Non-stationary Bandit Convex Optimization: A Comprehensive Study

Xiaoqi Liu[*]   Dorian Baudry[*†]   Julian Zimmert[‡]   Patrick Rebeschini[*]   Arya Akhavan[*§]

Abstract

Bandit Convex Optimization is a fundamental class of sequential decision-making problems, where the learner selects actions from a continuous domain and observes a loss (but not its gradient) at only one point per round. We study this problem in non-stationary environments, and aim to minimize the regret under three standard measures of non-stationarity: the number of switches S in the comparator sequence, the total variation Δ of the loss functions, and the path-length P of the comparator sequence. We propose a polynomial-time algorithm, Tilted Exponentially Weighted Average with Sleeping Experts (TEWA-SE), which adapts the sleeping experts framework from online convex optimization to the bandit setting. For strongly convex losses, we prove that TEWA-SE is minimax-optimal with respect to known S and Δ by establishing matching upper and lower bounds. By equipping TEWA-SE with the Bandit-over-Bandit framework, we extend our analysis to environments with unknown non-stationarity measures. For general convex losses, we introduce a second algorithm, clipped Exploration by Optimization (cExO), based on exponential weights over a discretized action space. While not polynomial-time computable, this method achieves minimax-optimal regret with respect to known S and Δ, and improves on the best existing bounds with respect to P.

1 Introduction

Many real-world decision-making problems, such as resource allocation, experimental design, or hyperparameter tuning require repeatedly selecting an action from a continuous space under uncertainty and limited feedback. These settings are naturally modeled as Bandit Convex Optimization (see [1] for an introduction), in which an adversary fixes a sequence of T loss functions $f_1, f_2, \ldots, f_T : \mathbb{R}^d \to \mathbb{R}$ beforehand, and a learner sequentially interacts with the adversary for T rounds. At each round t, the learner selects an action $\boldsymbol{z}_t$ from a continuous arm set $\Theta \subseteq \mathbb{R}^d$, assumed to be convex and compact. The learner then incurs a loss $f_t(\boldsymbol{z}_t)$ and observes a noisy feedback:

$$y_t = f_t(\boldsymbol{z}_t) + \xi_t \,, \tag{1}$$

where ξ_t is a sub-Gaussian noise variable (Definition 1). The goal is to minimize the learner's *regret* with respect to (w.r.t.) a performance benchmark. In the online learning literature [2, 3], the benchmark is typically the best *static* action in hindsight, with cumulative loss $\min_{\boldsymbol{z} \in \Theta} \sum_{t=1}^{T} f_t(\boldsymbol{z})$.

However, non-stationarity arises in many applications where different actions may work well during different time intervals. Hence, a line of works [4–8] propose to compare the learner's actions against

---

[*]University of Oxford. Correspondence to: `{shirley.liu, arya.akhavan}@stats.ox.ac.uk`.

[†]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France.

[‡]Google Research.

[§]École Polytechnique de Paris, IP Paris.

a sequence of comparators $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T \in \Theta$, leading to a regret defined as

$$R(T, \boldsymbol{u}_{1:T}) \coloneqq \sum_{t=1}^{T} \mathbf{E}\left[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u}_t)\right], \tag{2}$$

where $\boldsymbol{u}_{1:T}$ denotes $(\boldsymbol{u}_t)_{t=1}^{T}$, and the expectation is taken w.r.t. the randomness in the learner's actions $\boldsymbol{z}_t$'s and the randomness of the noise variables ξ_t's, similarly to the standard notion of pseudo-regret in the bandit literature, see e.g., [9, Section 4.8]. Choosing the regret-maximizing comparator in (2) gives rise to the notion of *dynamic regret*, defined as

$$R^{\mathsf{dyn}}(T) \coloneqq \max_{\boldsymbol{u}_{1:T} \in \Theta^T} R(T, \boldsymbol{u}_{1:T}). \tag{3}$$

While addressing non-stationarity through dynamic regret has been extensively studied in multi-armed bandits (e.g., [10–12]), it remains relatively underexplored in continuum bandits [7, 13, 14]. This work aims to bridge this gap by proposing algorithms for Bandit Convex Optimization that achieve sublinear dynamic regret. Such a rate is generally unattainable without imposing structural constraints on the environment, i.e., the comparator sequence and the loss function sequence [6]. For the comparator sequence, two commonly studied constraints are the number of switches [15] and the path-length [4], defined respectively as

$$S(\boldsymbol{u}_{1:T}) \coloneqq 1 + \sum_{t=2}^{T} \mathbf{1}\left(\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\right) \leq S, \qquad P(\boldsymbol{u}_{1:T}) \coloneqq \sum_{t=2}^{T} \|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\| \leq P. \tag{4}$$

For the loss function sequence, a popular constraint is the total variation [7], defined as

$$\Delta(f_{1:T}) \coloneqq \sum_{t=2}^{T} \max_{\boldsymbol{z} \in \Theta} |f_t(\boldsymbol{z}) - f_{t-1}(\boldsymbol{z})| \leq \Delta. \tag{5}$$

The constraints that the upper bounds S, P and Δ respectively impose on the comparators or on the loss functions lead to different notions of regret. We call the regret for environments constrained by S the *switching regret*, which we define as

$$R^{\mathsf{swi}}(T, S) \coloneqq \max_{\boldsymbol{u}_{1:T}: S(\boldsymbol{u}_{1:T}) \leq S} R(T, \boldsymbol{u}_{1:T}). \tag{6}$$

Similarly, we call the regret for environments constrained by P the *path-length regret*, denoted by $R^{\mathsf{path}}(T, P)$. We also use $R^{\mathsf{dyn}}(T, \Delta)$ and $R^{\mathsf{dyn}}(T, \Delta, S)$ to denote the dynamic regret, where the arguments after T specify environment constraints. See (8) for rigorous definitions.

We detail in Section 1.3 conversion results between the different regret definitions that we introduced: a sublinear switching regret $R^{\mathsf{swi}}(T, S)$ implies sublinear $R^{\mathsf{dyn}}(T, \Delta)$, $R^{\mathsf{dyn}}(T, \Delta, S)$ and $R^{\mathsf{path}}(T, P)$, as illustrated in Figure 1 (see also [16, 17]). Furthermore, the upper bounds on the switching regret presented in this work are derived from upper bounds on the *adaptive regret* [18, 19], which is defined using an interval length $\mathsf{B} \in [T]$ as follows,

$$R^{\mathsf{ada}}(\mathsf{B}, T) \coloneqq \max_{\substack{p,q \in [T], \\ 0 < q-p \leq \mathsf{B}}} \max_{\boldsymbol{u} \in \Theta} \sum_{t=p}^{q} \mathbf{E}\left[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})\right]. \tag{7}$$

With an appropriate tuning of B depending on S, an adaptive regret sublinear in B implies a switching regret sublinear in T through a simple reduction, see e.g., discussions in [19]. We illustrate the relations between these regret notions in Figure 1.
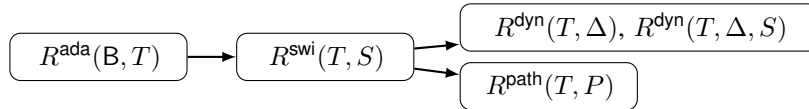

Figure 1: Conversions between regrets: $R_1 \longrightarrow R_2$ means that if regret R_1 is sublinear in T (or B), then regret R_2 is also sublinear in T, see Proposition 1 for precise mathematical statements.

We conclude this section by detailing the main notation and assumptions used throughout the paper.

Notation For $k \in \mathbb{N}^+$, we denote by $[k]$ the set of positive integers $\leq k$. We denote by $\|\cdot\|$ the Euclidean norm, $\mathbb{B}^d = \{u \in \mathbb{R}^d : \|u\| \leq 1\}$ the unit Euclidean ball, and $\Pi_\Theta(x) = \arg\min_{w \in \Theta} \|w - x\|$ the Euclidean projection of x to Θ. We use $a \vee b \equiv \max(a, b)$ and $a \wedge b \equiv \min(a, b)$. If A, B depend on T, we use $A = \mathcal{O}(B)$ (resp. $A = \Omega(B)$) when there exists $c > 0$ s.t. $A \leq cB$ (resp. $\geq$) with c independent of T, d, S, Δ and P. To hide polylogarithmic factors in T, we use interchangeably $A = \widetilde{\mathcal{O}}(B)$ and $A \lesssim B$ (resp. $A = \widetilde{\Omega}(B)$ and $A \gtrsim B$), e.g., $A \leq T \log T \implies A = \widetilde{\mathcal{O}}(T)$. Moreover, $A = o(B)$ means $A/B \to 0$ as $T \to \infty$.

Assumptions For simplicity, we assume that the time horizon T is known in advance; the case of unknown T can be handled using the standard doubling trick [20]. For some $\sigma > 0$, the noise variables $(\xi_t)_{t=1}^T$ are σ-sub-Gaussian. For all $t \in [T]$, $\max_{x \in \Theta} |f_t(x)| \leq 1$. We consider two cases: (i) general convex losses f_t, where we assume Lipschitz continuity with constant K, and (ii) the special strongly-convex case, where we assume β-smoothness. The domain Θ is assumed to contain a ball of radius r for some constant $r > 0$, and has a bounded diameter $\operatorname{diam}(\Theta) := \sup\{\|x - w\| : x, w \in \Theta\} \leq D$ for some constant $D > 0$.

1.1 Main contributions

Existing works on non-stationary Bandit Convex Optimization study different aspects of the problem in isolation: [7, 14] focus on dynamic regret $R^{\mathsf{dyn}}(T, \Delta)$, while [13, 21] address path-length regret $R^{\mathsf{path}}(T, P)$. The present work aims to systematically unify and extend previously scattered results, establishing a complete picture of the state-of-the-art regret bounds w.r.t. all three non-stationarity measures S, Δ and P.

Our first contribution is a polynomial-time algorithm called Tilted Exponentially Weighted Average with Sleeping Experts (TEWA-SE), which we design by adapting a series of works from online convex optimization [22–24] to the bandit setting with zeroth-order feedback. It addresses the absence of gradient information by employing the randomized perturbation technique from [25, 26] to estimate gradients, combined with the design of quadratic surrogate loss functions depending on a uniform upper bound on the norm of the gradient estimates.

Following [22–24], TEWA-SE runs multiple expert algorithms with different learning rates in parallel, and combines them using a tilted exponentially weighted average. This allows TEWA-SE to adapt to the curvature of the loss function f_t's without prior knowledge of parameters such as the strong-convexity parameter. For a given interval length B, an appropriately tuned TEWA-SE simultaneously achieves an adaptive regret of the order $\sqrt{d}\mathsf{B}^{\frac{3}{4}}$ for general convex losses and $d\sqrt{\mathsf{B}}$ for strongly-convex losses (Theorem 1). Consequently, for a known S, we prove that an optimal tuning of TEWA-SE yields a switching regret upper bound of order $\sqrt{d}S^{\frac{1}{4}}T^{\frac{3}{4}}$ for general convex losses (Corollary 1). In the same result, we further prove that if the losses are strongly convex, and that Δ is known and incorporated in the tuning of TEWA-SE, the algorithm simultaneously satisfies a $\min\{d\sqrt{ST}, d^{\frac{2}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}}\}$ dynamic regret bound. Importantly, TEWA-SE does *not* need to know the actual strong-convexity parameter, inheriting the adaptivity properties of the framework developed in [22–24]. We prove that this dynamic regret upper bound is minimax-optimal in T, d, S and Δ by establishing a matching lower bound (Theorem 2). Finally, still for strongly-convex losses, we prove that TEWA-SE can also achieve a path-length regret of the order $d^{\frac{2}{3}}P^{\frac{1}{3}}T^{\frac{2}{3}}$ when P is known. We summarize these results in Table 1. To overcome the restriction of knowing S, Δ and P to optimally tune TEWA-SE, we also analyze a variant equipped with the Bandit-over-Bandit framework [27].

