is the diameter of the feasible set $\mathcal{K}$ and L is the Lipschitz constant of the loss functions. It is also known that this bound is tight, i.e., there exists a sequence of loss functions such that the regret is $\Omega(DL\sqrt{T})$ [11].

Recent research [4, 12, 27, 28] has sought to extend the framework of online convex optimization (OCO) to settings where the decision variables lie on Riemannian manifolds and the loss functions exhibit geodesic convexity. Extending OCO from Euclidean spaces to Riemannian manifolds raises a fundamental question:

> **Research Question:** *How does the intrinsic difficulty of the problem change in transitioning from Euclidean geometry to a manifold setting?*

The objective of this work is to offer a refined and rigorous answer to this fundamental question. To date, our understanding of this question has remained limited. Wang et al. [27] proposed Riemannian OGD (R-OGD), which achieves a regret upper bound of $O(DL\sqrt{\zeta T})$, where $\zeta > 1$ is a parameter depending on the manifold's sectional curvature and the diameter of the feasible region. However, the established lower bound is only $\Omega(DL\sqrt{T})$, which implies that there remains a gap of factor $\sqrt{\zeta}$ between the upper and lower bounds for first-order methods. This stands in sharp contrast to the full-information setting, where Roux et al. [21] recently achieved a curvature-independent regret bound of $O(DL\sqrt{T})$ using an Online Mirror Descent type update. The same fundamental uncertainty persists in the offline setting of optimization over manifolds. While some recent work has established lower bounds incorporating ζ for specific algorithms, such as gradient descent with a Polyak step size [6], no matching lower bounds have been found for gradient descent with the step size $\alpha_t = \frac{D}{L\sqrt{\zeta t}}$ as proposed by Zhang and Sra [29]. More broadly, for many first-order methods on Riemannian manifolds, tight characterizations of complexity—particularly lower bounds—remain elusive. This highlights a fundamental gap in our theoretical understanding of optimization in curved spaces, both in the online and offline settings.

1.1. Contributions

The first contribution of this work is to provide a tight bound on the (worst-case) regret achievable by Online Gradient Descent (OGD) in Riemannian online convex optimization, in terms of the geometric structure of the underlying manifold. Specifically, we show that for Riemannian OGD (R-OGD),

Table 1: Comparison of upper and lower bounds for Riemannian online convex optimization problem. Regret bounds in gray boxes are new results in this paper.

Feasible set	Euclidean	Riemannian	
Feedback type	full-information/first-order	full-information	first-order
Upper bounds	$O(DL\sqrt{T})$: OGD[30],FTRL[9]	$O(DL\sqrt{T})$: RIOD[21] $O(DL\sqrt{T})$: R-FTRL	$O(DL\sqrt{\zeta T})$: R-OGD[27]
Lower bounds	$\Omega(DL\sqrt{T})$[19]	$\Omega(DL\sqrt{T})$[27]	$\Omega(DL\sqrt{T})$[27] $\Omega(DL\sqrt{\zeta T})$: R-OGD

there exists a sequence of loss functions under which the regret is lower bounded by $\Omega(DL\sqrt{\zeta T})$. This lower bound holds uniformly over all predetermined step-size schedules. Combined with the previously established upper bound of $O(DL\sqrt{\zeta T})$ [27], this result implies that the worst-case regret of OGD is $\Theta(DL\sqrt{\zeta T})$, thereby quantifying the effect of the manifold geometry through a multiplicative factor of $\sqrt{\zeta}$. It is important to note that the regret lower bound discussed above applies specifically to the R-OGD algorithm. Thus, it does not necessarily characterize the inherent difficulty of the problem itself. Nevertheless, since R-OGD constitutes a natural extension of the classical OGD algorithm—which achieves tight bounds in the Euclidean setting—understanding the regret behavior of R-OGD represents a significant step toward elucidating the intrinsic complexity of the problem and the influence of manifold geometry.

Our second main contribution is to propose Riemannian FTRL (R-FTRL) for the full-information feedback model. We prove that R-FTRL achieves a regret bound of $O(DL\sqrt{T})$, which is independent of the manifold's geometric structure. The lower bound $\Omega(DL\sqrt{T})$ was established in [27], and our contribution lies in demonstrating that the FTRL framework can also match this bound from above. The proposed algorithm is based on the *Follow-the-Regularized-Leader (FTRL)* framework, widely used in the Euclidean setting, and we refer to our extension as R-FTRL. Unlike the Euclidean case, where first-order approximations (i.e., gradients) of the loss functions are often sufficient, our algorithm directly utilizes the original loss functions. As a result, it requires access to full-information feedback rather than merely first-order information.

The regret bounds from prior work and our contributions are summarized in Table 1. From this, several noteworthy contrasts between the Euclidean and manifold settings can be observed in terms of worst-case regret: (i) In the Euclidean case, OGD is minimax optimal, whereas this is not the case on manifolds. (ii) In Euclidean spaces, OGD and FTRL exhibit the same (optimal) performance, while in the manifold setting, there is a performance gap of a factor of $\sqrt{\zeta}$ between them. (iii) In the Euclidean setting, there appears to be no distinction between the first-order and full-information feedback models in terms of minimax regret. In contrast, in the manifold setting, [21] and our results highlight that a performance gap exists between these two feedback models.

2. Preliminaries

Let $(\mathcal{M}, g)$ be a Riemannian manifold. At each point $x \in \mathcal{M}$, the tangent space $T_x\mathcal{M}$ is equipped with an inner product $\langle \cdot, \cdot \rangle$ and a norm $\| \cdot \|$ defined by the Riemannian metric g. We denote the distance between $x, y \in \mathcal{M}$ by $d(x, y)$. A geodesic is a curve that locally minimizes the length.

We assume that the manifold $\mathcal{M}$ is (geodesically) complete. Then, the exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ is defined for $\xi \in T_x\mathcal{M}$ by $\exp_x(\xi) = \gamma(1)$, where $\gamma : [0,1] \to \mathcal{M}$ is the geodesic satisfying $\gamma(0) = x$ and $\dot\gamma(0) = \xi$. A set $\mathcal{K} \subseteq \mathcal{M}$ is called geodesically (totally) convex (g-convex) if for any two points $x, y \in \mathcal{K}$, any geodesic connecting x and y is contained in $\mathcal{K}$. Additionally, if a geodesic is unique, the set is called *uniquely geodesically convex*. A function $f : \mathcal{K} \to \mathbb{R}$ is called *g-convex* if for any geodesic $\gamma : [0,1] \to \mathcal{K}$, it holds for all $t \in [0,1]$ that $f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1))$. A function $f : \mathcal{K} \to \mathbb{R}$ is called *geodesically L-Lipschitz (g-L-Lipschitz)* if there exists a constant $L \geq 0$ such that it holds for any $x, y \in \mathcal{K}$ that $|f(x) - f(y)| \leq L \cdot d(x,y)$. A simply connected complete manifold having non-positive sectional curvature everywhere is called a Hadamard manifold. A Hadamard manifold is uniquely geodesically convex. By Cartan-Hadamard theorem, the exponential map $\exp_x$ and its inverse $\exp_x^{-1}$ are diffeomorphisms. On a Hadamard manifold, the identity $\|\exp_x^{-1}(y)\| = d(x,y)$ holds. Moreover, the function $x \mapsto d(x,y)$ is geodesically convex. The subdifferential of a function $f : \mathcal{K} \to \mathbb{R}$ at a point $x \in \mathcal{K}$ is a subset of $T_x\mathcal{M}$ defined by $\partial f(x) = \{g \in T_x\mathcal{M} \mid f(y) \geq f(x) + \langle g, \exp_x^{-1}(y)\rangle,\ \forall y \in \mathcal{K}\}$, where $g \in \partial f(x)$ is called a subgradient of f at x. For any geodesically convex function on a Hadamard manifold, the subdifferential at each point is non-empty (see [8, p. 139] and [25, Chapter 3, Theorem 4.5]).

