Towards Characterizing the Complexity of Riemannian Online Convex Optimization

Hibiki Fukushima
The University of Tokyo and RIKEN

Hiroshi Hirai
Nagoya University

Shinji Ito
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 predetermined step-size schedules and for certain types of adaptive step-size schedules. 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. We also analyze a Riemannian extension of Follow-the-Regularized-Leader, which we term R-FTRL, in the full-information setting. R-FTRL achieves a regret bound of $O(DL\sqrt{T})$, independent of the curvature. This complements recent curvature-independent guarantees for full-information methods obtained by different algorithmic approaches. Together with our lower bound for R-OGD, our results support a separation between first-order and full-information models in non-Euclidean settings, and highlight the subtle interactions between feedback structure, algorithm design, and geometry.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

1 INTRODUCTION

The theory and methodology of Online Convex Optimization (OCO) (Hazan et al., 2016; Orabona, 2019) have developed significantly, both through their diverse applications—such as sequential prediction (Vovk, 2001), model ensemble (boosting) (Littlestone and Warmuth, 1994; Freund and Schapire, 1997), sequential portfolio selection (Cover, 1991; Hazan et al., 2006)—and through their role as a mathematical foundation in fields like online learning theory (Cesa-Bianchi and Lugosi, 2006), algorithmic game theory (Roughgarden, 2010), and reinforcement learning (Lattimore and Szepesvári, 2020). In OCO problems, a player interacts with an adversary in a sequential manner, where the player selects a point x_t from a compact convex set $\mathcal{K} \subseteq \mathbb{R}^d$ and incurs a loss based on the adversary's choice of convex function $f_t : \mathcal{K} \to \mathbb{R}$. A central object of interest in OCO is a performance metric known as regret Reg_T, which measures the difference between the player's cumulative loss and the best possible loss that could have been achieved by selecting the best point u^* in hindsight:

$$\mathrm{Reg}_T(u) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(u)$$

$$\mathrm{Reg}_T = \mathrm{Reg}_T(u^*) \qquad \left(u^* \in \arg\min_{u\in\mathcal{K}} \sum_{t=1}^{T} f_t(u)\right).$$

One of the most notable algorithms for this problem is Online Gradient Descent (OGD) (Zinkevich, 2003), 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 $\Omega(DL\sqrt{T})$ (Hazan et al., 2016).

Recent research (Wang et al., 2023; Hu et al., 2023; Wang et al., 2025; Chen and Sun, 2024) 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. This line of work is motivated by the growing number of applications in which the feasible domain is inherently non-Euclidean, such as optimization over probability simplices, positive definite matrices, or orthonormal matrices, where modeling the underlying geometric constraints is essential. Such extensions are expected to provide both theoretical insights and practical benefits, including better exploitation of problem structure, improved algorithmic stability, and broader applicability to tasks in machine learning, signal processing, and statistical inference.

A key geometric factor distinguishing these non-Euclidean domains from standard Euclidean space is the sectional curvature. When optimizing over spaces with nonpositive curvature, the geometry exhibits exponential volume growth as the curvature becomes more negative. Consequently, local gradient information evaluated in the tangent space becomes a significantly less reliable reflection of the global objective landscape. This geometric distortion naturally raises a fundamental question:

Research Question:

How does the fundamental difficulty of online convex optimization problems change when moving from Euclidean spaces to manifolds? In particular, which geometric properties of manifolds govern or characterize the complexity of such problems?

The objective of this work is to offer a refined and rigorous answer to this fundamental question. In particular, we focus on how this difficulty differs between first-order and full-information methods.

To date, our understanding of this question has remained somewhat limited. In prior work by Wang et al. (2023), a natural extension of Online Gradient Descent (OGD) from Euclidean spaces to Riemannian manifolds, referred to as Riemannian OGD (R-OGD), was proposed. It was shown that this algorithm achieves a regret upper bound of $O(DL\sqrt{\zeta T})$, where $\zeta \geq 1$ is a parameter defined in (1) that depends on κ and D. Notably, ζ is the only term in the upper bound that accounts for the gap between the manifold setting and the Euclidean case. On the other hand, they also established a regret lower bound of $\Omega(DL\sqrt{T})$, which does not involve the parameter ζ. Note that this lower bound is universal in the sense that it holds for all Hadamard manifolds. This implies that there exists no Hadamard manifold for which the intrinsic difficulty of the problem becomes fundamentally easier than in the Euclidean case, as the regret cannot improve beyond $\Omega(DL\sqrt{T})$. In contrast, whether such curvature dependence is inherent for natural first-order methods on manifolds remained unclear, since a gap by a factor of $\sqrt{\zeta}$ remained between the known upper and lower bounds for R-OGD, even though curvature-independent regret is known to be achievable in the full-information setting via different algorithmic approaches (Roux et al., 2025).

Related questions also arise in the offline setting of optimization over manifolds. The research of first-order methods of geodesically convex optimization on non-Euclidean manifolds started with a seminal work of Zhang and Sra (2016). Their analysis revealed that the convergence rate of gradient descent is intimately connected to both the sectional curvature of the manifold and the diameter of the feasible domain. In particular, they demonstrated that choosing the step size as $\alpha_t = \frac{D}{L\sqrt{\zeta t}}$ yields an upper bound of $\mathcal{O}(DL\sqrt{\frac{\zeta}{T}})$ on the optimality gap. Subsequent studies have investigated both upper and lower complexity bounds, as well as the potential for algorithmic acceleration in this non-Euclidean context(Criscitiello and Boumal, 2022; Criscitiello et al., 2023; Hamilton and Moitra, 2021). For example, Criscitiello and Boumal (2023) showed that in the case of gradient descent with Polyak step size, the curvature-diameter constant ζ also appears in the *lower bounds* of the iteration complexity, suggesting that this dependency may be inherent to the geometry of the manifold. Nevertheless, for gradient descent with the step size $\alpha_t = \frac{D}{L\sqrt{\zeta t}}$ as proposed by Zhang and Sra (2016), no matching lower bounds have yet been established. 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

In this work, as in the aforementioned study by Wang et al. (2023), we primarily consider online optimization problems with geodesically convex loss functions over Hadamard manifolds. We separately analyze two feedback models: the *first-order feedback* model, in which only a subgradient of the loss function at x_t is available, and the *full-information feedback* model, in which the learner has access to the entire loss function at each round. It is worth noting that, by definition, the full-information model provides strictly more feedback than the first-order model and can therefore be regarded as a strictly easier setting. In the classical Euclidean OCO setting, however, the minimax regret is $\Theta(DL\sqrt{T})$ under both models, and hence the distinction between the two is often omitted in the literature.