Table 1: Regret bounds we obtain for $R^{\mathsf{swi}}(T, S)$, $R^{\mathsf{dyn}}(T, \Delta)$ and $R^{\mathsf{path}}(T, P)$, respectively, for algorithms tuned with known S, Δ and P (polylogarithmic factors omitted). Straight underlines indicate minimax-optimal rates. A wavy underline indicates the result is either new to the literature (strongly-convex case) or improves on the best-known $P^{\frac{1}{4}}T^{\frac{3}{4}}$ rate [13] (general convex case).

	TEWA-SE (Algorithm 1)	cExO (Algorithm 3)
Convex	$\sqrt{d}S^{\frac{1}{4}}T^{\frac{3}{4}},\ d^{\frac{2}{5}}\Delta^{\frac{1}{5}}T^{\frac{4}{5}},\ d^{\frac{2}{5}}P^{\frac{1}{5}}T^{\frac{4}{5}}$	$d^{\frac{5}{2}}\sqrt{ST},\ d^{\frac{5}{3}}\underline{\Delta^{\frac{1}{3}}T^{\frac{2}{3}}}, d^{\frac{5}{3}}\underwave{P^{\frac{1}{3}}T^{\frac{2}{3}}}$
Strongly convex	$\underline{d\sqrt{ST}},\ \underline{d^{\frac{2}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}}},\ \underwave{d^{\frac{2}{3}}P^{\frac{1}{3}}T^{\frac{2}{3}}}$	

For general convex losses with known S, Δ and P, TEWA-SE achieves a suboptimal $T^{\frac{3}{4}}$ rate (Corollary 1), matching the rates in similar analysis for the static regret [25, 26]. Thus, the second contribution of this work is the clipped Exploration by Optimization (cExO) algorithm with improved guarantees for this setting, which uses exponential weights on a discretized action space Θ with clipping [28]. For a given interval length B, this algorithm with an optimally tuned learning rate w.r.t. B attains an order $d^{\frac{5}{2}}\sqrt{B}$ adaptive regret (Theorem 3). When S, Δ and P are known beforehand, this algorithm with an optimally tuned learning rate achieves the minimax-optimal dynamic regret w.r.t. S and Δ simultaneously, and attains a $P^{\frac{1}{3}}T^{\frac{2}{3}}$ path-length regret (Corollary 2), improving on the previous best $P^{\frac{1}{4}}T^{\frac{3}{4}}$ [13]. While this algorithm is not polynomial-time computable and has suboptimal rates w.r.t. the problem dimension d, it provides insights that may guide future research toward developing efficient algorithms with optimal guarantees for the convex case.

1.2 Related work

The literature on *Bandit Convex Optimization* (BCO) has traditionally focused on minimizing the static regret, see the recent monograph [1] for a comprehensive historical overview. Both convex and strongly convex objective functions have attracted significant attention, beginning with the foundational work of [25] and further developed in subsequent studies such as [29–36]. Minimizing regret in non-stationary environments has only received attention more recently [7, 13, 14, 21], see also [1, Section 2.4] for an overview for this topic. Among these works, [7, 14] study $R^{\mathsf{dyn}}(T, \Delta)$, whereas [13, 21] focus on $R^{\mathsf{path}}(T, P)$. As we explained above (and formalize in Section 1.3), the switching regret $R^{\mathsf{swi}}(T, S)$ can induce guarantees on both $R^{\mathsf{dyn}}(T, \Delta)$ and $R^{\mathsf{path}}(T, P)$, but the reverse does not necessarily hold. Therefore, the results in these works cannot be readily extended to provide regret guarantees w.r.t. all three measures S, Δ and P.

Minimizing regret in environments with non-stationarity measures such as S, Δ and P have been addressed with greater depth in *Online Convex Optimization* (OCO), where the learner has direct access to gradient information and can query the gradient or function value at multiple points of the loss function per round. The state-of-the-art algorithm with optimal adaptive regret guarantees is MetaGrad with sleeping experts [24], which queries only one gradient per round, and adapts to curvature information of the loss function such as strong-convexity when available. Our polynomial-time algorithm TEWA-SE builds upon [24] and its precursors [22, 23], adapting this approach to BCO by replacing the exact gradient per round with an approximate gradient estimate, and by designing a quadratic surrogate loss. The approach in [24] follows a long line of successive developments in OCO from expert tracking methods [15, 20, 37–41] to the study of adaptive regret [18, 19, 42–46], with recent advances [24, 47–50] reducing the query complexity from $\mathcal{O}(\log T)$ to $\mathcal{O}(1)$ per round, while achieving optimal adaptive regret or dynamic regret. The adaptivity of [24] directly inherits from MetaGrad [22] and its extension [23], which themselves build on earlier adaptive methods [51, 52].

For general convex functions, the approach of substituting a one-point gradient estimate for the exact gradient in each round of an OCO algorithm often yields suboptimal $T^{\frac{3}{4}}$ rates, both in static regret [25, 26] and dynamic regret [13, 21]; see also our Corollary 1. A series of breakthroughs [28, 32, 53–56] indicate that $\sqrt{T}$ rates (up to logarithms) are attainable for static regret, at the cost of a higher dependency on d. Our cExO algorithm follows this line of work, using exponential weights on a discretized action space [28]. By playing inside a clipped domain, we transform the algorithm from one with $\sqrt{T}$ static regret into one with $\sqrt{B}$ adaptive regret (modulo logarithms) for intervals of length $\leq B$, which in turn leads to regret guarantees w.r.t. S, Δ and P.

Finally, we mention that non-stationarity has been widely studied in the *Multi-Armed Bandit* (MAB) literature. A substantial body of work has focused on adapting standard policies—such as UCB [57, 58], EXP3 [59], and Thompson Sampling [60–62]—to perform effectively under non-stationarity. These adaptations often employ mechanisms to discard outdated information, either *actively* (e.g., change-detection methods [12, 63–67]), or *passively* (e.g., discounted rewards [10, 68], sliding windows [69, 70], or scheduled restarts [11]), but are not straightforward to adapt to BCO.

1.3 Conversions between different regret definitions

We present the key conversions between different regret notions, illustrated in Figure 1 above. Using the definition of $R^{\mathsf{dyn}}(T)$ in (3), we overload notation slightly to define

$$R^{\mathsf{dyn}}(T, \Delta) \coloneqq \sup_{f_{1:T}:\Delta(f_{1:T})\leq\Delta} \sum_{t=1}^{T} \mathbf{E}\left[f_t(\boldsymbol{z}_t) - \min_{\boldsymbol{z}\in\Theta} f_t(\boldsymbol{z})\right], \tag{8}$$

and $R^{\mathsf{dyn}}(T, \Delta, S)$ additionally constrains $1 + \sum_{t=2}^{T} \min_{(\boldsymbol{z}_t^*, \boldsymbol{z}_{t-1}^*)\in(\mathcal{Z}_t^*, \mathcal{Z}_{t-1}^*)} \mathbf{1}(\boldsymbol{z}_t^* \neq \boldsymbol{z}_{t-1}^*) \leq S$ where $\mathcal{Z}_t^* \coloneqq \arg\min_{\boldsymbol{z}\in\Theta} f_t(\boldsymbol{z})$ for all $t \in [T]$. In Proposition 1, we show how the adaptive regret $R^{\mathsf{ada}}(\mathsf{B}, T)$ can be used to bound the switching regret $R^{\mathsf{swi}}(T, S)$, which in turn can be used to bound the dynamic regret $R^{\mathsf{dyn}}(T, \Delta, S)$ and path-length regret $R^{\mathsf{path}}(T, P)$. Consequently, $R^{\mathsf{ada}}(\mathsf{B}, T)$ and $R^{\mathsf{swi}}(T, S)$ are the primary objects to analyze.

Proposition 1. *Suppose that an algorithm can be calibrated to satisfy $R^{\mathsf{ada}}(\mathsf{B}, T) \leq C\mathsf{B}^{\kappa}$, for any interval length $\mathsf{B} \in [T]$, for some factor $C > 0$ that is at most polynomial in d and $\log(T)$, and $\kappa \in [0, 1)$.*

Then, for any $S, S_\Delta, S_P \in [T]$, an appropriate choice of B yields the following regret guarantees:

Switching: $\mathsf{B} = \left\lceil \frac{T}{S} \right\rceil$ *guarantees that $R^{\mathsf{swi}}(T, S) \leq 2^{1+\kappa} C S^{1-\kappa} T^{\kappa}$.*

Dynamic: $\mathsf{B} = \left\lceil \frac{T}{S} \right\rceil \vee \left\lceil \frac{T}{S_\Delta} \right\rceil$ *yields $R^{\mathsf{dyn}}(T, \Delta, S) \leq R^{\mathsf{swi}}(T, S) \wedge \left(R^{\mathsf{swi}}(T, S_\Delta) + \Delta \left\lceil \frac{T}{S_\Delta} \right\rceil \right)$.*

Path-length: $\mathsf{B} = \left\lceil \frac{T}{S_P} \right\rceil$ *ensures that $R^{\mathsf{path}}(T, P) \leq R^{\mathsf{swi}}(T, S_P) + \frac{P}{r} \cdot \left\lceil \frac{T}{S_P} \right\rceil$.*

The proof is provided in Appendix B. We note that the reduction from $R^{\mathsf{path}}(T, P)$ to $R^{\mathsf{swi}}(T, S)$ in Proposition 1 is new and employs simple geometric arguments (see Lemma 2 in Appendix B). This reduction simplifies the analysis of $R^{\mathsf{path}}(T, P)$, though it can yield slightly looser bounds on $R^{\mathsf{path}}(T, P)$ than a direct analysis, as discussed in [17].

2 The TEWA-SE algorithm

In this section, we develop a polynomial-time algorithm called Tilted Exponentially Weighted Average with Sleeping Experts (TEWA-SE, Algorithm 1), building on the two-layer structure of previous experts-based algorithms [18, 19, 43]. Each expert in TEWA-SE is uniquely defined by its lifetime and learning rate. We denote the active experts at time t by $E_1, E_2, \ldots, E_{n_t}$, where E_i operates over interval I_i with learning rate η_i. In each round t, the active experts each propose an action, denoted by $\boldsymbol{x}_{t,I_i}^{\eta_i}$, and a meta-algorithm aggregates them into a single meta-action $\boldsymbol{x}_t$ by computing their tilted exponentially weighted average [22, 24], see line 7 in the pseudo-code. Then the algorithm receives a noisy evaluation of f_t at $\boldsymbol{x}_t$ and constructs an approximate *gradient estimate* $\boldsymbol{g}_t \in \mathbb{R}^d$ of f_t at $\boldsymbol{x}_t$. Both $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$ are shared with all experts, who update their actions via online gradient descent on their *surrogate* loss functions defined using $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$.

TEWA-SE employs the Geometric Covering scheme from [19, 24] to schedule experts across different time intervals, and the exponential grid from [22, 24] to assign varied learning rates to the experts. These deterministic schemes ensure that only a *logarithmic* number of experts are active per round, maintaining computational efficiency. Intuitively, the meta-algorithm achieves low adaptive regret on the original loss function because, for each subinterval of times, there exists at least one individual expert with low static regret on their surrogate loss functions on this subinterval. This is guaranteed by the careful design of the exponential grid of learning rates. While full details of TEWA-SE is deferred to Appendix C.1, we highlight below the distinctions between this paper and prior works.