3. Analysis of Riemannian online gradient descent

Riemannian Online Gradient Descent (R-OGD) [27] is a natural extension of OGD [30] to Riemannian manifolds, which works given the first-order feedback. The procedure of R-OGD is summarized in Algorithm 1. In this section, we first review the algorithm and its regret upper bound. Then, we show that the regret lower bound of R-OGD is $\Omega(DL\sqrt{\zeta T})$, where ζ is a parameter depending on the sectional curvature of the manifold and the diameter of the feasible set.

Algorithm 1: Riemannian Online Gradient Descent (R-OGD) [27]

Data: A feasible set $\mathcal{K} \subseteq \mathcal{M}$, time horizon T, step sizes $\{\alpha_t\}_{t=1}^T$, and an initial point $x_1 \in \mathcal{K}$.
for $t \leftarrow 1$ **to** T **do**
    Play x_t and observe a subgradient $g_t \in \partial f_t(x_t)$ of f_t at x_t;
    Update x_{t+1} with $x_{t+1} = \mathcal{P}_\mathcal{K}(\exp_{x_t}(-\alpha_t g_t))$, where $\mathcal{P}_\mathcal{K}$ is the Riemannian projection mapping of x onto $\mathcal{K}$, i.e., $\mathcal{P}_\mathcal{K}(x) \in \arg\min_{y\in\mathcal{K}} d(x,y)$;
end

3.1. Review of existing results: regret upper bound for R-OGD

Wang et al. [27] showed that R-OGD achieves the following regret upper bound:

Theorem 4 (Theorem 5 of Wang et al. [27]) *Suppose that Assumptions 1, 2, and 3 hold. Algorithm 1 (R-OGD) with step sizes* $\left\{\alpha_t = \frac{D}{L\sqrt{\zeta t}}\right\}$ *achieves* $\mathrm{Reg}_T \leq \frac{3}{2}DL\sqrt{\zeta T}$ *for any T and any initial point $x_1 \in \mathcal{K}$, where $\zeta = \zeta(\kappa, D) \geq 1$ is defined as*

$$\zeta(\kappa, D) = \frac{\sqrt{|\kappa|}D}{\tanh\left(\sqrt{|\kappa|}D\right)}\ (\kappa < 0), \quad \zeta(\kappa, D) = 1\ (\kappa \geq 0). \tag{1}$$

This theorem can be proved in a similar way to the Euclidean case, but there are several points requiring careful attention. The proof begins by upper bounding the regret as:

$$f_t(x_t) - f_t(u) \leq \left\langle -g_t, \exp_{x_t}^{-1}(u) \right\rangle = \tilde{f}_t(x_t) - \tilde{f}_t(u), \quad \text{where } \tilde{f}_t(x) = \left\langle g_t, \exp_{x_t}^{-1}(x) \right\rangle. \tag{2}$$

We note that this expression essentially corresponds to the first-order characterization of convexity in the Euclidean case, namely the following inequality:

$$f_t(x_t) - f_t(u) \leq \langle g_t, x_t - u \rangle = \tilde{f}_t(x_t) - \tilde{f}_t(u), \qquad \text{where } \tilde{f}_t(x) = \langle g_t, x - x_t \rangle. \tag{3}$$

From (2), it becomes clear that it suffices to bound the regret with respect to $\tilde{f}_t$ instead of f_t. Note, however, that while $\tilde{f}_t$ in (3) is a linear function—and hence convex—$\tilde{f}_t$ in equation (2) is not even linear, [1] nor is it g-convex. This observation indicates that bounding the regret for $\tilde{f}_t$ in our case requires a different analytical technique that generalizes beyond the Euclidean setting.

3.2. Regret lower bound for R-OGD

In this subsection, we present one of the main results of this paper: the $O(DL\sqrt{\zeta T})$ regret upper bound established in Theorem 4 is tight, as long as one uses R-OGD.

Theorem 5 *For any T and any predetermined sequence of step sizes $\{\alpha_t\}_{t=1}^T$, there exists $\mathcal{K} \subseteq \mathcal{M}$ and a sequence of loss functions $\{f_t\}_{t=1}^T$ satisfying Assumptions 1, 2, and 3 for which the regret of R-OGD (Algorithm 1) is lower bounded as $\mathrm{Reg}_T = \Omega\left(DL\sqrt{\zeta T}\right)$, where ζ is defined in (1).*

Together with Theorem 4, this result provides a $\Theta(DL\sqrt{\zeta T})$ characterization of the worst-case regret achievable by R-OGD with any fixed step-size sequence, which is tight up to constant factors. At the same time, it reveals that the parameter ζ serves as an essential quantity capturing how the structure of the Hadamard manifold affects the worst-case performance of R-OGD.

The following proposition is crucial for establishing the lower bound in Theorem 5

Proposition 6 (Lower bound for first-order algorithm) *Fix an arbitrary $\mathcal{M}$ satisfying Assumption 3 and and arbitrary $\mathcal{K} \subset \mathcal{M}$ satisfying Assumption 2. For any first-order deterministic algorithm, the worst case regret under Assumptions 1 is bounded from below as follows:*

$$\sup_{\{f_t\}_{t=1}^T \in \mathcal{F}} \max_{u \in \mathcal{K}} \sum_{t=1}^T \left(f_t(x_t) - f_t(u)\right) \geq \sup_{\{g_t\}_{t=1}^T \in \mathcal{G}} \max_{u \in \mathcal{K}} \sum_{t=1}^T \left(\tilde{f}_t(x_t) - \tilde{f}_t(u)\right), \tag{4}$$

where $\mathcal{F}$ denotes the set of all loss function sequences satisfying Assumption 1, $\mathcal{G}$ denotes the set of all sequences $\{g_t\}$ such that $g_t \in T_{x_t}\mathcal{M}$ and $\|g_t\| \leq L/\sqrt{2}$, and $\tilde{f}_t$ is defined in (2).

At this point, we emphasize the distinction between the Euclidean and Riemaniann settings. In the Euclidean case, when the loss functions are convex, the surrogate loss $\tilde{f}_t$ appearing on the right-hand side of (4) can be expressed as in (3). Then since $\tilde{f}_t$ is linear, it remains convex, and thus the problem of minimizing regret with respect to $\tilde{f}_t$ still falls within the scope of online convex optimization. In contrast, in the setting of g-convex functions over Hadamard manifolds, the surrogate loss $\tilde{f}_t$ appearing on the right-hand side of (4) is *not necessarily g-convex*. As a result, the problem of

1. On a Riemannian manifold, a "linear" (or affine) function is typically defined as one that is both g-convex and g-concave. However, on non-Euclidean Riemannian manifolds, nontrivial linear functions rarely exist; see e.g., [13].

minimizing regret with respect to $\tilde{f}_t$ goes beyond the class of online convex optimization problems and can be interpreted as belonging to a broader class of online learning problems. Importantly, this proposition holds not only for R-OGD but for any first-order algorithm, and it captures the intrinsic difficulty gap between the first-order feedback and the full-information feedback models.

4. Follow the regularized leader

In this section, we introduce a Riemannian extension of the Follow-the-Regularized-Leader (FTRL) algorithm. FTRL is a classical and widely used algorithm in online convex optimization, known for achieving minimax-optimal regret bounds [11, 18]. At each round, it outputs the minimizer of the cumulative loss (typically approximated by surrogate losses) plus a regularization term.

In the Euclidean setting, it is standard to apply FTRL to linear surrogate losses $\tilde{f}_t$ given in (3), which arise from first-order approximations of the original convex losses. This formulation not only preserves convexity but also ensures that the update at each round is uniquely defined. However, extending this approach to Riemannian manifolds presents fundamental challenges. In particular, while linear functions are convex in Euclidean spaces, the analogous surrogate losses $\tilde{f}_t$ in the Riemannian setting given in (2) are not necessarily g-convex, even when the original loss functions are. As a result, the minimization problem at each step may fail to have a unique solution, or may even fall outside the class of g-convex optimization problems.