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), 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 and for certain types of adaptive step-size schedules. Combined with the previously established upper bound of $O(DL\sqrt{\zeta T})$ (Wang et al., 2023), this result implies that the worst-case regret of R-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. Indeed, we show that the first-order setting can be viewed as a subclass of a broader family of online learning problems that extends beyond convex optimization. Our new lower bound for R-OGD naturally leads to the following question: *How do the regret bounds change when we consider algorithms beyond OGD?* Under the full-information feedback model, curvature-independent regret bounds are achievable, and in this paper we analyze an FTRL-based algorithm with this property.

Our second contribution is to analyze a Riemannian extension of 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. We show that R-FTRL achieves a regret bound of $O(DL\sqrt{T})$, independent of the curvature. This provides an FTRL-based full-information algorithm with curvature-independent regret, complementing recent results obtained by different algorithmic approaches. Together with our lower bound for R-OGD, this result supports a qualitative separation between first-order and full-information methods in the Riemannian setting.

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:

- In the Euclidean case, OGD is minimax optimal, whereas R-OGD is not minimax optimal on manifolds.
- In Euclidean spaces, OGD and FTRL exhibit the same (optimal) performance, while in the manifold setting, our results highlight a performance gap of a factor of $\sqrt{\zeta}$ between R-OGD and FTRL-based methods.
- 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, these results support a gap between these two feedback models.

1.2 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 (2016) 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. (2023) 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 (Sahinoglu and Shahrampour, 2025).

More recently, Roux et al. (2025) studied online learning on Hadamard manifolds using an implicit optimistic update scheme. Their results show that curvature-independent regret bounds are achievable without dependence on geometric constants such as the curvature

Table 1: Comparison of upper and lower bounds for Riemannian online convex optimization. Regret bounds in gray boxes are new results in this paper. Note that the $\Omega(DL\sqrt{\zeta T})$ lower bound shown in the bottom-right applies exclusively to the specific algorithm R-OGD.

Feasible set	Euclidean	Riemannian	
Feedback type	full-information/first-order	full-information	first-order
Upper bounds	$O(DL\sqrt{T})$: OGD(Zinkevich, 2003), FTRL(Hazan and Kale, 2010)	$O(DL\sqrt{T})$: RIOD(Roux et al., 2025) $O(DL\sqrt{T})$: R-FTRL	$O(DL\sqrt{\zeta T})$: R-OGD(Wang et al., 2023)
Lower bounds	$\Omega(DL\sqrt{T})$(Orabona and Pál, 2018)	$\Omega(DL\sqrt{T})$(Wang et al., 2023)	$\Omega(DL\sqrt{T})$(Wang et al., 2023) $\Omega(DL\sqrt{\zeta T})$: R-OGD

parameter, matching the best known Euclidean rates up to constants. Their framework also allows for more refined optimistic guarantees and inexact updates. In contrast, our full-information result is based on an FTRL formulation, and our primary new contribution is the lower bound for R-OGD, which helps clarify the gap between natural first-order methods and full-information methods in the Riemannian setting.

Meanwhile, Antonakopoulos et al. (2020) 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 (2023) conducted a systematic investigation into the role of curvature in the complexity of offline optimization on manifolds. In particular, they established a general lower bound for all deterministic first-order algorithms, though this bound leaves a gap with the known upper bounds. They also showed tight lower bounds for gradient descent with the Polyak step size, but this result does not cover gradient descent with standard predetermined step sizes such as those proposed by Zhang and Sra (2016). 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 Shu et al. (2025); Martínez-Rubio et al. (2024).

In this work, we address some of these gaps by adapting analytical ideas from Criscitiello and Boumal (2023) to the online optimization setting. Our contributions are distinct from, and in some aspects complementary to, both Criscitiello and Boumal (2023) and Roux et al. (2025). First, while the lower bounds of Criscitiello and Boumal (2023) 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. (2023), which is the same step size in Zhang and Sra (2016). Furthermore, our results apply to all R-OGD algorithms with predetermined step size sequences, thereby covering a broader class of methods.

2 PRELIMINARIES

A manifold $\mathcal{M}$ of dimension n is a topological space that is locally homeomorphic to the n-dimensional Euclidean space $\mathbb{R}^n$. At each point $x \in \mathcal{M}$, the tangent space $T_x\mathcal{M}$ is the vector space of all possible tangent vectors. A Riemannian manifold is a manifold equipped with a Riemannian metric g, which is a smoothly varying inner product on the tangent space. Given the inner product $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$ over $T_x\mathcal{M}$, we denote $\langle \xi, \eta \rangle = g_x(\xi, \eta)$ and $\|\xi\| = \sqrt{g_x(\xi, \xi)}$ for $\xi, \eta \in T_x\mathcal{M}$. For tangent vectors $\xi, \eta \in T_x\mathcal{M}$, the angle θ between them is defined by $\cos(\theta) = \frac{\langle \xi, \eta \rangle}{\|\xi\|\|\eta\|}$. For linearly independent vectors ξ, η, the sectional curvature $K(\xi, \eta)$ is defined by $K(\xi, \eta) = \frac{\langle R(\xi,\eta)\eta,\xi \rangle}{\langle \xi,\xi \rangle \langle \eta,\eta \rangle - \langle \xi,\eta \rangle^2}$, where R is the Riemann curvature tensor. The length of a smooth curve $\gamma : [a, b] \to \mathcal{M}$ is defined as $L(\gamma) = \int_a^b \sqrt{g_{\gamma(t)}\left(\dot{\gamma}(t), \dot{\gamma}(t)\right)}\, dt$. The distance between two points $x, y \in \mathcal{M}$ is defined as $d(x, y) = \inf\{L(\gamma) \mid \gamma$ is a smooth curve connecting x and $y\}$. For simplicity, we also use the following notation: $d^2(x, y) = (d(x, y))^2$. A geodesic is a curve that locally minimizes the length. We assume that the manifold $\mathcal{M}$ is (geodesically) complete. Then the exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ sends any tangent vector to a point on the manifold. More precisely, for $\xi \in T_x\mathcal{M}$, we define $\exp_x(\xi) = \gamma(1)$, where $\gamma : [0, 1] \to \mathcal{M}$ is a geodesic with $\gamma(0) = x$ 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 (uniquely g-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

$$f(\gamma(t)) \le (1 - t)f(\gamma(0)) + t f(\gamma(1)) \quad (\forall t \in [0, 1]).$$

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

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)) - \frac{\mu}{2}t(1-t)d^2(\gamma(0), \gamma(1))$$

for all $t \in [0, 1]$. 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 for any $x, y \in \mathcal{K}$, the following inequality holds:

$$|f(x) - f(y)| \leq L \cdot d(x, y) \quad (\forall x, y \in \mathcal{K}).$$

A simply connected complete manifold having non-positive sectional curvature everywhere is called a Hadamard manifold. A Hadamard manifold is uniquely geodesically convex (uniquely g-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, and the squared distance function $x \mapsto \frac{1}{2}d^2(x, y)$ is 1-strongly geodesically convex ((Lee, 2018, Lemma 12.15) and (Bačák, 2014, Remark 2.2.2)). 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 defined on a Hadamard manifold, the subdifferential at each point is non-empty (see Greene and Shiohama (1981, p. 139) and Udriste (1994, Chapter 3, Theorem 4.5)).