Construction of one-point gradient estimate For a fixed parameter $h \in (0, r)$, we define the clipped domain $\tilde{\Theta} = \{\boldsymbol{u} \in \Theta : \boldsymbol{u} + h\mathbb{B}^d \subset \Theta\}$, where $h < r$ ensures $\tilde{\Theta} \neq \emptyset$. In each round t, we select a meta-action $\boldsymbol{x}_t \in \tilde{\Theta}$ and query the function at a perturbed point $\boldsymbol{x}_t + h\boldsymbol{\zeta}_t$, receiving noisy feedback $y_t = f_t(\boldsymbol{x}_t + h\boldsymbol{\zeta}_t) + \xi_t$, where $\boldsymbol{\zeta}_t \in \mathbb{R}^d$ is sampled uniformly from the unit sphere $\partial\mathbb{B}^d$. This allows us to construct the gradient estimate $\boldsymbol{g}_t = (d/h)y_t\boldsymbol{\zeta}_t$. As implied by [25, Lemma 1], the

---

Algorithm 1 Tilted Exponentially Weighted Average with Sleeping Experts (TEWA-SE)

Input: $d, T, \mathrm{B}, h = \min\left(\sqrt{d}\mathrm{B}^{-\frac{1}{4}}, r\right), \tilde{\Theta} = \{\boldsymbol{u} \in \Theta : \boldsymbol{u} + h\mathbb{B}^d \subset \Theta\}, G$ as in (10), expert algorithm $E(I, \eta)$ defined in Algorithm 2, and $(n_t)_{t \in [T]}$ and $(I_i, \eta_i)_{i \in [n_t]} \, \forall t \in [T]$

1: **for** $t = 1, 2, \ldots, T$ **do**
2:      **for** $E_i \equiv E_i(I_i, \eta_i) \in \{E_1, E_2, \ldots, E_{n_t}\}$ **do** $\triangleright$ n_t experts active at t
3:          Receive action $\boldsymbol{x}_{t,I_i}^{\eta_i}$ from expert E_i
4:          **if** $\min\{\tau : \tau \in I_i\} = t$ **then** initialize $L_{t-1,I_i}^{\eta_i} = 0$, clipped domain $\tilde{\Theta}$ and parameter G
5:          **end if**
6:      **end for**
7:      Set meta-action as $\boldsymbol{x}_t = \sum_{i=1}^{n_t} \eta_i \exp(-L_{t-1,I_i}^{\eta_i})\boldsymbol{x}_{t,I_i}^{\eta_i} / \sum_{j=1}^{n_t} \eta_j \exp(-L_{t-1,I_j}^{\eta_j})$
8:      Sample $\boldsymbol{\zeta}_t$ uniformly from $\partial\mathbb{B}^d$
9:      Query point $\boldsymbol{z}_t = \boldsymbol{x}_t + h\boldsymbol{\zeta}_t$ to obtain $y_t = f_t(\boldsymbol{z}_t) + \xi_t$
10:      Construct gradient estimate $\boldsymbol{g}_t = (d/h)y_t\boldsymbol{\zeta}_t$
11:      **for** $i = 1, 2, \ldots, n_t$ **do**
12:          Send meta-action $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$ to E_i
13:          Increment cumulative loss $L_{t,I_i}^{\eta_i} = L_{t-1,I_i}^{\eta_i} + \ell_t^{\eta_i}(\boldsymbol{x}_{t,I_i}^{\eta_i})$ $\triangleright$ $\ell_t^{\eta_i}(\cdot)$ depends on $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$
14:      **end for**
15: **end for**

---

vector $\boldsymbol{g}_t$ is an unbiased gradient estimate of a spherically smoothed version of f_t at $\boldsymbol{x}_t$, satisfying

$$\mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t] = \nabla\hat{f}_t(\boldsymbol{x}_t), \quad \text{where} \quad \hat{f}_t(\boldsymbol{x}) = \mathbf{E}\left[f_t(\boldsymbol{x} + h\tilde{\boldsymbol{\zeta}})\right] \, \forall \, \boldsymbol{x} \in \tilde{\Theta}, \tag{9}$$

with $\tilde{\boldsymbol{\zeta}}$ distributed uniformly on the unit ball $\mathbb{B}^d$. Importantly, $\hat{f}_t$ inherits the convexity properties of f_t [71, Lemmas A.2–A.3]. Our approach differs from related works in OCO [22–24, 47, 48] that use exact gradients in two key ways: i) in each round, we query the perturbed point $\boldsymbol{z}_t = \boldsymbol{x}_t + h\boldsymbol{\zeta}_t$ rather than $\boldsymbol{x}_t$, accumulating regret at the perturbed point, and ii) we constrain $\boldsymbol{x}_t$ inside the clipped domain $\tilde{\Theta}$ to ensure all perturbed $\boldsymbol{z}_t$ remain feasible.

In our setting, under the high probability event $\Lambda_T = \left\{|\xi_t| \leq 2\sigma\sqrt{\log(T+1)}, \, \forall t \in [T]\right\}$, we have

$$\|\boldsymbol{g}_t\| = (d/h)|f_t(\boldsymbol{x}_t + h\boldsymbol{\zeta}_t) + \xi_t| \leq (d/h)\left(1 + 2\sigma\sqrt{\log(T+1)}\right) =: G, \quad \forall t \in [T]. \tag{10}$$

This implies a fundamental tradeoff in selecting the smoothing (and clipping) parameter h: larger values reduce G (and the variance of $\boldsymbol{g}_t$), but increase both the approximation error between $\hat{f}_t$ and f_t and the error due to clipping, while smaller values reduce bias at the cost of a higher variance in $\boldsymbol{g}_t$. In Theorem 1 and Corollary 1, we establish the optimal h and the resulting regret guarantees.

Algorithm 2 Expert algorithm $E(I, \eta)$: projected online gradient descent (OGD)

Input: $I = [r, s], \eta, G$, clipped domain $\tilde{\Theta}$, and surrogate loss $\ell_t^{\eta}(\cdot)$ defined in (11) $\forall t \in \mathbb{N}^+$
Initialize: $\boldsymbol{x}_{r,I}^{\eta}$ be any point in $\tilde{\Theta}$

1: **for** $t = r, r+1, \ldots, s$ **do**
2:      Send action $\boldsymbol{x}_{t,I}^{\eta}$ to Algorithm 1
3:      Receive meta-action $\boldsymbol{x}_t$ and $\boldsymbol{g}_t$ from Algorithm 1
4:      Update $\boldsymbol{x}_{t+1,I}^{\eta} = \Pi_{\tilde{\Theta}}\left(\boldsymbol{x}_{t,I}^{\eta} - \mu_t\nabla\ell_t^{\eta}(\boldsymbol{x}_{t,I}^{\eta})\right)$, where $\mu_t = 1/(2\eta^2 G^2(t - r + 1))$
5: **end for**

Design of expert algorithms and surrogate losses We choose projected online gradient descent (OGD) as the expert algorithms (Algorithm 2), i.e., each expert $E(I, \eta)$ runs OGD during its lifetime I. In the full-information setting, where experts observe f_t and gradients are evaluated at all of their actions, each expert could simply run OGD on the true loss functions. In contrast, for the bandit setting, with only one gradient estimate $\boldsymbol{g}_t$ of the smoothed loss $\hat{f}_t$ per round, we need to construct surrogate losses for the experts. The simplest option is the linear surrogate loss $\ell_t(\boldsymbol{x}) = -\boldsymbol{g}_t^\top(\boldsymbol{x}_t - \boldsymbol{x})$, but this fails to leverage curvature information and leads to a large $\widetilde{\mathcal{O}}(\sqrt{|I|})$ static regret for each expert, ultimately yielding linear adaptive regret.

To address these limitations, inspired by [22–24], we design the following strongly-convex surrogate loss $\ell_t^\eta : \mathbb{R}^d \to \mathbb{R}$:

$$\ell_t^\eta(\boldsymbol{x}) = -\eta \boldsymbol{g}_t^\top (\boldsymbol{x}_t - \boldsymbol{x}) + \eta^2 G^2 \|\boldsymbol{x}_t - \boldsymbol{x}\|^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^d, \tag{11}$$

where G is the upper bound (10) on $\|\boldsymbol{g}_t\|$, and η is the learning rate of the expert. We highlight that our choice of the quadratic term in (11) differs from the $\eta^2 \|\boldsymbol{g}_t\|^2 \|\boldsymbol{x}_t - \boldsymbol{x}\|^2$ and $\eta^2(\boldsymbol{g}_t^\top (\boldsymbol{x}_t - \boldsymbol{x}))^2$ in [24] and [22]. The latter necessitates an additional condition relating $\mathbf{E}[\|\boldsymbol{g}_t\|]$ and $\mathbf{E}[\|\boldsymbol{g}_t\|^2]$ (or $\mathbf{E}[\boldsymbol{g}_t \boldsymbol{g}_t^\top]$) to be satisfied in the analysis, see e.g., [22, Theorem 2], and may yield suboptimal rates in dimension d for strongly-convex losses, similar to [22]. Our choice of the quadratic term, similar to [23], eliminates these limitations and simplifies the proof.

For a comparator $\boldsymbol{u} \in \Theta$, (11) implies that the linearized regret associated with $\hat{f}_t$ on interval I can be bounded as:

$$\sum_{t \in I} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \rangle \leq \frac{1}{\eta} \underbrace{\sum_{t \in I} \mathbf{E}\big[\ell_t^\eta(\boldsymbol{x}_t) - \ell_t^\eta(\boldsymbol{u}) \mid \boldsymbol{x}_t, \Lambda_T\big]}_{:= \mathsf{A}} + \eta G^2 \sum_{t \in I} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2. \tag{12}$$

Due to the strong-convexity of ℓ_t^η, each expert attains only an $\mathcal{O}(\log |I|)$ static regret under OGD with an optimally tuned step size μ_t (see line 4 of Algorithm 2, and Lemma 6 in Appendix C.4). This ensures term A above is also of $\mathcal{O}(\log |I|)$. By the convexity of $\hat{f}_t$ we have

$$\sum_{t \in I} \mathbf{E}\big[\hat{f}_t(\boldsymbol{x}_t) - \hat{f}_t(\boldsymbol{u}) \mid \Lambda_T\big] \leq \mathbf{E}\Big[\tfrac{1}{\eta}\mathsf{A} + (\eta G^2 - \tfrac{\alpha}{2}) \sum_{t \in I} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2 \,\big|\, \Lambda_T\Big], \tag{13}$$

where $\alpha = 0$ for general convex $\hat{f}_t$ (and f_t), and $\alpha > 0$ for strongly-convex. Since both α and $\sum_{t \in I} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2$ are unknown a priori, we use a deterministic exponential grid of η values [19, 24], ensuring at least one expert covering I effectively minimize the RHS of (13), ultimately yielding a sublinear adaptive regret w.r.t. f_t. We present this result in the following theorem.

Theorem 1. *For any $T \in \mathbb{N}^+$ and $\mathsf{B} \in [T]$, Algorithm 1 with $h = \min(\sqrt{d}\mathsf{B}^{-\frac{1}{4}}, r)$ satisfies*

$$R^{ada}(\mathsf{B}, T) \lesssim \sqrt{d}\mathsf{B}^{\frac{3}{4}} + d^2, \tag{14}$$

and if f_t is α-strongly-convex with $\arg\min_{\boldsymbol{x} \in \mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$ for all $t \in [T]$,[1] it furthermore holds that

$$R^{ada}(\mathsf{B}, T) \lesssim \tfrac{d}{\alpha}\sqrt{\mathsf{B}} + \tfrac{1}{\alpha}d^2, \tag{15}$$

where $\lesssim$ conceals polylogarithmic terms in B and T, independent of d and α.