To address this issue, we consider a Riemannian Follow-the-Regularized-Leader algorithm (R-FTRL) that directly uses the original loss functions rather than their surrogates. The core of this method is the following update rule for each step $t = 1, \ldots, T$:

$$x_t \in \arg\min_{x \in \mathcal{K}} \left\{ \psi_t(x) + \sum_{s=1}^{t-1} f_s(x) \right\}. \tag{5}$$

This approach avoids the geometric inconsistency introduced by surrogate approximations and ensures that the algorithm operates within the well-defined structure of Riemannian geometry. While this design requires access to full-information feedback, it allows us to derive curvature-independent regret bounds, as shown in the following theorem:

Theorem 7 *Suppose that Assumptions 1, 2, and 3 hold. Let p be an arbitrary point in $\mathcal{K}$. Then, R-FTRL with regularizer $\psi_t(x) = \frac{L\sqrt{t}}{2D} d^2(x, p)$ achieves $\mathrm{Reg}_T \leq \frac{3}{2} DL\sqrt{T}$.*

Theorem 7 establish that R-FTRL achieves the optimal regret bound for Riemannian online convex optimization. Notably, the bound matches that of FTRL in the Euclidean setting, namely $O(DL\sqrt{T})$, and does not depend on the curvature parameter ζ. This implies that R-FTRL is robust to the geometry of the underlying manifold, in contrast to R-OGD, whose regret grows with $\sqrt{\zeta}$.

This observation highlights a sharp distinction between the Euclidean and Riemannian settings. In Euclidean spaces, OGD and FTRL are both minimax optimal and yield the same regret bounds under standard assumptions. However, in the Riemannian setting, this equivalence breaks down: while R-OGD suffers from curvature-dependent regret, R-FTRL achieves curvature-independent performance. This mirrors the behavior of [21] and confirms a fundamental separation in the effectiveness of first-order and full-information algorithms under non-Euclidean geometry.

Acknowledgement

Hiroshi Hirai was supported by JSPS KAKENHI Grant Number JP24K21315. Shinji Ito was was supported by JSPS KAKENHI Grant Number JP25K03184.

References

[1] Kimon Antonakopoulos, Elena Veronica Belmega, and Panayotis Mertikopoulos. Online and stochastic optimization beyond lipschitz continuity: A riemannian approach. In *ICLR 2020-International Conference on Learning Representations*, pages 1–20, 2020.

[2] Miroslav Bačák. *Convex analysis and optimization in Hadamard spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. De Gruyter, Berlin, 2014. ISBN 978-3-11-036103-2; 978-3-11-036162-9. doi: 10.1515/9783110361629. URL https://doi.org/10.1515/9783110361629.

[3] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[4] Hengchao Chen and Qiang Sun. Decentralized online riemannian optimization with dynamic environments. *arXiv preprint arXiv:2410.05128*, 2024.

[5] Thomas M Cover. Universal portfolios. *Mathematical finance*, 1(1):1–29, 1991.

[6] Christopher Criscitiello and Nicolas Boumal. Curvature and complexity: Better lower bounds for geodesically convex optimization. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2969–3013. PMLR, 2023.

[7] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[8] R. E. Greene and K. Shiohama. Convex functions on complete noncompact manifolds: topological structure. *Invent. Math.*, 63(1):129–157, 1981. ISSN 0020-9910,1432-1297. doi: 10.1007/BF01389196. URL https://doi.org/10.1007/BF01389196.

[9] Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine Learning*, 80:165–188, 2010.

[10] Elad Hazan, Adam Kalai, Satyen Kale, and Amit Agarwal. Logarithmic regret algorithms for online convex optimization. In *International Conference on Computational Learning Theory*, pages 499–513. Springer, 2006.

[11] Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[12] Zihao Hu, Guanghui Wang, and Jacob D Abernethy. Riemannian projection-free online learning. *Advances in Neural Information Processing Systems*, 36:41980–42014, 2023.

[13] Nobuhiro Innami. Splitting theorems of Riemannian manifolds. *Compositio Math.*, 47(3):237–247, 1982. ISSN 0010-437X,1570-5846. URL http://www.numdam.org/item?id=CM_1982__47_3_237_0.

[14] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[15] John M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, second edition, 2018. ISBN 978-3-319-91754-2; 978-3-319-91755-9.

[16] Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.

[17] David Martínez-Rubio, Christophe Roux, and Sebastian Pokutta. Convergence and trade-offs in riemannian gradient descent and riemannian proximal point. In *Proceedings of the 41st International Conference on Machine Learning, ICML 2024 Vienna, Austria, July 21-27*, volume 235 of *Proceedings of Machine Learning Research*, pages 34920–34948. PMLR, 2024. URL https://arxiv.org/abs/2403.10429.

[18] Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

[19] Francesco Orabona and Dávid Pál. Scale-free online learning. *Theoretical Computer Science*, 716:50–69, 2018.

[20] Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.

[21] Christophe Roux, David Martínez-Rubio, and Sebastian Pokutta. Implicit riemannian optimism with applications to min-max problems. In *Forty-second International Conference on Machine Learning*, 2025.

[22] Emre Sahinoglu and Shahin Shahrampour. Online optimization on hadamard manifolds: Curvature independent regret bounds on horospherically convex objectives. *arXiv preprint arXiv:2509.11236*, 2025.

[23] Takashi Sakai. *Riemannian geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1996. ISBN 0-8218-0284-4. doi: 10.1090/mmono/149. URL https://doi.org/10.1090/mmono/149. Translated from the 1992 Japanese original by the author.

[24] Yunlu Shu, Jiaxin Jiang, Lei Shi, and Tianyu Wang. Revisit gradient descent for geodesically convex optimization, 2025. URL https://arxiv.org/abs/2504.06814.

[25] Constantin Udriste. *Convex functions and optimization methods on Riemannian manifolds*, volume 367. Springer Science & Business Media, 1994.

[26] Volodya Vovk. Competitive on-line statistics. *International Statistical Review*, 69(2):213–248, 2001.

[27] Xi Wang, Zhipeng Tu, Yiguang Hong, Yingyi Wu, and Guodong Shi. Online optimization over riemannian manifolds. *Journal of Machine Learning Research*, 24(84):1–67, 2023.

[28] Xi Wang, Deming Yuan, Yiguang Hong, Zihao Hu, Lei Wang, and Guodong Shi. Riemannian online optimistic algorithms with dynamic regret. *IEEE Transactions on Automatic Control*, 2025.

[29] Hongyi Zhang and Suvrit Sra. First-order methods for geodesically convex optimization. In *Conference on Learning Theory*, pages 1617–1638. PMLR, 2016.

[30] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, pages 928–936, 2003.

Appendix A. Related work

Optimization on Riemannian manifolds has attracted growing attention due to its theoretical significance and practical applications in non-Euclidean domains. Zhang and Sra [29] pioneered the analysis of first-order methods on Hadamard manifolds by introducing novel analytical techniques tailored to the geometric structure of such spaces. They established convergence rate guarantees for Riemannian gradient descent applied to geodesically convex functions, demonstrating that the complexity bounds mirror those in Euclidean spaces, but with additional factors depending explicitly on the curvature and the geometry of the manifold.

Building upon these foundational techniques, Wang et al. [27] extended the analysis to the online setting, introducing Riemannian Online Gradient Descent (R-OGD) for online geodesically convex optimization over Hadamard manifolds. Their results showed that, as in the offline case, the regret upper bounds achieved by R-OGD contain curvature-dependent terms, reflecting the intrinsic difficulty introduced by the manifold geometry. This line of work highlighted that regret rates for first-order methods on curved spaces can be substantially affected by the curvature. One approach to circumventing this dependence is to impose stronger assumptions on the objective functions. It has been shown that under the condition of horospherical convexity—a stronger notion than geodesic convexity—R-OGD can achieve curvature-independent regret bounds [22].

Meanwhile, Antonakopoulos et al. [1] addressed the online optimization of non-Lipschitz convex functions in the Euclidean sense. By leveraging a suitable Riemannian manifold on which these functions become Riemannian-Lipschitz continuous, they applied the FTRL algorithm to establish regret bounds.