3 PROBLEM SETTINGS

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:

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

In this paper, as in the previous work by Wang et al. (2023), 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 $\kappa \leq 0$.*

4 ANALYSIS OF RIEMANNIAN ONLINE GRADIENT DESCENT

Riemannian Online Gradient Descent (R-OGD) (Wang et al., 2023) is a natural extension of OGD (Zinkevich, 2003) 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) (Wang et al., 2023)

Require: A feasible set $\mathcal{K} \subseteq \mathcal{M}$, time horizon T, a sequence (or a schedule) of step sizes $\{\alpha_t\}_{t=1}^T$, and an initial point $x_1 \in \mathcal{K}$.

1: **for** $t = 1$ to T **do**
2: Play x_t and observe a subgradient $g_t \in \partial f_t(x_t)$ of f_t at x_t.
3: Update x_{t+1} with

$$\tilde{x}_{t+1} = \exp_{x_t}(-\alpha_t g_t), \quad x_{t+1} = \mathcal{P}_\mathcal{K}(\tilde{x}_{t+1}),$$

where $\mathcal{P}_\mathcal{K}$ is the Riemannian projection mapping of x onto $\mathcal{K}$, i.e., $\mathcal{P}_\mathcal{K}(x) \in \arg\min_{y\in\mathcal{K}} d(x, y)$.
4: **end for**

4.1 Review of Existing Results: Regret Upper Bound for R-OGD

Wang et al. (2023) showed that R-OGD achieves the following regret upper bound:

Theorem 1 (Theorem 5 of Wang et al. (2023))**.** *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* $\mathrm{Reg}_T \le \frac{3}{2}DL\sqrt{\zeta T}$ *for any T and any initial point* $x_1 \in \mathcal{K}$*, where* $\zeta = \zeta(\kappa, D) \ge 1$ *is defined as*

$$\zeta(\kappa, D) = \begin{cases} \frac{\sqrt{|\kappa|}D}{\tanh\left(\sqrt{|\kappa|}D\right)} & \text{if } \kappa < 0, \\ 1 & \text{if } \kappa \ge 0. \end{cases} \tag{1}$$

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

$$f_t(x_t) - f_t(u) \le \left\langle -g_t, \exp_{x_t}^{-1}(u)\right\rangle = \tilde{f}_t(x_t) - \tilde{f}_t(u), \tag{2}$$

where $\tilde{f}_t(x) = \left\langle g_t, \exp_{x_t}^{-1}(x)\right\rangle$. We note that this expression essentially corresponds to the first-order characterization of convexity in the Euclidean case, namely the following inequality:

$$f_t(x_t) - f_t(u) \le \langle g_t, x_t - u\rangle = \tilde{f}_t(x_t) - \tilde{f}_t(u), \tag{3}$$

where $\tilde{f}_t(x) = \langle g_t, x - x_t\rangle$. 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. What proves to be useful at this point is the following lemma:

Lemma 1 (Special case of Lemma 5 of Zhang and Sra (2016))**.** *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*

$$a^2 \le \zeta(\kappa, c)b^2 + c^2 - 2bc\cos A,$$

where $\zeta(\kappa, c) \ge 1$ is defined in (1)*.*

Using this lemma, the regret can be upper bounded as follows, following a line of reasoning similar to that in the Euclidean case:

$$\begin{aligned} &\mathrm{Reg}_T(u) \\ &= \sum_{t=1}^{T} \left(f_t(x_t) - f_t(u)\right) \le \sum_{t=1}^{T} \left(\tilde{f}_t(x_t) - \tilde{f}_t(u)\right) \text{ (by (2))} \\ &\le \sum_{t=1}^{T} \Bigg(\frac{1}{2\alpha_t}\left(d^2(x_t, u) - d^2(\tilde{x}_{t+1}, u)\right) \\ &+ \frac{\alpha_t}{2}\zeta(\kappa, d(x_t, u))\|g_t\|^2 \Bigg) \\ &\le \sum_{t=1}^{T} \left(\frac{1}{2\alpha_t}\left(d^2(x_t, u) - d^2(x_{t+1}, u)\right) + \frac{\alpha_t}{2}\zeta(\kappa, D)L^2 \right), \end{aligned} \tag{4}$$

where the second inequality follows from Lemma 1 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 (Bačák, 2014, 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 1 follows by applying this inequality (4) together with the specification of step sizes α_t in the theorem and the bound $d(x_t, u) \le D$.

Furthermore, a similar line of analysis yields the following regret upper bound in the case of strong convexity.

Theorem 2 (Theorem 6 of Wang et al. (2023))**.** *Suppose that Assumptions 1, 2, and 3 hold, and that f_t is μ-strongly g-convex for any $t = 1, \ldots, T$. Then, Algorithm 1 (R-OGD) with step sizes* $\left\{\alpha_t = \frac{1}{\mu t}\right\}$ *achieves* $\mathrm{Reg}_T \le \frac{L^2\zeta}{2\mu}(1 + \log T)$ *for any T and any initial point* $x_1 \in \mathcal{K}$*, where* $\zeta = \zeta(\kappa, D) \ge 1$ *is defined in* (1)*.*

4.2 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 1 is tight, as long as one uses R-OGD.

Theorem 3. *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 follows:*

$$\mathrm{Reg}_T = \Omega\left(DL\sqrt{\zeta T}\right),$$

where ζ is defined in (1)*.*

Together with Theorem 1, this result provides a $\Theta(DL\sqrt{\zeta T})$ characterization of the worst-case regret

[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., Innami (1982).

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.

To establish the lower bound in Theorem 3, it suffices to identify a problem instance for which each inequality used in the proof of Theorem 1, e.g., inequalities in the proof sketch (4), holds with a constant-factor gap. The first critical step concerns the gap in the first inequality in (4) of this chain, which can be analyzed through the following lemma:

Lemma 2. *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 (2023). All results established in this work but stated without proof in the main text, including this lemma, are proved in the Appendix.