The proof of Theorem 1 can be found in Appendix C.2. We emphasize that TEWA-SE does not require knowledge of the strong-convexity parameter α. This parameter is only used in the analysis and appear in the upper bound (15). Compared to the $\mathcal{O}(\sqrt{\mathsf{B} \log T})$ and $\mathcal{O}(\frac{1}{\alpha} \log T \log \mathsf{B})$ adaptive regrets in [24] for general convex and strongly-convex losses respectively, our bounds in Theorem 1 reflect the separation between online first-order and zeroth-order optimization. This mirrors the established gap in static regret analyses, see e.g. [74] vs. [72]. We further note that our bound for the strongly-convex case has a $\frac{1}{\alpha}$ dependency, which is suboptimal compared to the $\frac{1}{\sqrt{\alpha}}$ dependency in [33, 73] for static regret in BCO for $\alpha \lesssim 1$.

Applying Proposition 1, the adaptive regret bounds in Theorem 1 lead to the following bounds for $R^{swi}(T, S)$, $R^{dyn}(T, \Delta, S)$ and $R^{path}(T, P)$. In Corollary 1, for clarity we drop the $\lceil \cdot \rceil$ operators from the expressions for B and assume without loss of generality B is an integer (proof in Appendix C.5).

Corollary 1. *Consider any horizon $T \in \mathbb{N}^+$ and assume that, for all $t \in [T]$, the loss f_t is convex, or strongly-convex with $\arg\min_{\boldsymbol{x} \in \mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$. We refer to the second scenario as the strongly-convex (SC) case. Then, Algorithm 1 tuned with parameter B satisfies the following regret guarantees:*

---

[1]The assumption that loss minimizers lie inside Θ is common in zeroth-order optimization, see e.g., [7, 72, 73]. Without it, our upper bound analysis would have an extra term depending on the gradients at the minimizers.

Switching. $\mathsf{B} = \frac{T}{S} \implies R^{swi}(T, S) \lesssim \begin{cases} \sqrt{d}S^{\frac{1}{4}}T^{\frac{3}{4}} + d^2 S \\ d\sqrt{ST} + d^2 S \quad \text{(SC)} \end{cases}$

Dynamic. $\begin{cases} \mathsf{B} = \frac{T}{S} \vee \left(\frac{\sqrt{d}T}{\Delta}\right)^{\frac{4}{5}} \Rightarrow R^{dyn}(T, \Delta, S) \lesssim R^{swi}(T, S) \wedge (d^{\frac{2}{5}}\Delta^{\frac{1}{5}}T^{\frac{4}{5}} + d^{\frac{8}{5}}\Delta^{\frac{4}{5}}T^{\frac{1}{5}}) \\ \mathsf{B} = \frac{T}{S} \vee \left(\frac{dT}{\Delta}\right)^{\frac{2}{3}} \Rightarrow R^{dyn}(T, \Delta, S) \lesssim R^{swi}(T, S) \wedge (d^{\frac{2}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}} + d^{\frac{4}{3}}\Delta^{\frac{2}{3}}T^{\frac{1}{3}}) \quad \text{(SC)} \end{cases}$

Path-length. $\begin{cases} \mathsf{B} = \left(\frac{r\sqrt{d}T}{P}\right)^{\frac{4}{5}} \Rightarrow R^{path}(T, P) \lesssim r^{-\frac{1}{5}}d^{\frac{2}{5}}P^{\frac{1}{5}}T^{\frac{4}{5}} + r^{-\frac{4}{5}}d^{\frac{8}{5}}P^{\frac{4}{5}}T^{\frac{1}{5}} \\ \mathsf{B} = \left(\frac{rdT}{P}\right)^{\frac{2}{3}} \Rightarrow R^{path}(T, P) \lesssim r^{-\frac{1}{3}}d^{\frac{2}{3}}P^{\frac{1}{3}}T^{\frac{2}{3}} + r^{-\frac{2}{3}}d^{\frac{4}{3}}P^{\frac{2}{3}}T^{\frac{1}{3}} \quad \text{(SC)} . \end{cases}$

2.1   Lower bound for strongly-convex loss functions

In this section, we derive a minimax lower bound on the dynamic regret and path-length regret, and discuss the optimality of TEWA-SE. To derive the lower bound for the dynamic regret, we adopt a standard minimax approach by constructing a class of hard functions, following [71, Theorem 6.1]. We assume that the adversary either (i) partitions the time horizon into S segments and assigns a different function from this class to each segment, or (ii) selects a sequence of functions with total variation bounded by Δ.

Theorem 2. *Let $\Theta = \mathbb{B}^d$. For $\alpha > 0$ denote by $\mathcal{F}_\alpha$ the class of α-strongly convex and smooth functions. Let $\pi = \{z_t\}_{t=1}^T$ be any randomized algorithm (see Appendix D for a definition). Then there exists $T_0 > 0$ such that for all $T \geq T_0$ it holds that*

$$\sup_{f_1,\ldots,f_T \in \mathcal{F}_\alpha} R^{dyn}(T, \Delta, S) \geq c_1 \cdot \left(d\sqrt{ST} \wedge d^{\frac{2}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}}\right), \tag{16}$$

where $c_1 > 0$ is a constant independent of $d, T,$ S and Δ.

We detail the proof in Appendix D. This lower bound establishes that TEWA-SE achieves the minimax-optimal dynamic regret (up to logarithms) for strongly convex and smooth functions w.r.t. d, T, S and Δ. We note that [7] derives a lower bound only in terms of T and Δ, matching (16), but it does not explicitly capture the dependence on d nor does it address the interplay between S and Δ. In the special case where $S = 1$, Theorem 2 recovers the classical minimax static regret of order $d\sqrt{T}$ [71, 72]. Interestingly, for $d = 1$ the scaling of the lower bound as function of T, S and Δ is the same as standard lower bounds in the non-stationary MAB literature [10, 11]. The proof of Theorem 2 can be readily adapted to consider only the measure S with the switching regret, yielding a rate of $d\sqrt{ST}$ and thereby establishing the minimax optimality of TEWA-SE's switching regret bound.

We also derive a lower bound for path-length regret analogously to that for dynamic regret. In Theorem 4 in Appendix D we show that under the same assumptions as in the statement of Theorem 2,

$$\sup_{f_1,\ldots,f_T \in \mathcal{F}_\alpha} R^{path}(T, P) \geq c_2 \cdot (d^2 P)^{\frac{2}{5}}T^{\frac{3}{5}}, \tag{17}$$

where $c_2 > 0$ is a constant independent of d, T and P. Hence, TEWA-SE may not achieve the optimal regret rate for path-length. Additionally, Eq. (17) improves upon the only existing $d\sqrt{PT}$ lower bound from [13] in terms of the horizon T, by leveraging a different construction of a hard instance. This improvement comes from assuming $P = o(T)$, which is necessary for sublinear regret.

2.2   Parameter-free guarantees

In Corollary 1, we showed that the knowledge of the non-stationarity measures S, Δ and P allows optimal tuning of TEWA-SE's parameter B. However, these measures can be hard to estimate. To obtain guarantees without such knowledge, we further analyze TEWA-SE under the *Bandit-over-Bandit* (BoB) framework from [27] (see Appendix C.6 for details), which divides the time horizon into epochs of suitable length L and uses an adversarial bandit algorithm (e.g., EXP3) to select B for TEWA-SE in each epoch from the set $\mathcal{B} = \{2^i : i = 0, 1, \ldots, \lfloor \log_2 T \rfloor\}$. In Corollary 3 in Appendix C.6, we adapt all the upper bounds in Corollary 1 to this framework, and show that this procedure costs an additional $d^{\frac{1}{3}}T^{\frac{5}{6}}$ term for the general convex case and $d^{\frac{1}{2}}T^{\frac{3}{4}}$ for the strongly-convex case. Our parameter-free path-length regret bound $P^{\frac{1}{5}}T^{\frac{4}{5}} + T^{\frac{5}{6}}$ for the general convex case improves on the $P^{\frac{1}{2}}T^{\frac{3}{4}}$ bound in [13] when $P = \Omega(T^{\frac{1}{6}})$.

Recent works on MAB [65–67, 75, 76] have proposed algorithms that achieve optimal dynamic regret without prior knowledge of S and Δ. However, they use procedures that crucially rely on the finiteness of the arm set, and are thus ill-suited for BCO. It remains open to determine if the minimax regret rate can be attained without such knowledge in the settings considered in this paper.

3   Clipped Exploration by Optimization

In this section, we propose a second algorithm (Algorithm 3) to improve upon the suboptimal rates for $R^{\mathsf{dyn}}(T, \Delta, S)$ and $R^{\mathsf{path}}(T, P)$ that TEWA-SE achieves for general convex loss functions. For ease of presentation, we assume in this section that the problem is noiseless, i.e., $\xi_t = 0$ for $t \in [T]$. We call this algorithm clipped Exploration by Optimization (cExO), which is built on Algorithm 8.3 (ExO) in [1]. The high level idea of ExO is to run exponential weights over a finite discretization of the feasible set, denoted by $\mathcal{C} \subset \Theta$. We assume the discretization $\mathcal{C}$ admits a worst-case error of $\varepsilon := \sup_{f \in \mathcal{F}_0} \min_{q \in \Delta(\mathcal{C})} \mathbf{E}_{z' \sim q} f(z') - \min_{z \in \Theta} f(z)$, where $\mathcal{F}_0$ denotes the class of convex and Lipschitz functions, and $\Delta(\mathcal{C})$ denotes the $(|\mathcal{C}| - 1)$-dimensional simplex.

With q_0 initialized as the uniform distribution, in each round t, given a loss estimate $\widehat{s}_t \in \mathbb{R}^{|\mathcal{C}|}$, ExO (in its mirror descent formulation) computes $q_t = \arg\min_{q \in \Delta(\mathcal{C})} \langle q, \widehat{s}_{t-1} \rangle + \frac{1}{\eta} \mathrm{KL}(q \| q_{t-1})$, where KL is the Kullback-Leibler divergence $\mathrm{KL}(q \| p) = \sum_{i=1}^{|\mathcal{C}|} q_i \log(q_i/p_i)$ for $q, p \in \Delta(\mathcal{C})$. The update rule in cExO departs from the vanilla ExO in this single step, by taking the minimum over the *clipped* simplex $\tilde{\Delta} = \Delta(\mathcal{C}) \cap [\gamma, 1]^{|\mathcal{C}|}$ where $\gamma \in (0, \frac{1}{|\mathcal{C}|})$ is a constant to be tuned, see line 2 of Algorithm 3. Clipping is a standard technique in mirror descent to ensure the algorithm does not commit too hard to any single action, and therefore detect changes in the environments more easily, yielding regret guarantees w.r.t. non-stationary measures [9, Chapter 31.1].

Given the reference distribution q_t, cExO selects a playing distribution $p_t \in \Delta(\mathcal{C})$ and an estimator function $E_t \in \mathcal{E}$ which returns an updated loss estimate for each action in $\mathcal{C}$, where $\mathcal{E}$ denotes the set of functions that map $\mathcal{C} \times [-1, 1]$ to $\mathbb{R}^{|\mathcal{C}|}$. It does so by solving an intractable optimization problem:[2]

$$\underset{p \in \Delta(\mathcal{C}), E \in \mathcal{E}}{\arg\min} \ \Lambda_\eta(q_t, p, E), \tag{18}$$

where, with $S_q(\eta \hat{s}) = \max_{q' \in \Delta(\mathcal{C})} \langle q - q', \eta \hat{s} \rangle - \mathrm{KL}(q' \| q)$, the objective function is defined by

$$\Lambda_\eta(q, p, E) := \sup_{p^\star \in \Delta(\mathcal{C})} \sup_{f \in \mathcal{F}_0} \mathbf{E}_{z \sim p} \Big[\langle p - p^\star, f \rangle + \langle p^\star - q, E(z, f(z)) \rangle + \frac{1}{\eta} S_q(\eta E(z, f(z))) \Big].$$

This optimization problem is intractable due to the large size of $\mathcal{E}$ and $\mathcal{F}_0$.[3] The role of this optimization problem is to tradeoff the worst-case cost of deviating from the desired distribution q_t versus the gain of improved exploration (hence the name Exploration by Optimization). Finally, cExO samples an action z_t according to p_t, observes the feedback $f(z_t)$ and constructs a loss estimate $\widehat{s}_t = E_t(z_t, f(z_t))$ to be used in the subsequent round. cExO achieves the adaptive regret guarantee stated in Theorem 3 below.