More recently, Criscitiello and Boumal [6] conducted a systematic investigation into the role of curvature in the complexity of offline optimization on manifolds, establishing lower bounds that reveal fundamental limitations imposed by negative curvature for certain classes of algorithms. However, their lower bounds apply only to a class of algorithms that encompasses the Polyak step size and its variants, which do not cover the methods of Zhang and Sra [29] or other first-order algorithms. Determining how curvature affects the complexity of g-convex optimization problems when using first-order methods remains an open problem; see, for example, recent progress in Martínez-Rubio et al. [17], Shu et al. [24].

In this work, we address some of these gaps by adapting analytical ideas from Criscitiello and Boumal [6] to the online optimization setting. Our contributions are distinct from, and in some aspects complementary to, Criscitiello and Boumal [6]. First, while their lower bounds are established in the offline setting, we focus on the online scenario and establish lower bounds on the regret of R-OGD under any predetermined step size sequence. In particular, we show that the curvature-dependent lower bound holds for the canonical step size choice $\eta_t = \frac{D}{L\sqrt{\zeta T}}$ introduced by Wang et al. [27], which is the same step size in Zhang and Sra [29]. Furthermore, our results apply to all R-OGD algorithms with predetermined step size sequences, thereby covering a broader class of methods.

Appendix B. Proof of Proposition 6

B.1. Review of the upper bound

First, we provide a brief overview of the proof of Theorem 4 by Wang et al. [27]. What proves to be useful at this point is the following lemma:

Lemma 8 (Special case of Lemma 5 of Zhang and Sra [29]) *Let $a > 0$, $b > 0$ and $c > 0$ be the sides (i.e., side lengths) of a geodesic triangle on a Riemannian manifold with curvature lower bounded by κ. Let A be the angle between sides b and c. We then have*

$$a^2 \leq \zeta(\kappa, c)b^2 + c^2 - 2bc\cos A, \tag{6}$$

where $\zeta(\kappa, c) \geq 1$ is defined in (1)

Let $\tilde{x}_{t+1} = \exp_{x_t}(-\alpha_t g_t)$ be the point obtained by the gradient update step. Using this lemma, the regret can be upper bounded as follows, following a line of reasoning similar to that in the Euclidean case.

$$\begin{aligned}
\mathrm{Reg}_T(u) &= \sum_{t=1}^{T} (f_t(x_t) - f_t(u)) \leq \sum_{t=1}^{T} \left(\tilde{f}_t(x_t) - \tilde{f}_t(u)\right) && \text{(from (2))} \\
&\leq \sum_{t=1}^{T} \left(\frac{1}{2\alpha_t} \left(d^2(x_t, u) - d^2(\tilde{x}_{t+1}, u)\right) + \frac{\alpha_t}{2}\zeta(\kappa, d(x_t, u))\|g_t\|^2 \right) \\
&\leq \sum_{t=1}^{T} \left(\frac{1}{2\alpha_t} \left(d^2(x_t, u) - d^2(x_{t+1}, u)\right) + \frac{\alpha_t}{2}\zeta(\kappa, D)L^2 \right), && (7)
\end{aligned}$$

where the second inequality follows from Lemma 8 applied to the geodesic triangle of vertices $x_t, \tilde{x}_t, u$ and the last inequality follows from the facts that the projection $\mathcal{P}_{\mathcal{K}}$ of any g-convex set $\mathcal{K}$ is distance-nonincreasing [2, Theorem 2.1.12], that $\zeta(\kappa, D)$ is monotone non-decreasing in $D > 0$, and that the norm $\|g_t\|$ of subgradient is at most the Lipschitz constant L. The regret upper bound of Theorem 4 follows by applying this inequality (7) together with the specification of step sizes α_t in the theorem and the bound $d(x_t, u) \leq D$.

B.2. Proof of the lower bound

We start by proving following lemma:

Lemma 9 *Let $\mathcal{M}$ be a Hadamard manifold, and $\mathcal{K} \subset \mathcal{M}$ be a g-convex set. Then, for any $x, u \in \mathcal{K}$, and $g \in T_x\mathcal{M}$, there exists a g-convex and $\sqrt{2}\|g\|$-Lipschitz function $f : \mathcal{K} \to \mathbb{R}$ such that $g \in \partial f(x)$ and*

$$f(x) - f(u) = \langle -g, \exp_x^{-1}(u) \rangle.$$

The proof of this lemma is inspired by the proof of Proposition 35 by Criscitiello and Boumal [6]. All results established in this work but stated without proof in the main text, including this lemma, are proved in the Appendix.

Lemma 9 implies that for any x_t, g_t, and u, there exists a loss function f_t such that the inequality in (2) holds with equality (and thus so does the first inequality in (7)). Moreover, under the first-order feedback model, the algorithm's output x_t depends only on the sequence $\{g_s\}_{s=1}^{t-1}$, and is invariant under changes to f_t, as long as the subgradient condition $g_t \in \partial f_t$ is satisfied. Therefore, in analyzing worst-case regret under first-order feedback, we may equivalently consider the regret induced by the surrogate functions $\tilde{f}_t$ defined in equation (2) in place of the original loss functions f_t.

Proof [Proof of Lemma 9] We can assume $u \neq x$. Let $g_{\|} = \frac{\langle g, \exp_x^{-1}(u)\rangle}{d(x,u)^2} \exp_x^{-1}(u)$ and $g_{\perp} = g - g_{\|}$. Note that $g_{\perp}$ is orthogonal to $g_{\|}$. Consider the geodesic line S passing through u and x, i.e., $S = \{\exp_x(t \exp_x^{-1}(u)) \mid t \in \mathbb{R}\}$. Define $u' \in \mathcal{M}$ by

$$u' = \begin{cases} \exp_x(-\exp_x^{-1}(u)) & \text{if } \langle g, \exp_x^{-1}(u)\rangle > 0, \\ u & \text{if } \langle g, \exp_x^{-1}(u)\rangle \leq 0. \end{cases}$$

Define $f_{\|}, f_{\perp}, f : \mathcal{M} \to \mathbb{R}$ by

$$\begin{aligned} f_{\|}(z) &= \|g_{\|}\| d(z, u') - \|g_{\|}\| d(x, u'), \\ f_{\perp}(z) &= \|g_{\perp}\| d(z, S), \\ f(z) &= f_{\|}(z) + f_{\perp}(z), \end{aligned}$$

where $d(z, S) = \min_{y \in S} d(z, y)$.

Then it holds that

$$g_{\perp} \in \partial f_{\perp}(x). \tag{8}$$

Indeed, consider arbitrary $z \in \mathcal{M}$, let θ and θ^* be the angles between $\exp_x^{-1} u$ and $\exp_x^{-1} z$, and between $\exp_x^{-1} z$ and $g_{\perp}$, respectively. Then we see (8) from

$$\begin{aligned} f_{\perp}(z) - f_{\perp}(x) &= \|g_{\perp}\| \min_{y \in S} d(z, y) \\ &\geq \|g_{\perp}\| \min_{y \in S} (d(x,z)^2 + d(x,y)^2 - 2d(x,z)d(x,y)\cos\theta)^{1/2} \\ &= \|g_{\perp}\| d(x,z) \sin\theta \\ &= \|g_{\perp}\| d(x,z) \cos(\pi/2 - \theta) \\ &\geq \|g_{\perp}\| d(x,z) \cos\theta^* \\ &= \langle g_{\perp}, \exp_x^{-1} v\rangle, \end{aligned}$$

where the first inequality follows from the CAT(0)-inequality for geodesic triangle of vertices x, y, z, and the second from the fact that the angle between $g_{\perp}$ and $\exp_x^{-1} u$ is $\pi/2$ and the triangle inequality $\theta^* + \theta \geq \pi/2$.

In addition, by [23, Proposition 4.8 (1)], it holds that

$$\nabla f_{\|}(x) = g_{\|}.$$

Therefore, $g = g_{\|} + g_{\perp} \in \partial f_{\|}(x) + \partial f_{\perp}(x) \subseteq \partial f(x)$.