Lemma 2 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 (4)). 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. This leads to the following proposition:

Proposition 1 (Lower bound for first-order algorithm)**.** *Fix an arbitrary $\mathcal{M}$ satisfying Assumption 3 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:*

$$\sup_{\{f_t\}_{t=1}^T \in \mathcal{F}} \max_{u \in \mathcal{K}} \sum_{t=1}^{T} (f_t(x_t) - f_t(u)) \geq \sup_{\{g_t\}_{t=1}^T \in \mathcal{G}} \max_{u \in \mathcal{K}} \sum_{t=1}^{T} \left(\tilde{f}_t(x_t) - \tilde{f}_t(u)\right), \tag{5}$$

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 once again the distinction between the Euclidean and Riemannian settings. In the Euclidean case, when the loss functions are convex, the surrogate loss $\tilde{f}_t$ appearing on the right-hand side of (5) can also 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 (5) is *not necessarily g-convex*. As a result, the problem of 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.

Thus far, we have shown that the first inequality in (4) can be made tight. It remains to show that the remaining inequalities can also be tightened up to constant factors. Specifically, we consider the inequality corresponding to Lemma 1, as well as the gap between $\zeta(\kappa, d(x_t, u))$ and $\zeta(\kappa, D)$, and identify conditions under which this gap becomes negligible (i.e., within a constant multiplicative factor).

To this end, we construct a specific instance in which the feasible set $\mathcal{K}$ is the unit ball in a hyperbolic space, and the gradients g_t are chosen such that the iterates x_t move along a fixed geodesic. The magnitudes of g_t are adapted according to the step sizes α_t, which allows us to control the position of x_t. We then select u at a constant-factor distance (relative to the radius) from the origin to induce a tight lower bound instance. We refer the reader to the Appendix for full details of the construction.

It is worth noting that our construction methodology differs from the approach of Criscitiello and Boumal (2023) for establishing lower bounds in the offline optimization setting. Their approach relies on a high-dimensional hyperbolic space ($d \geq T$), where the objective function is designed such that its gradient continually points in a new orthogonal direction. As a result, their analysis compels the algorithm to explore increasingly higher dimensions and critically depends on this property, together with the structure of the Polyak step size.

In contrast, as discussed above, our worst-case instance only requires a fixed two-dimensional hyperbolic space, which suffices to restrict the iterates to a single geodesic. This shows that the intrinsic difficulty of the problem arises directly from the geometry of negative curvature, rather than from an artificially inflated state space dimension.

A key ingredient enabling this construction is Proposition 1. Since deterministic first-order algorithms in R-OCO must contend with linear approximations ($\tilde{f}_t$) that are generally not g-convex, we can instead construct an instance using alternating linear functions. This forces the algorithm to incur large regret without requiring any orthogonal expansion. Furthermore, the emergence of the curvature-dependent factor $\sqrt{\zeta}$ can be explicitly understood through the geometry of hyperbolic right triangles. Consider the geodesic triangle formed by the current iterate x_t, the origin, and the comparator u. In hyperbolic space, even when the iterate approaches the origin, the angle at x_t increases more slowly than in Euclidean space due to negative curvature. By exploiting this geometric effect, our choice of gradients along a fixed geodesic precisely captures the $\sqrt{\zeta}$ dependence, providing a clear geometric explanation for the degraded performance of R-OGD in curved spaces.

4.3 Regret Bounds for R-OGD with Adaptive Stepsize

The preceding analysis assumed a predetermined step-size schedule, which requires knowledge of parameters such as the time horizon T and the Lipschitz constant L. A natural alternative is to use an adaptive step size that does not depend on such prior knowledge. Here, we show that an AdaGrad-style step size (Auer et al., 2002; McMahan and Streeter, 2010; Duchi et al., 2011) achieves a similar regret bound, which is also tight.

Theorem 4. *Suppose that Assumptions 1, 2, and 3 hold. Algorithm 1 (R-OGD) with step sizes* $\left\{\alpha_t = \frac{D}{\sqrt{\zeta \sum_{i=1}^{t} \|g_i\|^2}}\right\}$ *achieves*

$$\mathrm{Reg}_T \le \frac{3}{2} D \sqrt{\zeta \sum_{t=1}^{T} \|g_t\|^2} \le \frac{3}{2} DL\sqrt{\zeta T}$$

for any T and any initial point $x_1 \in \mathcal{K}$.

This upper bound matches the one in Theorem 1 up to constant factors, but with the advantage of not requiring T and L in advance. Furthermore, we can show that this bound is tight for this class of adaptive algorithms.

Theorem 5. *Consider Algorithm 1 (R-OGD) with step sizes* $\left\{\alpha_t = \frac{\alpha}{\sqrt{\sum_{i=1}^{t} \|g_i\|^2}}\right\}$. *For any $\alpha > 0$, there exists $\mathcal{K} \subseteq \mathcal{M}$ and a sequence of loss functions $\{f_t\}_{t=1}^T$ satisfying Assumptions 1, 2, and 3, such that the regret of the algorithm is bounded as follows:*

$$\mathrm{Reg}_T = \Omega(DL\sqrt{\zeta T}).$$

Taken together, these results demonstrate that using an adaptive step size also leads to a $\Theta(DL\sqrt{\zeta T})$ characterization of the worst-case regret. This reinforces the conclusion that the curvature-dependent factor ζ is an intrinsic element governing the performance of R-OGD on Hadamard manifolds, irrespective of whether the step-size schedule is fixed or adaptive.

5 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 (Hazan et al., 2016; Orabona, 2019). 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 surrogate losses $\tilde{f}_t$ given in (3), which are convex approximations of the original losses. Moreover, if the regularizer is strictly convex, 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, as presented in Algorithm 2. 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 yields curvature-independent regret bounds, as shown in the following theorems:

Algorithm 2 Riemannian Follow the Regularized Leader Algorithm (R-FTRL)

Require: A sequence of regularizers $\psi_1, \dots, \psi_T : \mathcal{K} \to \mathbb{R}$.
1: **for** $t = 1$ to T **do**
2: Output $x_t \in \arg\min_{x \in \mathcal{K}} \psi_t(x) + \sum_{s=1}^{t-1} f_s(x)$.
3: Receive $f_t : \mathcal{K} \to \mathbb{R}$.
4: **end for**

Theorem 6. *Suppose that Assumptions 1, 2, and 3 hold. Let p be an arbitrary point in $\mathcal{K}$. Then, Algo-*

rithm 2 (R-FTRL) with regularizer $\psi_t(x) = \frac{L\sqrt{t}}{2D} d^2(x, p)$ *achieves the following regret bound:*

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

Theorem 7. *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, Algorithm 2 (R-FTRL) with* $\psi_t(x) = 0$ *achieves the following regret bound:*

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

Theorems 6 and 7 show that R-FTRL achieves curvature-independent regret bounds under both general and strongly convex losses. In particular, the bound in Theorem 6 matches the Euclidean $O(DL\sqrt{T})$ rate and does not depend on the curvature parameter ζ. This shows 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 an important contrast 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. In contrast, in the Riemannian setting, our results indicate that this equivalence does not extend to natural first-order methods: while R-OGD suffers from curvature-dependent regret, R-FTRL achieves curvature-independent performance. Together with existing full-information results, this supports a qualitative separation in the effectiveness of first-order and full-information algorithms under non-Euclidean geometry.