Theorem 3. *For $T \in \mathbb{N}^+$ and $\mathsf{B} \in [T]$, Algorithm 3 calibrated with $\varepsilon = \frac{1}{T}$, $\gamma = \frac{1}{T|\mathcal{C}|}$, $\eta = \sqrt{\log(\gamma^{-1})/(d^4 \log(dT)\mathsf{B})}$ and $\log|\mathcal{C}| = \mathcal{O}(d \log(dT^2))$ satisfies*

$$R^{\mathsf{ada}}(\mathsf{B}, T) \lesssim d^{\frac{5}{2}} \sqrt{\mathsf{B}}. \tag{19}$$

We then use Proposition 1 to convert the bound of Theorem 3 into the following regret guarantees w.r.t. S, Δ and P. Like in Corollary 1, we omit $\lceil \cdot \rceil$ from the expressions for B for clarity.

Corollary 2. *For any horizon $T \in \mathbb{N}^+$, Algorithm 3 calibrated as in Theorem 3 and tuned with interval size B (which determines η) satisfies the following regret guarantees:*

> ***Switching:*** $\quad \mathsf{B} = \frac{T}{S} \implies R^{\mathsf{swi}}(T, S) \lesssim d^{\frac{5}{2}} \sqrt{ST}$,
>
> ***Dynamic:*** $\quad \mathsf{B} = \frac{T}{S} \vee (d^{\frac{5}{2}} T / \Delta)^{\frac{2}{3}} \implies R^{\mathsf{dyn}}(T, \Delta, S) \lesssim R^{\mathsf{swi}}(T, S) \wedge d^{\frac{5}{3}} \Delta^{\frac{1}{3}} T^{\frac{2}{3}}$,
>
> ***Path-length:*** $\quad \mathsf{B} = (r d^{\frac{5}{2}} T / P)^{\frac{2}{3}} \implies R^{\mathsf{path}}(T, P) \lesssim r^{-\frac{1}{3}} d^{\frac{5}{3}} P^{\frac{1}{3}} T^{\frac{2}{3}}$.

---

[2]For detailed discussions about these functions, we refer the reader to [28].

[3]One can in theory bound the domain of $\mathcal{E}$ and discretize $\mathcal{E}$, $\mathcal{F}_0$ and $\Delta(\mathcal{C})$. The optimization problem is hence computable, though not in polynomial time.

Algorithm 3 clipped Exploration by Optimization (cExO)

Input: d, T, B, feasible set Θ, a finite covering set $\mathcal{C} \subset \Theta$ of Θ, discretization error ε, learning rate η, clipping parameter $\gamma \in (0, \frac{1}{|\mathcal{C}|})$, and $\tilde{\Delta} = \Delta(\mathcal{C}) \cap [\gamma, 1]^{|\mathcal{C}|}$

Initialize: $q_{0,i} = \frac{1}{|\mathcal{C}|} \; \forall i \in [|\mathcal{C}|]$.

1: **for** $t = 1, \ldots, T$ **do**
2:     Compute $\boldsymbol{q}_t = \arg\min_{\boldsymbol{q} \in \tilde{\Delta}} \langle \boldsymbol{q}, \widehat{\boldsymbol{s}}_{t-1} \rangle + \frac{1}{\eta} \mathrm{KL}(\boldsymbol{q} || \boldsymbol{q}_{t-1})$
3:     Find distribution $\boldsymbol{p}_t \in \Delta(\mathcal{C})$ and $E_t \in \mathcal{E}$ s.t. $\Lambda_\eta(\boldsymbol{q}_t, \boldsymbol{p}_t, E_t) \leq \inf_{\substack{\boldsymbol{p} \in \Delta(\mathcal{C}), \\ E \in \mathcal{E}}} \Lambda_\eta(\boldsymbol{q}_t, \boldsymbol{p}, E) + \eta d$
4:     Sample $\boldsymbol{z}_t \sim \boldsymbol{p}_t$ and observe $f_t(\boldsymbol{z}_t)$
5:     Compute $\widehat{\boldsymbol{s}}_t = E_t(\boldsymbol{z}_t, f_t(\boldsymbol{z}_t))$
6: **end for**

The proofs of Theorem 3 and Corollary 2 are presented in Appendix E. By comparing these results to the lower bounds in Section 2.1, we obtain that for known S, Δ and P, cExO achieves minimax-optimal rates in T, S and Δ, but remains suboptimal in d (for all measures), and potentially for the path-length bound (see Eq. (17)). This suboptimal dependence on d is unsurprising since even the best known static regret bounds of [31] and [34] suffer from similar dimensional dependence. Moreover, the gap between cExO's path-length regret bound and our minimax lower bound of order $d^{\frac{4}{5}} P^{\frac{2}{5}} T^{\frac{3}{5}}$ may stem from either (i) looseness in the lower bound, or (ii) sub-optimality of cExO, which runs OMD in distribution space rather than directly on the action set. The latter may allow us to bound path-length regret more directly and sharply.

To adapt to unknown non-stationarity measures, cExO equipped with the BoB framework yields the upper bounds in Corollary 2 with an additional $d^{\frac{5}{4}} T^{\frac{3}{4}}$ term (see Corollary 4 in Appendix E). Our path-length regret of $P^{\frac{1}{3}} T^{\frac{2}{3}}$ and $P^{\frac{1}{3}} T^{\frac{2}{3}} + T^{\frac{3}{4}}$ for known and unknown P, respectively, improves on the $P^{\frac{1}{4}} T^{\frac{3}{4}}$ and $P^{\frac{1}{2}} T^{\frac{3}{4}}$ rates in [13] in terms of T.

4 Conclusion

In this work, we develop and analyze two approaches for non-stationary Bandit Convex Optimization. For strongly convex losses, our polynomial-time TEWA-SE algorithm achieves minimax-optimal dynamic regret w.r.t. S and Δ without knowing the strong-convexity parameter, but incurs a sub-optimal $T^{\frac{3}{4}}$ rate for general convex losses. To address this, we propose a second algorithm, cExO, which achieves minimax-optimality for S and Δ. However, this algorithm is not polynomial-time computable and has an increased dimension dependence. Our matching lower bounds confirm the optimality results, but also reveal potentially suboptimal guarantees w.r.t. the path-length P. This work highlights a central open challenge: designing algorithms that are simultaneously minimax-optimal and computationally efficient for general convex losses in non-stationary environments. A promising stepstone towards this goal is to incorporate second-order information, akin to the online Newton methods from [31, 34] that achieve state-of-the-art static regret guarantees for adversarial convex bandits. In particular, a restart criterion, similar to the one in line 15 of [35, Algorithm 1] or line 11 of [31, Algorithm 1], may enable tracking capabilities and lead to improved regret bounds.

Acknowledgments and Disclosure of Funding

References

[1] Tor Lattimore. Bandit convex optimisation. *arXiv:2402.06535*, 2024.

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

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

[4] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936, 2003.

[5] Aryan Mokhtari, Shahin Shahrampour, Ali Jadbabaie, and Alejandro Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *Conference on Decision and Control*, pages 7195–7201. IEEE, 2016.

[6] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pages 398–406. PMLR, 2015.

[7] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.

[8] Eric C Hall and Rebecca M Willett. Online convex optimization in dynamic environments. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):647–662, 2015.

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

[10] Aurélien Garivier and Eric Moulines. On upper-confidence bound policies for switching bandit problems. In *Proceedings of the 22nd International Conference on Algorithmic Learning Theory*, 2011.

[11] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in Neural Information Processing Systems*, 27, 2014.

[12] Lilian Besson, Emilie Kaufmann, Odalric-Ambrym Maillard, and Julien Seznec. Efficient change-point detection for tackling piecewise-stationary bandits. *Journal of Machine Learning Research*, 23(77):1–40, 2022.

[13] Peng Zhao, Guanghui Wang, Lijun Zhang, and Zhi-Hua Zhou. Bandit convex optimization in non-stationary environments. *Journal of Machine Learning Research*, 22(125):1–45, 2021.

[14] Yining Wang. On adaptivity in nonstationary stochastic optimization with bandit feedback. *Operations Research*, 73(2):819–828, 2025.

[15] Mark Herbster and Manfred K. Warmuth. Tracking the Best Expert. *Machine Learning*, 32(2):151–178, August 1998.

[16] Lijun Zhang, Tianbao Yang, Zhi-Hua Zhou, et al. Dynamic regret of strongly adaptive methods. In *International conference on machine learning*, pages 5882–5891. PMLR, 2018.

[17] Lijun Zhang, Shiyin Lu, and Tianbao Yang. Minimizing dynamic regret and adaptive regret simultaneously. In *International Conference on Artificial Intelligence and Statistics*, pages 309–319. PMLR, 2020.

[18] Elad Hazan and C. Seshadhri. Efficient learning algorithms for changing environments. In *International Conference on Machine Learning*, volume 382 of *ACM International Conference Proceeding Series*, pages 393–400. ACM, 2009.

[19] Amit Daniely, Alon Gonen, and Shai Shalev-Shwartz. Strongly adaptive online learning. In *International Conference on Machine Learning*, pages 1405–1411. PMLR, 2015.

[20] Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, May 1997.

[21] Tianyi Chen and Georgios B Giannakis. Bandit convex optimization for scalable and dynamic IoT management. *IEEE Internet of Things Journal*, 6(1):1276–1286, 2018.

[22] Tim van Erven, Wouter M. Koolen, and Dirk van der Hoeven. Metagrad: Adaptation using multiple learning rates in online learning. *Journal of Machine Learning Research*, 22(161):1–61, 2021.

[23] Guanghui Wang, Shiyin Lu, and Lijun Zhang. Adaptivity and optimality: A universal algorithm for online convex optimization. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 659–668. PMLR, 2020.

[24] Lijun Zhang, Guanghui Wang, Wei-Wei Tu, Wei Jiang, and Zhi-Hua Zhou. Dual adaptivity: a universal algorithm for minimizing the adaptive regret of convex functions. In *International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2021.

[25] Abraham Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 385–394. SIAM, 2005.

[26] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *International Conference on Neural Information Processing Systems*, page 697–704. MIT Press, 2004.

[27] Wang Chi Cheung, David Simchi-Levi, and Ruihao Zhu. Learning to optimize under non-stationarity. In *International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1079–1087. PMLR, 2019.

[28] Tor Lattimore and Andras Gyorgy. Mirror descent and the information ratio. In *Conference on Learning Theory*, pages 2965–2992. PMLR, 2021.

[29] Alekh Agarwal, Dean P Foster, Daniel J Hsu, Sham M Kakade, and Alexander Rakhlin. Stochastic convex optimization with bandit feedback. *Advances in Neural Information Processing Systems*, 24, 2011.

[30] Ankan Saha and Ambuj Tewari. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *International conference on artificial intelligence and statistics*, pages 636–642. JMLR Workshop and Conference Proceedings, 2011.

[31] Hidde Fokkema, Dirk van der Hoeven, Tor Lattimore, and Jack J Mayo. Online newton method for bandit convex optimisation. In *Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 1713–1714. PMLR, 2024.

[32] Sébastien Bubeck, Ronen Eldan, and Yin Tat Lee. Kernel-based methods for bandit convex optimization. *Journal of the ACM*, 68(4):1–35, 2021.

[33] Elad Hazan and Kfir Levy. Bandit convex optimization: Towards tight bounds. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[34] Arun Suggala, Y Jennifer Sun, Praneeth Netrapalli, and Elad Hazan. Second order methods for bandit optimization and control. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 4691–4763. PMLR, 2024.