Since S is a convex set on a Hadamard manifold, $f_{\perp}$ is g-convex (see [2, Example 2.2.4]). Also $f_{\|}$ is g-convex. Thus, f is g-convex. Clearly, $f(x) - f(u) = \langle -g, \exp_x^{-1}(u)\rangle$ holds by construction. Hence, the Lipschitz constant for f satisfies

$$\|g_{\|}\| + \|g_{\perp}\| \leq \sqrt{2(\|g_{\|}\|^2 + \|g_{\perp}\|^2)} = \sqrt{2}\|g\|$$

where the first inequality follows from the Cauchy–Schwarz inequality, and the second equality holds due to the orthogonality of $g_{\|}$ and $g_{\perp}$. Therefore, f is g-convex and $\sqrt{2}\|g\|$-Lipschitz, completing the proof. ∎

We are now ready to show Proposition 6.

Proof [Proof of Proposition 6] For any first-order algorithm, the sequence $\{x_t\}$ remains unchanged even if $\{f_t\}$ varies, as long as $\{g_t\}$ remain unchanged. Thus, for any sequence of $\{g_t\}$ such that $g_t \in T_{\mathcal{M}}(x_t)$ and $\|g_t\| \leq L/\sqrt{2}$, by keeping g_t fixed and defining a function as in Lemma 9 corresponding to u, we can construct a sequence of L-Lipschitz g-convex functions $f_{t,u}$ that ensures

$$f_{t,u}(u) = \langle -g_t, \exp_{x_t}^{-1}(u)\rangle$$

for all t. For such $\{f_{t,u}\}_{t=1}^{T}$ we obtain the regret bound as follows:

$$\mathrm{Reg}_T(u) = \sum_{t=1}^{T} f_{t,u}(x_t) - f_{t,u}(u) = \sum_{t=1}^{T} \langle -g_t, \exp_{x_t}^{-1}(u)\rangle.$$

■

Appendix C. Proof of Theorem 5

For the proof, consider a hyperbolic space with curvature κ. Consider a geodesic triangle of side lengths a, b, c where the angle of edges of lengths b and c is A. Then the following hyperbolic law of cosines holds:

$$\cosh(\sqrt{|\kappa|}a) = \cosh(\sqrt{|\kappa|}b)\cosh(\sqrt{|\kappa|}c) - \sinh(\sqrt{|\kappa|}b)\sinh(\sqrt{|\kappa|}c)\cos(A).$$

Before the proof, we consider right triangles on hyperbolic space. Consider a right-angled triangle with vertices p, x, u, where right angle is at vertex p, and the angle at x is θ. Let $a = d(p, x)$, $b = d(p, u)$, and $c = d(x, u)$. Then, by the hyperbolic law of cosines, we have:

$$\cosh(\sqrt{|\kappa|}c) = \cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b), \tag{9}$$

$$\cosh(\sqrt{|\kappa|}b) = \cosh(\sqrt{|\kappa|}c)\cosh(\sqrt{|\kappa|}a) - \sinh(\sqrt{|\kappa|}c)\sinh(\sqrt{|\kappa|}a)\cos(\theta). \tag{10}$$

From (9), we have

$$c = \frac{1}{\sqrt{|\kappa|}}\cosh^{-1}(\cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)) \geq \frac{1}{\sqrt{|\kappa|}}\cosh^{-1}(\cosh(\sqrt{|\kappa|}b)) = b. \tag{11}$$

From (10),

$$\begin{aligned}
\cos(\theta) &= \frac{\cosh(\sqrt{|\kappa|}c)\cosh(\sqrt{|\kappa|}a) - \cosh(\sqrt{|\kappa|}b)}{\sinh(\sqrt{|\kappa|}c)\sinh(\sqrt{|\kappa|}a)} \\
&= \frac{\cosh^2(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b) - \cosh(\sqrt{|\kappa|}b)}{\sinh(\sqrt{|\kappa|}c)\sinh(\sqrt{|\kappa|}a)} \\
&= \frac{(\sinh^2(\sqrt{|\kappa|}a) + 1)\cosh(\sqrt{|\kappa|}b) - \cosh(\sqrt{|\kappa|}b)}{\sinh(\sqrt{|\kappa|}c)\sinh(\sqrt{|\kappa|}a)} \\
&= \frac{\sinh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)}{\sinh(\sqrt{|\kappa|}c)} \\
&= \frac{\sinh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)}{\sinh(\cosh^{-1}(\cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)))},
\end{aligned}$$

where the second and fifth equalities follow from (9). Because $\frac{1}{\sinh(\cosh^{-1}(x))} \geq \frac{1}{x}$ for $x \geq 1$, we have

$$
\begin{aligned}
\cos(\theta) &= \frac{\sinh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)}{\sinh(\cosh^{-1}(\cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)))} \\
&\geq \frac{\sinh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)}{\cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)} \\
&= \tanh(\sqrt{|\kappa|}a). \qquad (12)
\end{aligned}
$$

Lemma 10 *Consider the two-dimensional hyperbolic plane $\mathbb{H}^2_\kappa$ with curvature κ. Let $p \in \mathbb{H}^2_\kappa$ and take an orthonormal basis e_1, e_2 of $T_p\mathbb{H}^2_\kappa$. For $r \geq 0$, define*

$$
x = \exp_p(re_1), \quad u = \exp_p\left(\frac{D}{2}e_2\right).
$$

Furthermore, define $g \in T_x\mathbb{H}^2_\kappa$ by

$$
-g = \frac{L}{\sqrt{2}}\frac{\exp_x^{-1}(p)}{d(x,p)}.
$$

If $r \geq \frac{D}{2\sqrt{2\zeta T}}$, $\sqrt{|\kappa|}D \geq 1$, and $\frac{1}{4}\sqrt{\frac{\zeta}{2T}} \leq 1$, then it holds that

$$
\langle -g, \exp_x^{-1}(u)\rangle \geq \frac{DL\sqrt{\frac{\zeta}{T}}}{32}.
$$

Proof We begin by applying (11) and (12) to the geodesic triangle with vertices p, x, u, where we have sides of length $a = d(p,x) = r$, $b = d(p,u) = D/2$, and $c = d(x,u)$.

First, from the law of cosines in (11), it holds that $d(x,u) = c \geq b = D/2$. Combining this result with the bound from (12), we can establish a lower bound for the inner product:

$$
\langle -g, \exp_x^{-1}(u)\rangle = d(x,u)\|g\|\cos(\theta) \geq \frac{D}{2}\frac{L}{\sqrt{2}}\tanh\left(\sqrt{|\kappa|}r\right). \qquad (13)
$$

Next, we use the given condition $r \geq \frac{D}{2\sqrt{2\zeta T}}$. Since the hyperbolic tangent function is monotonically increasing for $x \geq 0$, we can substitute this lower bound for r into the argument to obtain

$$
\frac{DL}{2\sqrt{2}}\tanh\left(\sqrt{|\kappa|}r\right) \geq \frac{DL}{2\sqrt{2}}\tanh\left(\frac{\sqrt{|\kappa|}D}{2\sqrt{2\zeta T}}\right). \qquad (14)
$$

By the assumption $\sqrt{|\kappa|}D \geq 1$ and the relation $\sqrt{|\kappa|}D = \zeta\tanh(\sqrt{|\kappa|}D)$, it follows that $\sqrt{|\kappa|}D > \zeta/2$, since $\tanh(x) > 1/2$ for $x \geq 1$. This provides a simpler lower bound for the argument:

$$
\frac{\sqrt{|\kappa|}D}{2\sqrt{2\zeta T}} > \frac{\zeta/2}{2\sqrt{2\zeta T}} = \frac{1}{4}\sqrt{\frac{\zeta}{2T}}.
$$

Applying the monotonicity of $\tanh$ once more yields

$$
\frac{DL}{2\sqrt{2}}\tanh\left(\frac{\sqrt{|\kappa|}D}{2\sqrt{2\zeta T}}\right) \geq \frac{DL}{2\sqrt{2}}\tanh\left(\frac{1}{4}\sqrt{\frac{\zeta}{2T}}\right). \qquad (15)
$$