6 CONCLUSION AND DISCUSSION

This work explores how the geometry of Riemannian manifolds influences the complexity of online convex optimization, focusing on regret under both first-order and full-information feedback models. Our findings suggest fundamental differences between Euclidean and Riemannian settings and contribute toward a deeper understanding of optimization in non-Euclidean spaces.

More specifically, our analysis uncovers structural differences between Euclidean and Riemannian OCO. In Euclidean spaces, first-order and full-information feedback yield the same minimax regret, and the surrogate loss $\tilde{f}_t$ remains convex, thus preserving the problem's convexity structure. In contrast, on Hadamard manifolds, the surrogate loss $\tilde{f}_t$ in (2) is not necessarily g-convex, and hence cannot be reduced to a g-convex full-information problem. Proposition 1 further indicates that, in the first-order setting, there exist worst-case instances where the regret with respect to such non-g-convex surrogate losses cannot be avoided. This suggests that the two feedback models can exhibit qualitatively different behavior in the Riemannian setting.

Based on these observations, we conjecture that an $\Omega(\sqrt{\zeta T})$ lower bound holds for any first-order algorithm, not just R-OGD. A promising route toward proving this would be to construct subgradient sequences $\{g_t\}$ that make all inequalities in the regret decomposition (e.g., (4)) tight, as we did for R-OGD. However, our construction heavily relies on the specific update dynamics of R-OGD, and generalizing it to arbitrary algorithms appears to be a nontrivial task.

In summary, our results highlight several distinctive features of online learning over Riemannian manifolds and open several directions for future research, including a better understanding of feedback-dependent complexity and the development of practical algorithms that combine geometric adaptivity with computational efficiency.

Acknowledgments

Hiroshi Hirai was supported by JSPS KAKENHI Grant Number JP24K21315. Shinji Ito was supported by JSPS KAKENHI Grant Number JP25K03184 and by JST PRESTO, Japan, Grant Number JPMJPR2511.

References

Antonakopoulos, K., Belmega, E. V., and Mertikopoulos, P. (2020). Online and stochastic optimization beyond lipschitz continuity: A Riemannian approach. In *ICLR 2020-International Conference on Learning Representations*, pages 1–20.

Auer, P., Cesa-Bianchi, N., and Gentile, C. (2002). Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75.

Bačák, M. (2014). *Convex analysis and optimization in Hadamard spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. De Gruyter, Berlin.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press.

Chen, H. and Sun, Q. (2024). Decentralized online Riemannian optimization with dynamic environments. *arXiv preprint arXiv:2410.05128*.

Cover, T. M. (1991). Universal portfolios. *Mathematical finance*, 1(1):1–29.

Criscitiello, C. and Boumal, N. (2022). Negative curvature obstructs acceleration for strongly geodesically convex optimization, even with exact first-order oracles. In *Conference on Learning Theory, 2-5 July*

2022, London, UK, volume 178 of *Proceedings of Machine Learning Research*, pages 496–542. PMLR.

Criscitiello, C. and Boumal, N. (2023). Curvature and complexity: Better lower bounds for geodesically convex optimization. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2969–3013. PMLR.

Criscitiello, C., Martínez-Rubio, D., and Boumal, N. (2023). Open problem: Polynomial linearly-convergent method for g-convex optimization? In *The Thirty Sixth Annual Conference on Learning Theory, COLT 2023, 12-15 July 2023, Bangalore, India*, volume 195 of *Proceedings of Machine Learning Research*, pages 5950–5956. PMLR.

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).

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

Greene, R. E. and Shiohama, K. (1981). Convex functions on complete noncompact manifolds: topological structure. *Invent. Math.*, 63(1):129–157.

Hamilton, L. and Moitra, A. (2021). No-go theorem for acceleration in the hyperbolic plane.

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

Hazan, E., Kalai, A., Kale, S., and Agarwal, A. (2006). Logarithmic regret algorithms for online convex optimization. In *International Conference on Computational Learning Theory*, pages 499–513. Springer.

Hazan, E. and Kale, S. (2010). Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine Learning*, 80:165–188.

Hu, Z., Wang, G., and Abernethy, J. D. (2023). Riemannian projection-free online learning. *Advances in Neural Information Processing Systems*, 36:41980–42014.

Innami, N. (1982). Splitting theorems of Riemannian manifolds. *Compositio Math.*, 47(3):237–247.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Lee, J. M. (2018). *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, second edition.

Littlestone, N. and Warmuth, M. K. (1994). The weighted majority algorithm. *Information and computation*, 108(2):212–261.

Martínez-Rubio, D., Roux, C., and Pokutta, S. (2024). 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.

McMahan, H. B. and Streeter, M. (2010). Adaptive bound optimization for online convex optimization. In *COLT*.

Orabona, F. (2019). A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*.

Orabona, F. and Pál, D. (2018). Scale-free online learning. *Theoretical Computer Science*, 716:50–69.

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

Roux, C., Martínez-Rubio, D., and Pokutta, S. (2025). Implicit Riemannian optimism with applications to min-max problems. In *International Conference on Machine Learning*, pages 52139–52172. PMLR.

Sahinoglu, E. and Shahrampour, S. (2025). Online optimization on hadamard manifolds: Curvature independent regret bounds on horospherically convex objectives. *arXiv preprint arXiv:2509.11236*.

Sakai, T. (1996). *Riemannian geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the 1992 Japanese original by the author.

Shu, Y., Jiang, J., Shi, L., and Wang, T. (2025). Revisit gradient descent for geodesically convex optimization.

Udriste, C. (1994). *Convex functions and optimization methods on Riemannian manifolds*, volume 367. Springer Science & Business Media.

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

Wang, X., Tu, Z., Hong, Y., Wu, Y., and Shi, G. (2023). Online optimization over Riemannian manifolds. *Journal of Machine Learning Research*, 24(84):1–67.