[35] Arun Sai Suggala, Pradeep Ravikumar, and Praneeth Netrapalli. Efficient bandit convex optimization: Beyond linear losses. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134, pages 4008–4067. PMLR, 2021.

[36] Tor Lattimore and András György. A second-order method for stochastic bandit convex optimisation. In *Conference on Learning Theory*, pages 2067–2094. PMLR, 2023.

[37] Olivier Bousquet and Manfred K Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3(Nov):363–396, 2002.

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

[39] V Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

[40] Yoav Freund, Robert Schapire, Yoram Singer, and Manfred Warmuth. Using and combining predictors that specialize. *Conference Proceedings of the Annual ACM Symposium on Theory of Computing*, 01 1997.

[41] Wouter M Koolen and Tim Van Erven. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pages 1155–1175. PMLR, 2015.

[42] Dmitry Adamskiy, Wouter M. Koolen, Alexey Chernov, and Vladimir Vovk. A closer look at adaptive regret. *Journal of Machine Learning Research*, 17(23):1–21, 2016.

[43] Kwang-Sung Jun, Francesco Orabona, Stephen Wright, and Rebecca Willett. Online learning for changing environments using coin betting. *arXiv preprint arXiv:1711.02545*, 2017.

[44] Ashok Cutkosky. Parameter-free, dynamic, and strongly-adaptive online learning. In *International Conference on Machine Learning*, volume 119, pages 2250–2259. PMLR, 2020.

[45] Zhou Lu, Wenhan Xia, Sanjeev Arora, and Elad Hazan. Adaptive gradient methods with local guarantees. *arXiv preprint arXiv:2203.01400*, 2022.

[46] Dheeraj Baby and Yu-Xiang Wang. Optimal dynamic regret in proper online learning with strongly convex losses and beyond. In *International Conference on Artificial Intelligence and Statistics*, pages 1805–1845. PMLR, 2022.

[47] Guanghui Wang, Dakuan Zhao, and Lijun Zhang. Minimizing adaptive regret with one gradient per iteration. In *International Joint Conference on Artificial Intelligence*, IJCAI'18, page 2762–2768. AAAI Press, 2018.

[48] Peng Zhao, Yan-Feng Xie, Lijun Zhang, and Zhi-Hua Zhou. Efficient methods for non-stationary online learning. *Advances in Neural Information Processing Systems*, 35:11573–11585, 2022.

[49] Lijun Zhang, Shiyin Lu, and Zhi-Hua Zhou. Adaptive online learning in dynamic environments. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 1330–1340. Curran Associates Inc., 2018.

[50] Wenhao Yang, Yibo Wang, Peng Zhao, and Lijun Zhang. Universal online convex optimization with 1 projection per round. In *Advances in Neural Information Processing Systems*, volume 37, pages 31438–31472. Curran Associates, Inc., 2024.

[51] Peter L. Bartlett, Elad Hazan, and Alexander Rakhlin. Adaptive online gradient descent. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*, page 65–72. Curran Associates Inc., 2007.

[52] Chuong B. Do, Quoc V. Le, and Chuan-Sheng Foo. Proximal regularization for online and batch learning. In *International Conference on Machine Learning*, page 257–264. Association for Computing Machinery, 2009.

[53] Daniel Russo and Benjamin Van Roy. Learning to optimize via information-directed sampling. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[54] Sébastien Bubeck, Ofer Dekel, Tomer Koren, and Yuval Peres. Bandit convex optimization: $\sqrt{T}$ regret in one dimension. In *Conference on Learning Theory*, pages 266–278. PMLR, 2015.

[55] Sébastien Bubeck and Ronen Eldan. Exploratory distributions for convex functions. *Mathematical Statistics and Learning*, 1(1):73–100, 2018.

[56] Tor Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. *Mathematical Statistics and Learning*, 2(3):311–334, 2020.

[57] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002.

[58] Olivier Cappé, Aurélien Garivier, Odalric-Ambrym Maillard, Rémi Munos, and Gilles Stoltz. Kullback-Leibler upper confidence bounds for optimal sequential allocation. *Annals of Statistics*, 41(3):1516–1541, 2013.

[59] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

[60] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[61] Shipra Agrawal and Navin Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory*, 2012.

[62] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *International conference on algorithmic learning theory*, pages 199–213. Springer, 2012.

[63] Fang Liu, Joohyun Lee, and Ness B. Shroff. A change-detection based framework for piecewise-stationary multi-armed bandit problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[64] Yang Cao, Zheng Wen, Branislav Kveton, and Yao Xie. Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit. In *International Conference on Artificial Intelligence and Statistics*, pages 418–427. PMLR, 2019.

[65] Peter Auer, Pratik Gajane, and Ronald Ortner. Adaptively tracking the best bandit arm with an unknown number of distribution changes. In *Conference on Learning Theory*, volume 99, pages 138–158. PMLR, 2019.

[66] Chen-Yu Wei and Haipeng Luo. Non-stationary reinforcement learning without prior knowledge: An optimal black-box approach. In *Conference on learning theory*, pages 4300–4354. PMLR, 2021.

[67] Joe Suk and Samory Kpotufe. Tracking most significant arm switches in bandits. In *Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 2160–2182, 2022.

[68] Yoan Russac, Claire Vernade, and Olivier Cappé. Weighted linear bandits for non-stationary environments. *Advances in Neural Information Processing Systems*, 32, 2019.

[69] Francesco Trovò, Marcello Restelli, and Nicola Gatti. Sliding-window thompson sampling for non-stationary settings. *Journal of Artificial Intelligence Research*, 68:311–364, 2020.

[70] Dorian Baudry, Yoan Russac, and Olivier Cappé. On limited-memory subsampling strategies for bandits. In *International Conference on Machine Learning*, pages 727–737. PMLR, 2021.

[71] Arya Akhavan, Massimiliano Pontil, and Alexandre Tsybakov. Exploiting higher order smoothness in derivative-free optimization and continuous bandits. *Advances in Neural Information Processing Systems*, 33:9017–9027, 2020.

[72] Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory*, pages 3–24. PMLR, 2013.

[73] Shinji Ito. An optimal algorithm for bandit convex optimization with strongly-convex and smooth loss. In *International Conference on Artificial Intelligence and Statistics*, pages 2229–2239. PMLR, 2020.

[74] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192, December 2007.

[75] Haipeng Luo, Chen-Yu Wei, Alekh Agarwal, and John Langford. Efficient contextual bandits in non-stationary worlds. In *Conference On Learning Theory*, pages 1739–1776. PMLR, 2018.

[76] Yifang Chen, Chung-Wei Lee, Haipeng Luo, and Chen-Yu Wei. A new algorithm for non-stationary contextual bandits: Efficient, optimal and parameter-free. In *Conference on Learning Theory*, pages 696–726. PMLR, 2019.

[77] Gilles Stoltz. *Incomplete information and internal regret in prediction of individual sequences*. Theses, Université Paris Sud - Paris XI, May 2005.

[78] Arya Akhavan, Karim Lounici, Massimiliano Pontil, and Alexandre B Tsybakov. Contextual continuum bandits: Static versus dynamic regret. *arXiv preprint arXiv:2406.05714*, 2024.

[79] Alexandre B. Tsybakov. *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York, 2009.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: In Section 1, we define our problem setting, goals of the paper, and assumptions we make throughout the paper. In Section 1.1 we describe our main contributions and results, while in Section 1.2 we contextualize our work by discussing related work. The abstract summarizes these.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In the main contributions section (Section 1.1), we describe the limitations of the work: the polynomial-time algorithm TEWA-SE we develop attains suboptimal regret bounds for general convex losses, and cExO, the second algorithm we propose, achieves optimal rates w.r.t. T, S and Δ, but is not polynomial-time computable and has suboptimal rates w.r.t. the problem dimension d. These limitations are discussed in detail after we state each theoretical result, and restated in the conclusion (Section 4). We also discuss the potential suboptimality of both algorithms w.r.t. the path-length P in our main results section and in the conclusion.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We state all assumptions we make in Section 1. Each theoretical result is followed by a reference to its detailed proof in the appendices. We provide all proofs in the appendices: definitions (Appendix A), proof of Proposition 1 (Appendix B), proofs for TEWA-SE (Appendix C), proofs of lower bounds (Appendix D), and proofs for cExO (Appendix E).

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: This paper is theoretical in nature and does not include experimental results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: This paper does not include experiments.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: This paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: This paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: This work has been conducted in a way that fully conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: This work is theoretical in nature and is not expected to have significant social impacts, positive or negative.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: This paper does not use such assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: This paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.

A Definitions

Definition 1. *Let $\sigma > 0$. A random variable ξ is σ-sub-Gaussian if for any $\lambda > 0$ we have $\mathbf{E}[\exp(\lambda\xi)] \leq \exp(\sigma^2\lambda^2/2)$.*

Definition 2. *Let $\alpha > 0$. A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is called α-strongly convex, if for $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d$, $f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{z} - \boldsymbol{x}) + \frac{\alpha}{2}\|\boldsymbol{z} - \boldsymbol{x}\|^2$.*

Definition 3. *Let $\beta > 0$. Function $f : \mathbb{R}^d \to \mathbb{R}$ is called β-smooth, if it is continuously differentiable and for any $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d$, $\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{z})\| \leq \beta\|\boldsymbol{x} - \boldsymbol{z}\|$.*

Definition 4. *Let $K > 0$. Function $f : \mathbb{R}^d \to \mathbb{R}$ is called K-Lipschitz if for any $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d$, $|f(\boldsymbol{x}) - f(\boldsymbol{z})| \leq K\|\boldsymbol{x} - \boldsymbol{z}\|$.*

B Proof of Proposition 1

We start this section by restating the proposition, before detailing its proof.

Proposition 1. *Suppose that an algorithm can be calibrated to satisfy $R^{ada}(\mathsf{B}, T) \leq C\mathsf{B}^\kappa$, for any interval length $\mathsf{B} \in [T]$, for some factor $C > 0$ that is at most polynomial in d and $\log(T)$, and $\kappa \in [0, 1)$.*

Then, for any $S, S_\Delta, S_P \in [T]$, an appropriate choice of B yields the following regret guarantees:

Switching: $\mathsf{B} = \lceil\frac{T}{S}\rceil$ *guarantees that $R^{swi}(T, S) \leq 2^{1+\kappa}CS^{1-\kappa}T^\kappa$.*

Dynamic: $\mathsf{B} = \lceil\frac{T}{S}\rceil \vee \lceil\frac{T}{S_\Delta}\rceil$ *yields $R^{dyn}(T, \Delta, S) \leq R^{swi}(T, S) \wedge \left(R^{swi}(T, S_\Delta) + \Delta\lceil\frac{T}{S_\Delta}\rceil\right)$.*

Path-length: $\mathsf{B} = \lceil\frac{T}{S_P}\rceil$ *ensures that $R^{path}(T, P) \leq R^{swi}(T, S_P) + \frac{P}{r}\cdot\lceil\frac{T}{S_P}\rceil$.*

Proof of Proposition 1. The proof follows two steps. First, we state in Lemma 1 the conversion between adaptive regret and switching regret. A similar conversion can be found in [19], but we detail the proof for completeness. Next, we prove in Lemma 2 that switching regret guarantees for appropriate number of switches convert into dynamic and path-length regret guarantees. ☐

In the remainder of this section, we detail the two supporting lemmas and their proof.