Finally, under the condition $\frac{1}{4}\sqrt{\frac{\zeta}{2T}} \leq 1$, we can apply the inequality $\tanh(x) \geq x/2$ for $x \in [0,1]$. This leads to our final result:

$$\frac{DL}{2\sqrt{2}} \tanh\left(\frac{1}{4}\sqrt{\frac{\zeta}{2T}}\right) \geq \frac{DL}{2\sqrt{2}} \cdot \frac{1}{2}\left(\frac{1}{4}\sqrt{\frac{\zeta}{2T}}\right) = \frac{DL\sqrt{\zeta}}{32\sqrt{T}} = \frac{DL}{32}\sqrt{\frac{\zeta}{T}}. \tag{16}$$

Combining inequalities (13) through (16), we obtain:

$$\langle -g, \exp_x^{-1}(u)\rangle \geq \frac{DL}{32}\sqrt{\frac{\zeta}{T}}.$$

∎

Proof of Theorem 5. Let $\mathbb{H}_\kappa^2$ be the two-dimensional hyperbolic space with curvature κ. Let $p \in \mathbb{H}_\kappa^2$ and take an orthonormal basis e_1, e_2 of $T_p\mathbb{H}_\kappa^2$. We define the g-convex set $\mathcal{K}$ as a closed ball of radius $D/2$ centered at p:

$$\mathcal{K} = \{x \in \mathbb{H}_\kappa^2 \mid d(p,x) \leq D/2\}.$$

Define $I = \{t \in \{1,2,\ldots,T\} | \alpha_t \geq \frac{D}{L\sqrt{\zeta T}}\}$ and consider cases based on $|I|$.
Case 1: $|I| \geq \frac{T}{2}$.
Assume $\sqrt{|\kappa|}D \geq 1$, $T \geq 2$ and $\frac{1}{4}\sqrt{\frac{\zeta}{2T}} \leq 1$. Define a geodesic γ such that $\gamma(0) = p, \dot\gamma(0) = e_1$. Let $x_1 = \gamma(\frac{D}{2})$ and define s_t such that $x_t = \gamma(s_t)$. Define

$$g_t = \begin{cases} \frac{L}{\sqrt{2}}\dot\gamma(s_t) & (t \in I \text{ and } s_t \geq 0), \\ -\frac{L}{\sqrt{2}}\dot\gamma(s_t) & (t \in I \text{ and } s_t < 0), \\ 0 & (t \notin I). \end{cases}$$

We here note that $\|g_t\| \leq L/\sqrt{2}$, and hence we can use Proposition 6. Additionally, let $u = \exp_p(\frac{D}{2}e_2)$. Then, consider the geodesic triangle formed by the points p, x_t, u. Let their corresponding side lengths be denoted by $a = d(p, x_t)$, $b = d(p,u)$, and $c = d(x_t, u)$. The sign of the inner product $\langle -g_t, \exp_{x_t}^{-1}(u)\rangle$ is determined by the sign of the cosine of the angle θ at the vertex x_t. From inequality (12), we have $\cos\theta \geq \tanh(\sqrt{|\kappa|}a) \geq 0$. This directly implies that $\langle -g_t, \exp_{x_t}^{-1}(u)\rangle \geq 0$. We next show that

$$\left|\left\{t \in I \;\middle|\; |s_t| \geq \frac{D}{2\sqrt{2\zeta T}}\right\}\right| \geq \frac{|I|}{2}. \tag{17}$$

For any $t \in I$, we consider the quantity $|s_{t+1} - s_t|$ and analyze it by distinguishing two cases depending on whether

$$\hat{x}_{t+1} := \exp_{x_t}(-\alpha_t g_t)$$

lies within the feasible set $\mathcal{K}$.

If $\hat{x}_{t+1} \in \mathcal{K}$, no projection is applied, and hence the change in s_t equals the step length, i.e.,

$$|s_{t+1} - s_t| = \alpha_t \|g_t\|.$$

Conversely, if $\hat{x}_{t+1} \notin \mathcal{K}$, the iterate is projected onto the boundary of $\mathcal{K}$. Because g_t is defined such that the direction $-\alpha_t g_t$ points toward the center p, any step that exits $\mathcal{K}$ must cross p. As a result, the updated point satisfies $|s_{t+1}| = D/2$ with an opposite sign to that of s_t, leading to

$$|s_{t+1} - s_t| = |s_t| + D/2 \geq D/2.$$

These two cases can be combined into a single, unified lower bound: $|s_{t+1} - s_t| \geq \min(\alpha_t \|g_t\|, D/2)$. Furthermore, using the condition for $t \in I$, namely $\alpha_t \geq \frac{D}{L\sqrt{\zeta T}}$, and the fact that $\|g_t\| = L/\sqrt{2}$, we have $\alpha_t \|g_t\| \geq \frac{D}{\sqrt{2\zeta T}}$. Since $\zeta \geq 1$ by definition and we assume $T \geq 2$, it follows that $\sqrt{2\zeta T} \geq 2$, which ensures that $\frac{D}{\sqrt{2\zeta T}} \leq \frac{D}{2}$. This leads to the final lower bound:

$$|s_{t+1} - s_t| \geq \min\left(\alpha_t \|g_t\|, \frac{D}{2}\right) \geq \min\left(\frac{D}{\sqrt{2\zeta T}}, \frac{D}{2}\right) = \frac{D}{\sqrt{2\zeta T}}.$$

Therefore, if both $|s_{t+1}| < \frac{D}{2\sqrt{2\zeta T}}$ and $|s_t| < \frac{D}{2\sqrt{2\zeta T}}$, then$|s_{t+1} - s_t| < \frac{D}{\sqrt{2\zeta T}}$, which contradicts the above inequality. Thus, for each $t \in I$, at least one of t or $t+1$ satisfies $|s_t| \geq \frac{D}{2\sqrt{2\zeta T}}$. Combining this with the fact that $s_{t+1} = s_t$ for all $t \notin I$, it follows that for any two consecutive elements of I, say t_i and t_{i+1}, at least one of them must satisfy $|s_t| \geq \frac{D}{2\sqrt{2\zeta T}}$. Therefore, equation (17) holds.

Using Lemma 10, we obtain that for any $t \in I$ with $|s_t| \geq \frac{D}{2\sqrt{2\zeta T}}$, it holds that

$$\langle -g_t, \exp_{x_t}^{-1}(u) \rangle \geq \frac{DL\sqrt{\frac{\zeta}{T}}}{32}. \tag{18}$$

Then, from Proposition 6, there exists $\{f_t\}$ satisfying Assumptions 1, 2 and 3 such that

$$\mathrm{Reg}_T(u) \geq \sum_{t=1}^{T} \langle -g_t, \exp_{x_t}^{-1}(u) \rangle. \tag{19}$$

Combining (17), the bound of (18) that holds for t such that $|s_t| \geq \frac{D}{2\sqrt{2\zeta T}}$, and the fact that $\langle -g_t, \exp_{x_t}^{-1}(u) \rangle \geq 0$ hold for all other t, we obtain

$$\sum_{t=1}^{T} \langle -g_t, \exp_{x_t}^{-1}(u) \rangle \geq \frac{DL\sqrt{\frac{\zeta}{T}}}{32} \frac{|I|}{2} \geq \frac{DL\sqrt{\frac{\zeta}{T}}}{64} \frac{T}{2} \geq \frac{DL\sqrt{\zeta T}}{128}.$$

Case 2: $|I| < \frac{T}{2}$.
Let $J = \{1, \ldots, T\} \setminus I$, and note that our condition implies $|J| > T/2$. Let the elements of J be ordered as $t_1 < t_2 < \cdots < t_{|J|}$. Define a geodesic γ such that $\gamma(0) = p, \dot{\gamma}(0) = e_1$. Assume $\zeta \leq T$. Let $x_1 = \gamma(-\frac{D}{2}e_1), u = \exp_p(\frac{D}{2}e_1)$, and $x_t = \gamma(s_t)$. Define

$$f_t(x) = \begin{cases} 0 & (t \in I), \\ L\sqrt{\frac{\zeta}{T}} d(x, u) & (t \in J). \end{cases}$$