Wang, X., Yuan, D., Hong, Y., Hu, Z., Wang, L., and Shi, G. (2025). Riemannian online optimistic algorithms with dynamic regret. *IEEE Transactions on Automatic Control*.

Zhang, H. and Sra, S. (2016). First-order methods for geodesically convex optimization. In *Conference on Learning Theory*, pages 1617–1638. PMLR.

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

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] Justification: The mathematical setting and assumptions are provided in Sections 2 and 3. The R-OGD and R-FTRL algorithms are presented in Algorithm 1 (Section 4) and Algorithm 2 (Section 5), respectively.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] Justification: The regret complexity of the presented algorithms is the central topic of our analysis, with main results stated in Theorems 3, 4, 5, 6, and 7.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable] Justification: This is a purely theoretical work and does not include source code.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes] Justification: Assumptions 1, 2, and 3, which apply to our theoretical results, are explicitly stated.
   (b) Complete proofs of all theoretical results. [Yes] Justification: Proofs for all theorems, propositions, and lemmas are provided in the Appendix.
   (c) Clear explanations of any assumptions. [Yes] Justification: The assumptions are standard and are clearly stated in Section 3.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable] Justification: This paper is theoretical and does not include empirical results.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable] Justification: This paper is theoretical and does not include empirical results.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable] Justification: This paper is theoretical and does not include empirical results.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable] Justification: This paper is theoretical and does not include empirical results.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable] Justification: Our work does not use any existing assets, nor do we release any new assets.
   (b) The license information of the assets, if applicable. [Not Applicable] Justification: Our work does not use any existing assets, nor do we release any new assets.
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable] Justification: Our work does not use any existing assets, nor do we release any new assets.
   (d) Information about consent from data providers/curators. [Not Applicable] Justification: Our work does not use any existing assets, nor do we release any new assets.
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable] Justification: Our work does not use any existing assets, nor do we release any new assets.

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable] Justification: Our research did not involve human subjects or crowdsourcing.
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable] Justification: Our research did not involve human subjects or crowdsourcing.
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable] Justification: Our research did not involve human subjects or crowdsourcing.

Towards Characterizing the Complexity of Riemannian Online Convex Optimization: Supplementary Materials

A PROOF OF PROPOSITION 1

This section presents the proof of Proposition 1. We start by proving Lemma 2

Proof of Lemma 2. 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

$$g_{\perp} \in \partial f_{\perp}(x). \tag{6}$$

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 (6) from

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

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 (Sakai, 1996, 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 (Bačák, 2014, 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 1.

Proof of Proposition 1. 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 2 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:

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

☐

B PROOF OF THEOREM 3

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:

$$\cosh(\sqrt{|\kappa|}c) = \cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b), \tag{7}$$

$$\cosh(\sqrt{|\kappa|}b) = \cosh(\sqrt{|\kappa|}c)\cosh(\sqrt{|\kappa|}a) - \sinh(\sqrt{|\kappa|}c)\sinh(\sqrt{|\kappa|}a)\cos(\theta). \tag{8}$$

From (7), we have

$$c = \frac{1}{\sqrt{|\kappa|}}\cosh^{-1}(\cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)) \geq \frac{1}{\sqrt{|\kappa|}}\cosh^{-1}(\cosh(\sqrt{|\kappa|}b)) = b. \tag{9}$$

From (8),

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

where the second and fifth equalities follow from (7). Because $\frac{1}{\sinh(\cosh^{-1}(x))} \geq \frac{1}{x}$ for $x \geq 1$, we have

$$\begin{aligned}\cos(\theta) &= \frac{\sinh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)}{\sinh(\cosh^{-1}(\cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)))} \\ &\geq \frac{\sinh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)}{\cosh(\sqrt{|\kappa|}a)\cosh(\sqrt{|\kappa|}b)} \\ &= \tanh(\sqrt{|\kappa|}a). \qquad (10)\end{aligned}$$

Lemma 3. *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 (9) and (10) 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 (9), it holds that $d(x,u) = c \geq b = D/2$. Combining this result with the bound from (10), we can establish a lower bound for the inner product:

$$\langle -g, \exp_x^{-1}(u)\rangle = d(x,u)\|g\|\cos(\theta) \geq \frac{D}{2}\frac{L}{\sqrt{2}}\tanh\left(\sqrt{|\kappa|}r\right). \qquad (11)$$

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

$$\frac{DL}{2\sqrt{2}}\tanh\left(\sqrt{|\kappa|}r\right) \geq \frac{DL}{2\sqrt{2}}\tanh\left(\frac{\sqrt{|\kappa|}D}{2\sqrt{2\zeta T}}\right). \qquad (12)$$

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

$$\frac{DL}{2\sqrt{2}}\tanh\left(\frac{\sqrt{|\kappa|}D}{2\sqrt{2\zeta T}}\right) \geq \frac{DL}{2\sqrt{2}}\tanh\left(\frac{1}{4}\sqrt{\frac{\zeta}{2T}}\right). \qquad (13)$$

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:

$$\frac{DL}{2\sqrt{2}}\tanh\left(\frac{1}{4}\sqrt{\frac{\zeta}{2T}}\right) \geq \frac{DL}{2\sqrt{2}}\cdot\frac{1}{2}\left(\frac{1}{4}\sqrt{\frac{\zeta}{2T}}\right) = \frac{DL\sqrt{\zeta}}{32\sqrt{T}} = \frac{DL}{32}\sqrt{\frac{\zeta}{T}}. \qquad (14)$$

Combining inequalities (11) through (14), we obtain:

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

□

We are now ready to prove Theorem 3.

Proof of Theorem 3. 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 1. 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 (10), 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

$$\left|\left\{t \in I \;\middle|\; |s_t| \geq \frac{D}{2\sqrt{2\zeta T}}\right\}\right| \geq \frac{|I|}{2}. \tag{15}$$

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 (15) holds.