Lemma 1. *Consider an algorithm that satisfies the adaptive regret guarantees of Proposition 1, then this algorithm calibrated with interval size $\mathsf{B} = \lceil\frac{T}{S}\rceil$ satisfies*

$$R^{swi}(T, S) \leq 2^{1+\kappa}CS^{1-\kappa}T^\kappa .$$

Proof of Lemma 1. Consider $\mathsf{B} = \lceil\frac{T}{S}\rceil$. Let $\boldsymbol{u}_{1:T} \in \Theta^T$ be a sequence of arbitrary comparators with at most S switches. We divide the horizon into intervals of length B (the last interval may be shorter than B), and further divide the intervals at the rounds where $\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}$. This ensures each of these intervals is associated with a constant comparator. By construction, these intervals are of length $\leq \mathsf{B}$ and the number of intervals is bounded by $2S$. Hence, we can apply the adaptive regret bound to each interval to obtain

$$R^{swi}(T, S) \leq 2S \cdot C\mathsf{B}^\kappa \leq 2CS \cdot \left(\frac{T}{S} + 1\right)^\kappa$$

$$= 2C \cdot S^{1-\kappa}T^\kappa \cdot \left(1 + \frac{S}{T}\right)^\kappa \leq 2^{1+\kappa}C \cdot S^{1-\kappa}T^\kappa .$$

☐

We now prove the conversion between switching regret and dynamic and path-length regrets.

Lemma 2. *Consider any fictitious number of switches $S' \in [T]$. Then the dynamic regret of environments constrained by Δ satisfies*

$$R^{dyn}(T, \Delta) \leq R^{swi}(T, S') + \Delta\lceil\tfrac{T}{S'}\rceil , \tag{20}$$

and the path-length regret satisfies

$$R^{path}(T, P) \leq R^{swi}(T, S') + \tfrac{P}{r}\lceil\tfrac{T}{S'}\rceil . \tag{21}$$

Proof of Lemma 2. For both upper bounds, the switching regret term comes from dividing the horizon $[T]$ into S' intervals, denoted by $(I_s)_{s \in [S']}$, each of length at most $\lceil \frac{T}{S'} \rceil$ (defining them precisely is not important for the following arguments). Recall the definition of $R(T, \boldsymbol{u}_{1:T})$ from (2). For any sequence of actions $\boldsymbol{z}_{1:T} \in \Theta^T$ chosen by the given algorithm, and for any arbitrary comparator sequences $\boldsymbol{u}_{1:T} \in \Theta^T$ and $\boldsymbol{v}_{1:S'} \in \Theta^{S'}$, it holds that

$$
\begin{aligned}
R(T, \boldsymbol{u}_{1:T}) &= \sum_{t=1}^{T} \mathbf{E}\left[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u}_t) \right] \\
&= \sum_{t=1}^{T} \sum_{s=1}^{S'} \mathbf{1}(t \in I_s) \cdot \left(\mathbf{E}\left[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{v}_s) \right] + (f_t(\boldsymbol{v}_s) - f_t(\boldsymbol{u}_t)) \right) \\
&\leq R^{\mathsf{swi}}(T, S') + \sum_{s=1}^{S'} \underbrace{\sum_{t \in I_s} (f_t(\boldsymbol{v}_s) - f_t(\boldsymbol{u}_t))}_{=:V_s} ,
\end{aligned}
\tag{22}
$$

where the last step holds by the definition of the switching regret. It thus remains to choose a suitable $\boldsymbol{v}_s \in \Theta$ and upper bound the term V_s for $s \in [S']$. We choose a different $\boldsymbol{v}_s$ for the proof of the dynamic regret bound vs. that of the path-length regret bound.

Dynamic regret. Consider the interval I_s for $s \in [S']$. Let L_s be its length and $\Delta_s = \sum_{t \in I_s} \max_{\boldsymbol{z} \in \Theta} |f_t(\boldsymbol{z}) - f_{t-1}(\boldsymbol{z})|$ be the total variation over this interval. Then, for any two time steps t and t' in I_s and any $\boldsymbol{z} \in \Theta$, it holds that $f_t(\boldsymbol{z}) - f_{t'}(\boldsymbol{z}) \leq \Delta_s$ by definition of total variation. Let $\bar{f}_s$ denote the average of the functions over the interval I_s and define $\boldsymbol{v}_s \in \arg\min_{\boldsymbol{z} \in \Theta} \bar{f}_s(\boldsymbol{z})$, then we have

$$
\forall s \in [S'], \quad V_s \leq \sum_{t \in I_s} (f_t(\boldsymbol{v}_s) - \bar{f}_s(\boldsymbol{u}_t) + \Delta_s) \leq \Delta_s L_s .
$$

Taking the sum over all intervals and using $L_s \leq \lceil \frac{T}{S'} \rceil$ completes the proof of (20).

Path-length regret. This proof proceeds similarly as that for the dynamic regret. Consider the interval I_s for some $s \in [S']$, and denote by L_s its length and $P_s = \sum_{t \in I_s} \|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\|$ the path-length of the comparator sequence on this interval. For the proof, we construct $\boldsymbol{v}_s \in \Theta$ differently from that in the proof of the dynamic regret. Before detailing the construction of $\boldsymbol{v}_s$, we first define a set of comparators $(\boldsymbol{u}'_t)_{t \in I_s} \in \Theta^{L_s}$ as follows: for some $\alpha_0 \in [0, 1]$ and any time $t \in I_s$, we define $\boldsymbol{u}'_t$ to satisfy $\boldsymbol{v}_s = \alpha_0 \boldsymbol{u}_t + (1 - \alpha_0)\boldsymbol{u}'_t$. Using this and by the convexity and boundedness of f_t, we can bound that

$$
f_t(\boldsymbol{v}_s) \leq f_t(\boldsymbol{u}_t) + (1 - \alpha_0)(f_t(\boldsymbol{u}'_t) - f_t(\boldsymbol{u}_t)) \leq f_t(\boldsymbol{u}_t) + 2(1 - \alpha_0) .
$$

We then proceed by choosing a suitable $\boldsymbol{v}_s$ and α_0 to make this bound depend on the path-length. Since the path-length is P_s, there exists an ℓ_2-ball of radius $\frac{P_s}{2}$ that contains all the comparators $(\boldsymbol{u}_t)_{t \in I_s}$, and its center $\boldsymbol{c}_u$ lies in the feasible domain Θ. By assumption (as we stated in Section 1), there also exists a ball with radius r and center $\boldsymbol{c}_r$ within the domain. We can thus construct $\boldsymbol{v}_s$ to satisfy

$$
\boldsymbol{v}_s = \alpha_0 \boldsymbol{c}_u + (1 - \alpha_0)\boldsymbol{c}_r \in \Theta , \qquad \text{which yields} \quad \boldsymbol{u}'_t = \boldsymbol{c}_r + \frac{\alpha_0}{1 - \alpha_0}(\boldsymbol{c}_u - \boldsymbol{u}_t) ,
\tag{23}
$$

where $\boldsymbol{v}_s \in \Theta$ due to the convexity of the domain. Our goal is then to choose α_0 as large as possible (to make $1 - \alpha_0$ small) such that all the comparators $(\boldsymbol{u}'_t)_{t \in I_s}$ belong to Θ. Eq. (23) implies that

$$
\|\boldsymbol{u}'_t - \boldsymbol{c}_r\| = \frac{\alpha_0}{1 - \alpha_0}\|\boldsymbol{c}_u - \boldsymbol{u}_t\| \leq \frac{\alpha_0}{1 - \alpha_0} \cdot \frac{P_s}{2} = \frac{\alpha_0 P_s}{2(1 - \alpha_0)} ,
$$

which by definition of the r-ball guarantees that $\boldsymbol{u}'_t \in \Theta$ as long as $\frac{\alpha_0 P_s}{2(1 - \alpha_0)} \leq r$. To satisfy this condition, we can thus pick $\alpha_0 = \frac{2r}{P_s + 2r}$, which guarantees by construction that

$$
\forall t \in I_s : \quad f_t(\boldsymbol{v}_s) \leq f_t(\boldsymbol{u}_t) + \frac{2P_s}{P_s + 2r} \leq f_t(\boldsymbol{u}_t) + \frac{P_s}{r} .
$$

The desired bound on V_s in (22) directly follows. The final result (21) then comes by summation over all intervals $(I_s)_{s \in [S']}$. $\qquad\square$

24

C Details and proofs for TEWA-SE

In this appendix, we provide additional details on TEWA-SE in Section C.1 and establish its theoretical guarantees in Sections C.4–C.6. We present the proof of Theorem 1 in Section C.2, followed by the supporting lemmas in Sections C.3 and C.4. We then provide the proof of Corollary 1 in Section C.5, and the parameter-free guarantees in Section C.6.

C.1 Additional details on TEWA-SE

As we described in Section 2, TEWA-SE handles non-stationary environments by employing the Geometric Covering (GC) scheme from [19] to schedule experts across different time intervals. Additionally, TEWA-SE assigns an exponential grid of learning rates to the multiple experts covering each GC interval, to adapt to the curvature of the loss functions. We first invoke the definition of GC intervals from [19].

Definition 5 (Geometric Covering (GC) intervals [19]). *For $k \in \mathbb{N}$, define the set of intervals*

$$\mathcal{I}_k = \left\{ [i \cdot 2^k, (i+1) \cdot 2^k - 1] : i \in \mathbb{N}^+ \right\}, \tag{24}$$

that is, $\mathcal{I}_k$ is a partition of $\mathbb{N}^+ \setminus [2^k - 1]$ into intervals of length 2^k. Then we call $\mathcal{I} = \bigcup_{k \in \mathbb{N}} \mathcal{I}_k$ the set of Geometric Covering (GC) intervals.

For any interval length $L \in \mathbb{N}^+$, we also define the exponential grid of learning rates as

$$\mathcal{S}(L) = \left\{ \frac{2^{-i}}{5GD} : i \in \left\{ 0, 1, \ldots, \left\lceil \frac{1}{2} \log_2 L \right\rceil \right\} \right\}, \tag{25}$$

where G is the uniform upper bound (10) on $\|\boldsymbol{g}_t\|$, and D is the diameter of the feasible set Θ. For each given GC interval $I = [r, s] \in \mathcal{I}$, TEWA-SE instantiates multiple experts in round r, each assigned a distinct learning rate $\eta \in \mathcal{S}(|I|)$ and surrogate loss ℓ_t^η as defined in (11). It removes these experts after round s. This scheduling scheme ensures at least one expert covering I effectively minimizes the linearized regret $\sum_{t \in I} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \rangle$ associated with $\hat{f}_t$ on the interval I (Lemma 5), ultimately yielding the regret guarantees in Theorem 1 and Corollary 1.

Polylogarithmic computational complexity For $t \in \mathbb{N}^+$, we use $\mathcal{C}_t = \{ I \in \mathcal{I} : t \in I \}$ to denote the set of GC intervals covering time t. From Definition 5 it is easy to verify that $|\mathcal{C}_t| = 1 + \lfloor \log_2 t \rfloor$. The longest interval in $\mathcal{C}_t$ has length at most t, which is associated with at most $|\mathcal{S}(t)| = 1 + \lceil \frac{1}{2} \log_2 t \rceil$ experts. With $\mathcal{A}_t = \{ E(I, \eta) : t \in I \}$ representing the set of experts active in round t, the number of active experts in round t, denoted by $n_t = |\mathcal{A}_t|$ in Algorithm 1, satisfies

$$n_t \leq (1 + \lfloor \log_2 t \rfloor) \cdot \left(1 + \lceil \tfrac{1}{2} \log_2 t \rceil \right). \tag{26}$$

This ensures that the computational complexity of TEWA-SE is only $\mathcal{O}(\log^2 T)$ per round.