Then, $f_t(x_t) \geq 0$, $f_t(u) = 0$, and

$$g_t(x_t) = \begin{cases} 0 & (t \in I), \\ -L\sqrt{\frac{\zeta}{T}} \dot{\gamma}(s_t) & (t \in J). \end{cases}$$

We can express

$$s_{t_j} - s_1 = \sum_{i=1}^{j-1} \alpha_{t_i} \|g_{t_i}\| = \sum_{i=1}^{j-1} \alpha_{t_i} L \sqrt{\frac{\zeta}{T}}.$$

Since $\alpha_i < \frac{D}{L\sqrt{\zeta T}}$ for $i \in J$, we have

$$\begin{aligned} s_{t_j} - s_1 &< \sum_{i=1}^{j-1} \frac{D}{L\sqrt{\zeta T}} L \sqrt{\frac{\zeta}{T}} \\ &= \sum_{i=1}^{j-1} \frac{D}{T} \\ &= \frac{D(j-1)}{T}. \end{aligned}$$

Because $s_1 = -\frac{D}{2}$, we obtain

$$s_{t_j} < \frac{D(j-1)}{T} - \frac{D}{2}. \tag{20}$$

Therefore, for all $t = 1, 2, \ldots, T$, we have $s_t < \frac{D(T-1)}{T} - \frac{D}{2} < \frac{D}{2}$. Then, we obtain

$$f_{t_j}(x_{t_j}) = L\sqrt{\frac{\zeta}{T}} d(x_{t_j}, u) = L\sqrt{\frac{\zeta}{T}} \left(\frac{D}{2} - s_{t_j} \right).$$

Consequently, combining the fact that $f_t(x_t) = 0$ for $t \in I$, we obtain

$$\text{Reg}_T(u) = \sum_{t=1}^{T} f_t(x_t) = \sum_{j=1}^{|J|} L\sqrt{\frac{\zeta}{T}} \left(\frac{D}{2} - s_{t_j} \right).$$

Because $|J| > \frac{T}{2}$ and $L\sqrt{\frac{\zeta}{T}} \left(\frac{D}{2} - s_{t_j} \right) \geq 0$, it holds that

$$\sum_{j=1}^{|J|} L\sqrt{\frac{\zeta}{T}} \left(\frac{D}{2} - s_{t_j} \right) \geq \sum_{j=1}^{\frac{T}{2}} L\sqrt{\frac{\zeta}{T}} \left(\frac{D}{2} - s_{t_j} \right).$$

Next, substituting the upper bound for s_{t_j} from (20) into this inequality yields:

$$\begin{aligned} \sum_{j=1}^{\frac{T}{2}} L\sqrt{\frac{\zeta}{T}} \left(\frac{D}{2} - s_{t_j} \right) &= \sum_{j=1}^{\frac{T}{2}} L\sqrt{\frac{\zeta}{T}} \left(\frac{D}{2} - \left(\frac{D(j-1)}{T} - \frac{D}{2} \right) \right) \\ &= \sum_{j=1}^{\frac{T}{2}} L\sqrt{\frac{\zeta}{T}} D \frac{(T-j+1)}{T} \\ &= L\sqrt{\frac{\zeta}{T}} D \frac{3T^2 + 2T}{8T} \\ &\geq \frac{3}{8} DL\sqrt{\zeta T}. \end{aligned}$$

Thus, we have

$$\mathrm{Reg}_T(u) \geq \frac{3}{8} DL\sqrt{\zeta T}.$$

Therefore, in both cases, we have regret lower bounds of $\Omega(DL\sqrt{\zeta T})$.

Appendix D. Proof of Theorems 7 and 11

A function f is called *μ-strongly g-convex* if for any geodesic $\gamma : [0,1] \to \mathcal{K}$, it holds

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)) - \frac{\mu}{2}t(1-t)d^2(\gamma(0), \gamma(1)) \quad (\forall t \in [0,1]).$$

On a Hadamard manifold, the squared distance function $x \mapsto \frac{1}{2}d^2(x,y)$ is 1-strongly geodesically convex ([15, Lemma 12.15] and [2, Remark 2.2.2]). When the objective function f_t is strongly g-convex, the following theorem holds.

Theorem 11 *Suppose that Assumptions 1, 2, and 3 hold. We also assume that f_t is μ-strongly g-convex for all $t \in [T]$. Then, R-FTRL defined in 5 with $\psi_t(x) = 0$ achieves the following regret bound:*

$$\mathrm{Reg}_T \leq \frac{L^2}{2\mu}(1 + \log(T)).$$

In this section, we provide the proofs for Theorems 7 and 11. To demonstrate the regret bound for FTRL on manifolds, we use the following lemma and proposition. In the following, $\mathcal{M}$ is a Hadamard manifold and $\mathcal{K}$ is a convex subset in $\mathcal{M}$.

Proposition 12 *For $f : \mathcal{K} \to \mathbb{R}$, the following conditions are equivalent:*

(i) f is μ-strongly g-convex.

(ii) For all $x, y \in \mathcal{K}$, $\partial f(x)$ is non-empty, and for all $g \in \partial f(x)$, it holds that

$$f(y) \geq f(x) + \langle g, \exp_x^{-1}(y)\rangle + \frac{\mu}{2}d^2(x,y).$$

Proof We apply the proof in [25] for the case of μ-strongly g-convex functions.
$(ii) \Rightarrow (i)$
Let $\gamma : [0,1] \to \mathcal{K}$ be the geodesic connecting $\gamma(0) = x$ and $\gamma(1) = y$, and define $\bar{\gamma}(t) = \gamma(1-t)$. Fix t and set $u(s) = t + s(1-t)$. Define $\alpha(s) = \gamma(u(s)) = \gamma(t + s(1-t))$ and $\beta(s) = \bar{\gamma}(1 - t + st)$. Then,

$$\alpha(0) = \gamma(t), \quad \frac{\mathrm{d}\alpha}{\mathrm{d}s}(0) = (1-t)\dot{\gamma}(t),$$
$$\beta(0) = \gamma(t), \quad \frac{\mathrm{d}\beta}{\mathrm{d}s}(0) = -t\dot{\gamma}(t).$$

By (ii), for any $g \in \partial f(\gamma(t))$, we obtain

$$f(y) \geq f(\gamma(t)) + (1-t)\langle g, \dot{\gamma}(t)\rangle + \frac{\mu}{2}(1-t)^2 d^2(x,y),$$
$$f(x) \geq f(\gamma(t)) - t\langle g, \dot{\gamma}(t)\rangle + \frac{\mu}{2}t^2 d^2(x,y).$$

Thus,

$$tf(y) + (1-t)f(x) \geq f(\gamma(t)) + \frac{\mu}{2}t(1-t)d^2(x,y).$$

$(i) \Rightarrow (ii)$
Since f is g-convex, $\partial f(x)$ is non-empty. For any $g \in \partial f(x)$, by the definition of subgradient,

$$\langle g, \exp_x^{-1}(\gamma(t))\rangle = t\langle g, \exp_x^{-1}(y)\rangle \leq f(\gamma(t)) - f(x).$$

Using the μ-strongly g-convexity of f, we get

$$\langle g, \exp_x^{-1}(y)\rangle \leq \frac{f(\gamma(t)) - f(x)}{t} \leq f(y) - f(x) - \frac{\mu}{2}(1-t)d^2(x,y).$$

Taking the limit as $t \to 0$ gives (ii). ■

Lemma 13 *Let $f : \mathcal{K} \to \mathbb{R}$ be a μ-strongly g-convex function. Then, for any $x, y \in \mathcal{K}$ and $g \in \partial f(x)$, it holds that*

$$f(x) - f(y) \leq \frac{1}{2\mu}\|g\|^2.$$

Proof

$$\begin{aligned} f(x) - f(y) &\leq -\langle g, \exp_x^{-1}(y)\rangle - \frac{\mu}{2}d^2(x,y) \quad (\because f \text{ is } \mu\text{-strongly g-convex}) \\ &= -\frac{\mu}{2}\|\exp_x^{-1}(y) + \frac{1}{\mu}g\|^2 + \frac{1}{2\mu}\|g\|^2 \\ &\leq \frac{1}{2\mu}\|g\|^2. \end{aligned}$$

■

Lemma 14 *Let $f : \mathcal{K} \to \mathbb{R}$ be g-convex. Then, $x^* \in \arg\min_{x\in\mathcal{K}} f(x)$ if and only if $0 \in \partial f(x^*)$.*

This follows from $x^* \in \arg\min_{x\in\mathcal{K}} f(x) \Leftrightarrow f(x) \geq f(x^*) = f(x^*) + \langle 0, \exp_{x^*}^{-1}(x)\rangle$ $(\forall x \in \mathcal{K})$.