Using Lemma 3, we obtain that for any $t \in I$ with $|s_t| \geq \frac{D}{2\sqrt{2\zeta T}}$, it holds that

$$\langle -g_t, \exp_{x_t}^{-1}(u)\rangle \geq \frac{DL\sqrt{\frac{\zeta}{T}}}{32}. \tag{16}$$

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

$$\mathrm{Reg}_T(u) \geq \sum_{t=1}^{T} \langle -g_t, \exp_{x_t}^{-1}(u)\rangle. \tag{17}$$

Combining (15), the bound of (16) 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 $J = \{1, \dots, T\} \setminus I$, 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

$$s_{t_j} < \frac{D(j-1)}{T} - \frac{D}{2}. \tag{18}$$

Therefore, for all $t = 1, 2, \dots, T$, 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

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

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 (18) 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 $\Omega(DL\sqrt{\zeta T})$. □

C PROOF OF THEOREMS 6 AND 7

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 2. *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 Udriste (1994) 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,

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

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

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

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 4. *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 5. *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 6 and 7. 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 6. *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 7. *Assume F_t is g-convex, and $F_t + f_t$ is λ_t-strongly g-convex. Then, for $g_t \in \partial f_t(x_t)$,*

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

Proof. Since $x_t \in \arg\min_{x\in\mathcal{K}} F_t(x)$, Lemma 5 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 4 with λ_t-strongly g-convexity of $F_t + f_t$, we have

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

□

Lemma 8. *Let $\psi : \mathcal{K} \to \mathbb{R}$ be μ-strongly g-convex. Let α_t be a sequence of positive numbers that satisfies $\alpha_t \leq \alpha_{t+1}$. Define $\psi_t : \mathcal{K} \to \mathbb{R}_{\geq 0}$ by*

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

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

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

Proof. From Lemma 7, we have

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

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

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

Together with Lemma 6, this gives

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

□

Corollary 1 (Restatement of Theorem 6)**.** *Suppose that Assumptions 1, 2, and 3 hold. Let p be an arbitrary point in $\mathcal{K}$. Then, Algorithm 2 (R-FTRL) with regularizer $\psi_t(x) = \frac{L\sqrt{t}}{2D} d^2(x, p)$ achieves the following regret bound:*

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

Proof. Since ψ is 1-strongly g-convex by Proposition 2, by applying Lemma 8, we derive the following upper bound on the regret:

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

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

Corollary 2 (Restatement of Theorem 7)**.** *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, Algorithm 2 (R-FTRL) with $\psi_t(x) = 0$ achieves the following regret bound:*

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

Proof. Noting that $F_t + f_t$ is a (μt)-strongly convex function, from Lemmas 6 and 7, we have

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

☐

These results demonstrate the effectiveness of FTRL on manifolds.

D PROOF OF THEOREMS 4 and 5

Proof of Theorem 4. From inequality (4), we have the initial bound on the regret:

$$\begin{aligned}
\mathrm{Reg}_T &\leq \sum_{t=1}^{T}\frac{1}{2\alpha_t}\left(d^2(x_t,u) - d^2(x_{t+1},u)\right) + \sum_{t=1}^{T}\frac{\alpha_t}{2}\zeta\|g_t\|^2 \\
&= \sum_{t=2}^{T}\left(\frac{1}{2\alpha_t} - \frac{1}{2\alpha_{t-1}}\right)d^2(x_t,u) + \frac{1}{2\alpha_1}d^2(x_1,u) - \frac{1}{2\alpha_T}d^2(x_{T+1},u) + \sum_{t=1}^{T}\frac{\alpha_t}{2}\zeta\|g_t\|^2 \\
&\leq \sum_{t=2}^{T}\left(\frac{1}{2\alpha_t} - \frac{1}{2\alpha_{t-1}}\right)d^2(x_t,u) + \frac{1}{2\alpha_1}d^2(x_1,u) + \sum_{t=1}^{T}\frac{\alpha_t}{2}\zeta\|g_t\|^2.
\end{aligned}$$

Assuming $\alpha_t \leq \alpha_{t-1}$ for all t, the term $\left(\frac{1}{2\alpha_t} - \frac{1}{2\alpha_{t-1}}\right)$ is non-negative. We can then use the bound $d(x_t, u) \leq D$ to get:

$$\begin{aligned}
\mathrm{Reg}_T &\leq \sum_{t=2}^{T} \left(\frac{1}{2\alpha_t} - \frac{1}{2\alpha_{t-1}}\right) D^2 + \frac{1}{2\alpha_1} D^2 + \sum_{t=1}^{T} \frac{\alpha_t}{2} \zeta \|g_t\|^2 \\
&= \frac{1}{2\alpha_T} D^2 + \sum_{t=1}^{T} \frac{\alpha_t}{2} \zeta \|g_t\|^2.
\end{aligned}$$

Now, we set the step size $\alpha_t = \frac{D}{\sqrt{\zeta \sum_{i=1}^{t} \|g_i\|^2}}$. Note that this choice of α_t satisfies $\alpha_t \leq \alpha_{t-1}$ for all t. Substituting this into our regret bound yields:

$$\begin{aligned}
\mathrm{Reg}_T &\leq \frac{1}{2\alpha_T} D^2 + \sum_{t=1}^{T} \frac{\alpha_t}{2} \zeta \|g_t\|^2 \\
&= \frac{\sqrt{\zeta \sum_{i=1}^{T} \|g_i\|^2}}{2D} D^2 + \sum_{t=1}^{T} \frac{\zeta}{2} \frac{D\|g_t\|^2}{\sqrt{\zeta \sum_{i=1}^{t} \|g_i\|^2}} \\
&= \frac{1}{2} D \sqrt{\zeta \sum_{i=1}^{T} \|g_i\|^2} + \frac{D\sqrt{\zeta}}{2} \sum_{t=1}^{T} \frac{\|g_t\|^2}{\sqrt{\sum_{i=1}^{t} \|g_i\|^2}}.
\end{aligned}$$

To bound the summation term, we use the following inequality for a sequence of non-negative numbers a_t (see, e.g., Duchi et al. (2011), Appendix C, Eq. 24):

$$\sum_{t=1}^{T} \frac{a_t}{\sqrt{\sum_{i=1}^{t} a_i}} \leq 2\sqrt{\sum_{i=1}^{T} a_i}.$$

This gives:

$$\begin{aligned}
\mathrm{Reg}_T &\leq \frac{1}{2} D \sqrt{\zeta \sum_{i=1}^{T} \|g_i\|^2} + \frac{D\sqrt{\zeta}}{2} \left(2\sqrt{\sum_{i=1}^{T} \|g_i\|^2}\right) \\
&= \frac{3}{2} D \sqrt{\zeta \sum_{i=1}^{T} \|g_i\|^2}.
\end{aligned}$$

Finally, using the assumption that $\|g_t\| \leq L$ for all t, we have $\sum_{i=1}^{T} \|g_i\|^2 \leq TL^2$, which leads to the final bound:

$$\mathrm{Reg}_T \leq \frac{3}{2} D \sqrt{\zeta \sum_{i=1}^{T} \|g_i\|^2} \leq \frac{3}{2} DL\sqrt{\zeta T}.$$

□

Proof of Theorem 5. Let $\mathcal{M} = \mathbb{H}^2_\kappa$ be the two-dimensional hyperbolic space with curvature κ. Let $p \in \mathbb{H}^2_\kappa$ and take an orthonormal basis e_1, e_2 of $T_p\mathbb{H}^2_\kappa$. We perform a case analysis based on α.
Case 1: $\alpha \geq \frac{D}{\sqrt{2\zeta}}$.
In this case, we define the g-convex set $\mathcal{K}$ as a closed ball of radius $D/2$ centered at p:

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

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

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

In this case, the step sizes α_t satisfy

$$\begin{aligned} \alpha_t &= \frac{\alpha}{\sqrt{\sum_{i=1}^t \|g_i\|^2}} \\ &= \frac{\alpha}{\sqrt{\sum_{i=1}^t \frac{L^2}{2}}} \\ &= \frac{\alpha}{L\sqrt{\frac{t}{2}}} \\ &\geq \frac{D}{L\sqrt{\zeta T}}. \end{aligned}$$

We next show that

$$\left| \left\{ t \in \{1, 2, \ldots, T\} \;\middle|\; |s_t| \geq \frac{D}{2\sqrt{2\zeta T}} \right\} \right| \geq \frac{T}{2}. \tag{19}$$

For any t, 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, at least one of t or $t+1$ satisfies $|s_t| \geq \frac{D}{2\sqrt{2\zeta T}}$. Therefore, equation (19) holds.

Using Lemma 3, we obtain that for any $t \in I$ with $|s_t| \geq \frac{D}{2\sqrt{2\zeta T}}$, it holds that

$$\langle -g_t, \exp_{x_t}^{-1}(u) \rangle \geq \frac{DL\sqrt{\frac{\zeta}{T}}}{32}. \tag{20}$$

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

$$\mathrm{Reg}_T(u) \geq \sum_{t=1}^{T} \langle -g_t, \exp_{x_t}^{-1}(u) \rangle. \tag{21}$$

Combining (19), the bound of (20) 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

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

Case 2: $\alpha < \frac{D}{\sqrt{2\zeta}}$.
Assume $\zeta \leq T$. Let γ be a geodesic in $\mathbb{H}_\kappa^2$ satisfying the initial conditions $\gamma(0) = p$ and $\dot{\gamma}(0) = e_1$. We consider the g-convex geodesic set

$$\mathcal{K} = \left\{ x \in \mathbb{H}_\kappa^2 \mid x = \gamma(s), 0 \leq s \leq D \right\},$$

with endpoints $u = \gamma(D)$ and $x_1 = p = \gamma(0)$. For any point $x_t = \gamma(s_t)$ on this segment, we define the vector g_t as

$$g_t = \begin{cases} \frac{L}{\sqrt{2}} e_2 & \text{for } t = 1, 2, \ldots, \frac{T}{2}, \\ -\frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} \dot{\gamma}(s_t) & \text{for } t = \frac{T}{2} + 1, \ldots, T. \end{cases}$$

In this case, $s_t = 0$ for all $t = 1, 2, \ldots, \frac{T}{2} + 1$, because $\mathcal{P}_{\mathcal{K}}\left(\exp_p(ae_2)\right) = p$, where $\mathcal{P}_{\mathcal{K}}$ is the Riemannian projection mapping of x onto $\mathcal{K}$. Also, we have

$$\langle -g_t, \exp_{x_t}^{-1}(u) \rangle = 0 \quad \text{for all } t = 1, 2, \ldots, \frac{T}{2}, \tag{22}$$

because e_1 and e_2 are orthogonal.
Next, for $t = \frac{T}{2} + 1, \frac{T}{2} + 2, \ldots, T$, the step sizes α_t satisfy

$$\begin{aligned} \alpha_t &= \frac{\alpha}{\sqrt{\sum_{i=1}^{t} \|g_i\|^2}} \\ &\leq \frac{\alpha}{\sqrt{\sum_{i=1}^{\frac{T}{2}} \frac{L^2}{2}}} \\ &= \frac{\alpha}{L\sqrt{\frac{T}{4}}}. \end{aligned}$$

Because $\alpha < \frac{D}{\sqrt{2\zeta}}$, it follows that $\alpha_t < \frac{D}{L}\sqrt{\frac{2}{\zeta T}}$. Using this bound for α_t, and recalling that $\|g_t\| = \frac{L}{\sqrt{2}}\sqrt{\frac{\zeta}{T}}$ for this range of t, the product $\alpha_t \|g_t\|$ is bounded by

$$\begin{aligned} \alpha_t \|g_t\| &\leq \frac{D}{L}\sqrt{\frac{2}{\zeta T}} \frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} \\ &= \frac{D}{T}. \end{aligned}$$

Therefore, for $t = \frac{T}{2} + 2, \frac{T}{2} + 3, \ldots, T$,

$$\begin{aligned} s_t &= \sum_{i=\frac{T}{2}+1}^{t-1} \alpha_i \|g_i\| \\ &\leq \left(t - \frac{T}{2} - 1\right) \frac{D}{T} \end{aligned} \tag{23}$$

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

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

We can now establish a lower bound for the regret $\mathrm{Reg}_T(u)$. From (22), the terms for $t = 1, 2, \ldots, \frac{T}{2}$ are zero. Thus, the summation simplifies to

$$\begin{aligned}\mathrm{Reg}_T(u) &\geq \sum_{t=\frac{T}{2}+1}^{T} \langle -g_t, \exp_{x_t}^{-1}(u) \rangle \\ &= \sum_{t=\frac{T}{2}+1}^{T} \frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} \left(D - s_t\right).\end{aligned}$$

We apply the upper bound on s_t derived in (23). This gives a lower bound on the regret:

$$\begin{aligned}\mathrm{Reg}_T(u) &\geq \sum_{t=\frac{T}{2}+1}^{T} \frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} \left(D - \left(t - \frac{T}{2} - 1\right) \frac{D}{T}\right) \\ &= \sum_{t=\frac{T}{2}+1}^{T} \frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} \left(D \frac{\frac{3}{2}T - t + 1}{T}\right).\end{aligned}$$

Evaluating the summation over the $T/2$ terms gives

$$\begin{aligned}\mathrm{Reg}_T(u) &\geq \sum_{t=\frac{T}{2}+1}^{T} \frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} \left(D \frac{\frac{3}{2}T - t + 1}{T}\right) \\ &= \frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} \left(D \frac{3T^2 + 2T}{8T}\right) \\ &\geq \frac{L}{\sqrt{2}} \sqrt{\frac{\zeta}{T}} D \frac{3T^2}{8T} \\ &= \frac{3}{8\sqrt{2}} DL\sqrt{\zeta T}.\end{aligned}$$

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