Proof of Theorems 7 and 11. Using these lemmas, we derive an upper bound for FTRL on a manifold. Define functions F_t as $F_t(x) = \psi_t(x) + \sum_{i=1}^{t-1} f_i(x)$, with $\psi_{T+1} = \psi_T$.

Lemma 15 *The following holds:*

$$\sum_{t=1}^{T}(f_t(x_t) - f_t(u)) = \psi_{T+1}(u) - \min_{x\in\mathcal{K}}\psi_1(x) + \sum_{t=1}^{T}[F_t(x_t) - F_{t+1}(x_{t+1}) + f_t(x_t)] + F_{T+1}(x_{T+1}) - F_{T+1}(u).$$

Proof

$$\begin{aligned}
-\sum_{t=1}^{T} f_t(u) &= \psi_{T+1}(u) - F_{t+1}(u) \\
&= \psi_{T+1}(u) - F_1(x_1) + F_1(x_1) - F_{T+1}(x_{T+1}) + F_{T+1}(x_{T+1}) - F_{T+1}(u) \\
&= \psi_{T+1}(u) - F_1(x_1) + \sum_{t=1}^{T}[F_t(x_t) - F_{t+1}(x_{t+1})] + F_{T+1}(x_{T+1}) - F_{T+1}(u).
\end{aligned}$$

Since $F_1(x_1) = \psi_1(x_1) = \min_{x\in\mathcal{K}} \psi_1(x)$, the lemma follows. ■

Lemma 16 *Assume F_t is g-convex, and $F_t + f_t$ is λ_t-strongly g-convex. Then, for $g_t \in \partial f_t(x_t)$,*

$$F_t(x_t) - F_{t+1}(x_{t+1}) + f_t(x_t) \le \frac{1}{2\lambda_t}\|g_t\|^2 + \psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1}) \quad (\forall g_t \in \partial f_t(x_t)).$$

Proof Since $x_t \in \arg\min_{x\in\mathcal{K}} F_t(x)$, Lemma 14 implies that $0 \in \partial F_t(x_t)$. In particular, $g_t \in \partial f_t(x_t)$ is also in $\partial(F_t + f_t)(x_t)$. By Lemma 13 with λ_t-strongly g-convexity of $F_t + f_t$, we have

$$\begin{aligned}
F_t(x_t) - F_{t+1}(x_{t+1}) + f_t(x_t) &= (F_t(x_t) + f_t(x_t)) - (F_t(x_{t+1}) + f_t(x_{t+1})) \\
&\quad + \psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1}) \\
&\le \frac{\|g_t\|^2}{2\lambda_t} + \psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1}).
\end{aligned}$$

■

Lemma 17 *Let $\psi : \mathcal{K} \to \mathbb{R}$ be μ-strongly g-convex. Let α_t be a sequence of positive numbers that satisfies $\alpha_t \le \alpha_{t+1}$. Define $\psi_t : \mathcal{K} \to \mathbb{R}_{\ge 0}$ by*

$$\psi_t(x) = \frac{\psi(x) - \min_{z\in\mathcal{K}} \psi(z)}{\alpha_{t-1}}.$$

Then, for R-FTLR using ψ_t as the regularization function, the following inequality holds:

$$\text{Reg}_T(u) \le \psi_T(u) + \frac{1}{2\mu}\sum_{t=1}^{T} \alpha_{t-1}\|g_t\|^2.$$

Proof From Lemma 16, we have

$$\begin{aligned}
&\sum_{t=1}^{T} [F_t(x_t) - F_{t+1}(x_{t+1}) + f_t(x_t)] \\
&\le \sum_{t=1}^{T} \left[\frac{\alpha_{t-1}\|g_t\|^2}{2\mu} + \psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1})\right] \\
&\le \sum_{t=1}^{T} \frac{\alpha_{t-1}\|g_t\|^2}{2\mu} \quad (\text{since } \alpha_t \le \alpha_{t+1} \Rightarrow \psi_t \ge \psi_{t+1}).
\end{aligned}$$

Also, since $x_{T+1} = \arg\min_{x\in\mathcal{K}} F_{T+1}(x)$, we have

$$F_{T+1}(x_{T+1}) \le F_{T+1}(u).$$

Together with Lemma 15, this gives

$$\begin{aligned}
\mathrm{Reg}_T(u) &= \sum_{t=1}^{T} (f_t(x_t) - f_t(u)) \\
&\le \psi_{T+1}(u) - \min_{x\in\mathcal{K}} \psi_1(x) + \sum_{t=1}^{T} [F_t(x_t) - F_{t+1}(x_{t+1}) + f_t(x_t)] + F_{T+1}(x_{T+1}) - F_{T+1}(u) \\
&\le \psi_T(u) + \frac{1}{2\mu}\sum_{t=1}^{T} \alpha_{t-1}\|g_t\|^2.
\end{aligned}$$

■

Corollary 18 (Restatement of Theorem 7) *Suppose that Assumptions 1, 2, and 3 hold. Let p be an arbitrary point in $\mathcal{K}$. Then, R-FTRL defined in 5 with regularizer $\psi_t(x) = \frac{L\sqrt{t}}{2D} d^2(x,p)$ achieves the following regret bound:*

$$\mathrm{Reg}_T \le \frac{3}{2} DL\sqrt{T}.$$

Proof Since ψ is 1-strongly convex from Proposition 12, By applying Lemma 17, we derive the following upper bound on the regret:

$$\begin{aligned}
\mathrm{Reg}_T(u) &\le \psi_T(u) + \frac{1}{2\mu}\sum_{t=1}^{T} \alpha_{t-1}\|g_t\|^2 \\
&\le \frac{\frac{d(u,p)^2}{2}}{\frac{D}{L\sqrt{T}}} + \frac{1}{2}\sum_{t=1}^{T} \frac{D}{L\sqrt{t}} L^2 \\
&\le \frac{DL\sqrt{T}}{2} + \frac{1}{2} DL \sum_{t=1}^{T} \frac{1}{\sqrt{t}} \\
&\le \frac{3}{2} DL\sqrt{T}.
\end{aligned}$$

Here, we used the fact that $\sum_{t=1}^{T} \frac{1}{\sqrt{t}} \le 2\sqrt{T}$. ■

Corollary 19 (Restatement of Theorem 11) *Suppose that Assumptions 1, 2, and 3 hold. We also assume that f_t is μ-strongly g-convex for all $t \in [T]$. Then, R-FTRL defined in 5 with $\psi_t(x) = 0$ achieves the following regret bound:*

$$\mathrm{Reg}_T \le \frac{L^2}{2\mu}(1 + \log(T)).$$

Proof Noting that $F_t + f_t$ is a (μt)-strongly convex function, from Lemmas 15 and 16, we have

$$\begin{aligned}\mathrm{Reg}_T(u) &\le \sum_{t=1}^{T} \frac{1}{2\mu t}\|g_t\|^2 \\ &\le \frac{L^2}{2\mu}\sum_{t=1}^{T}\frac{1}{t} \\ &\le \frac{L^2}{2\mu}(1+\log T).\end{aligned}$$

■

These results demonstrate the effectiveness of FTRL on manifolds.