Tilted Exponentially Weighted Average In each round t, TEWA-SE aggregates the actions proposed by the active experts $E(I, \eta) \in \mathcal{A}_t$ using exponential weights, tilted by their respective learning rates, by computing

$$\boldsymbol{x}_t = \frac{\sum_{E(I,\eta) \in \mathcal{A}_t} \eta \exp(-L_{t-1,I}^\eta) \boldsymbol{x}_{t,I}^\eta}{\sum_{E(\tilde{I},\tilde{\eta}) \in \mathcal{A}_t} \tilde{\eta} \exp(-L_{t-1,\tilde{I}}^{\tilde{\eta}})}, \tag{27}$$

where for $I = [r, s]$ and $t \in [r+1, s]$, $L_{t-1,I}^\eta = \sum_{\tau=r}^{t-1} \ell_\tau^\eta(\boldsymbol{x}_{\tau,I}^\eta)$ represents the cumulative surrogate loss accrued by expert $E(I, \eta)$ over the interval $[r, t-1]$. Note that (27) is equivalent to line 7 of Algorithm 1, rewritten with notation better suited for our proof.

In what follows, we prove some theoretical guarantees for TEWA-SE.

C.2 Proof of Theorem 1

In this section, we first restate Theorem 1 and provide its complete proof, which relies on several supporting lemmas. For clarity of exposition, we defer the statements and proofs of these supporting lemmas to the following sections.

Theorem 1. *For any $T \in \mathbb{N}^+$ and $\mathsf{B} \in [T]$, Algorithm 1 with $h = \min(\sqrt{d}\mathsf{B}^{-\frac{1}{4}}, r)$ satisfies*

$$R^{\mathsf{ada}}(\mathsf{B}, T) \lesssim \sqrt{d}\mathsf{B}^{\frac{3}{4}} + d^2, \tag{14}$$

and if f_t is α-strongly-convex with $\arg\min_{\boldsymbol{x} \in \mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$ for all $t \in [T]$,[4] it furthermore holds that

$$R^{\mathsf{ada}}(\mathsf{B}, T) \lesssim \frac{d}{\alpha}\sqrt{\mathsf{B}} + \frac{1}{\alpha}d^2, \tag{15}$$

where $\lesssim$ conceals polylogarithmic terms in B and T, independent of d and α.

Proof of Theorem 1. We prove (14) for the general convex case and (15) for the strongly-convex case similarly. To bound $R^{\mathsf{ada}}(\mathsf{B}, T)$, we will uniformly bound $\sum_{t=p}^q \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})]$ across all comparators $\boldsymbol{u} \in \Theta$ and intervals $[p, q]$ shorter than B.

Common setup: Invoking the event $\Lambda_T = \{|\xi_t| \le 2\sigma\sqrt{\log(T+1)}, \forall t \in [T]\}$ defined above (10), since $\{\xi_t\}_{t=1}^T$ are σ-sub-Gaussian, we have

$$\mathbf{P}(\Lambda_T^c) \le \sum_{t=1}^T \mathbf{P}\left(|\xi_t| > 2\sigma\sqrt{\log(T+1)}\right) \le 2\sum_{t=2}^{T+1} T^{-2} = 2T^{-1}. \tag{28}$$

By the law of total expectation we can write for any $\boldsymbol{u} \in \Theta$,

$$\sum_{t=p}^q \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})] = \sum_{t=p}^q \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u}) \mid \Lambda_T]\mathbf{P}(\Lambda_T) + \sum_{t=p}^q \mathbf{E}[\underbrace{f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})}_{\le 2} \mid \Lambda_T^c]\underbrace{\mathbf{P}(\Lambda_T^c)}_{\le 2T^{-1}}$$

$$\le \sum_{t=p}^q \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u}) \mid \Lambda_T] + 4. \tag{29}$$

To bound the first term in the last display, we consider the following decomposition

$$\sum_{t=p}^q \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u}) \mid \Lambda_T]$$

$$= \underbrace{\sum_{t=p}^q \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{x}_t) \mid \Lambda_T]}_{\text{term I}} + \underbrace{\sum_{t=p}^q \mathbf{E}[f_t(\boldsymbol{x}_t) - \hat{f}_t(\boldsymbol{x}_t) \mid \Lambda_T]}_{\text{term II}} +$$

$$\underbrace{\sum_{t=p}^q \mathbf{E}[\hat{f}_t(\boldsymbol{x}_t) - \hat{f}_t(\boldsymbol{u}) \mid \Lambda_T]}_{\text{term III}} + \underbrace{\sum_{t=p}^q \mathbf{E}[\hat{f}_t(\boldsymbol{u}) - f_t(\boldsymbol{u}) \mid \Lambda_T]}_{\text{term IV}}. \tag{30}$$

Since f_t is convex, by Jensen's inequality we obtain that term II is negative (c.f. [71, Lemma A.2 (ii)]). In what follows, we bound terms I, III and IV in this decomposition separately for the general convex case and the strongly-convex case.

General convex and Lipschitz case: Recall that $(\boldsymbol{\zeta}_t)_{t=1}^T$ denote uniform samples from the unit sphere $\partial\mathbb{B}^d$, and $\tilde{\boldsymbol{\zeta}}$ denotes a uniform sample from the unit ball $\mathbb{B}^d$, while $\hat{f}_t(\boldsymbol{x}) = \mathbf{E}[f_t(\boldsymbol{x} + h\tilde{\boldsymbol{\zeta}})]\ \forall \boldsymbol{x} \in \tilde{\Theta}$. Since $(f_t)_{t=1}^T$ are K-Lipschitz, $\|\boldsymbol{\zeta}_t\| = 1$, and $\mathbf{E}[\|\tilde{\boldsymbol{\zeta}}\|] \le 1$, we can bound term I and term IV by

$$\text{term I} = \sum_{t=p}^q \mathbf{E}\left[\mathbf{E}[f_t(\boldsymbol{x}_t + h\boldsymbol{\zeta}_t)|\boldsymbol{x}_t] - f_t(\boldsymbol{x}_t) \mid \Lambda_T\right] \le K(q - p + 1)h, \tag{31}$$

$$\text{term IV} = \sum_{t=p}^q \mathbf{E}\left[\mathbf{E}[f_t(\boldsymbol{u} + h\tilde{\boldsymbol{\zeta}})] - f_t(\boldsymbol{u}) \mid \Lambda_T\right] \le K(q - p + 1)h. \tag{32}$$

---

[4]The assumption that loss minimizers lie inside Θ is common in zeroth-order optimization, see e.g., [7, 72, 73]. Without it, our upper bound analysis would have an extra term depending on the gradients at the minimizers.

To bound term III, recall that $\boldsymbol{g}_t$ denotes the gradient estimate of $\hat{f}_t$ at $\boldsymbol{x}_t$. We use the convexity of $\hat{f}_t$ and apply Lemma 3 to obtain that for any $\boldsymbol{u} \in \Theta$,

$$
\text{term III} \leq \mathbf{E}\left[\sum_{t=p}^{q} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u}\rangle \,\big|\, \Lambda_T\right]
$$

$$
\leq \mathbf{E}\left[10GDa_{p,q}b_{p,q} + 3G\sqrt{a_{p,q}b_{p,q}}\sqrt{\sum_{t=p}^{q} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2} \,\big|\, \Lambda_T\right]
$$

$$
\leq 10GDa_{p,q}b_{p,q} + 3GD\sqrt{a_{p,q}b_{p,q}}\sqrt{q-p+1}\,, \tag{33}
$$

where all constants are explicit in the statement of the lemma. By combining the bounds for all four terms in (30) with (29), and using $h = \min\left(\sqrt{d}\mathsf{B}^{-\frac{1}{4}}, r\right)$, $G = \frac{d}{h}(1 + 2\sigma\sqrt{\log(T+1)})$, $a_{p,q} = \frac{1}{2} + 2\log(2q) + \frac{1}{2}\log(q-p+1) \leq 6\log(T+1)$, and $b_{p,q} = 2\lceil\log_2(q-p+2)\rceil \leq 6\log(\mathsf{B}+1)$, we establish that

$$
R^{\mathsf{ada}}(\mathsf{B}, T) \leq 10GDa_{p,q}b_{p,q} + 3GD\sqrt{a_{p,q}b_{p,q}}\sqrt{q-p+1} + 2K(q-p+1)h + 4
$$

$$
\leq \begin{cases} C\sqrt{d}\mathsf{B}^{\frac{3}{4}} + 4 & \text{if } h = \sqrt{d}\mathsf{B}^{-\frac{1}{4}} \\ C_1 d^2 + C_2 d + 4 & \text{if } h = r \end{cases}
$$

$$
\leq C\sqrt{d}\mathsf{B}^{\frac{3}{4}} + C_1 d^2 + C_2 d + 4\,, \tag{34}
$$

where $C, C_1, C_2 > 0$ are polylogarithmic in T, independent of d, defined with $M_T = 1 + 2\sigma\sqrt{\log(T+1)}$ and $N_{T,\mathsf{B}} = \log(T+1)\log(\mathsf{B}+1)$ as

$$
C = 18DM_T\left(\sqrt{N_{T,\mathsf{B}}} + 20N_{T,\mathsf{B}}\mathsf{B}^{-\frac{1}{2}}\right) + 2K \tag{35}
$$

$$
C_1 = \left(18DM_T\sqrt{N_{T,\mathsf{B}}} + 2K\right)/r^3 \tag{36}
$$

$$
C_2 = 360DM_T N_{T,\mathsf{B}}/r\,. \tag{37}
$$

This concludes the proof of (14).

Strongly-convex and smooth case: Due to the β-smoothness of f_t and the fact that $\mathbf{E}[\boldsymbol{\zeta}_t] = \mathbf{E}[\tilde{\boldsymbol{\zeta}}] = \mathbf{0}$ and $\mathbf{E}[\|\tilde{\boldsymbol{\zeta}}\|^2] \leq \mathbf{E}[\|\boldsymbol{\zeta}_t\|^2] = 1$, we can bound term I and term IV each by $\frac{\beta}{2}(q-p+1)h^2$. When the f_t's are strongly-convex, we can derive a tighter bound on term III than that in (33) by restricting the comparator $\boldsymbol{u}$ to the clipped domain $\tilde{\Theta}$ and using the fact that when f_t is α-strongly convex on Θ, $\hat{f}_t$ is α-strongly convex on $\tilde{\Theta}$ (c.f. [71, Lemma A.3]). That is, we have for any $\boldsymbol{u} \in \tilde{\Theta}$,

$$
\text{term III} \leq \mathbf{E}\left[\sum_{t=p}^{q} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u}\rangle - \frac{\alpha}{2}\sum_{t=p}^{q} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2 \,\big|\, \Lambda_T\right]
$$

$$
\leq \mathbf{E}\left[10GDa_{p,q}b_{p,q} + \underbrace{3G\sqrt{a_{p,q}b_{p,q}}\sqrt{\sum_{t=p}^{q} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2} - \frac{\alpha}{2}\sum_{t=p}^{q} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2}_{=:\delta} \,\big|\, \Lambda_T\right]
$$

$$
\leq \left(10GD + \frac{18}{\alpha}G^2\right)a_{p,q}b_{p,q}\,, \tag{38}
$$

where the last inequality holds because term δ is uniformly bounded as follows:

$$
\delta \leq \begin{cases} \frac{18}{\alpha}G^2 a_{p,q}b_{p,q} & \text{if } \sqrt{\sum_{t=p}^{q} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2} \leq \frac{6}{\alpha}G\sqrt{a_{p,q}b_{p,q}}\,, \\ 0 & \text{otherwise.} \end{cases}
$$

Combining (38) with (29)–(30) and simplifying yields for any $\boldsymbol{u} \in \tilde{\Theta}$,

$$
\sum_{t=p}^{q} \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})] \leq \left(10GD + \frac{18}{\alpha}G^2\right)a_{p,q}b_{p,q} + \beta(q-p+1)h^2 + 4\,. \tag{39}
$$

The final step is to handle the case where the comparator $\boldsymbol{u} \in \Theta \setminus \tilde{\Theta}$. Consider the worst case when the comparator is $\boldsymbol{u}^* \in \arg\min_{\boldsymbol{u} \in \Theta} \sum_{t=p}^{q} f_t(\boldsymbol{u})$ with $\boldsymbol{u}^* \in \Theta \setminus \tilde{\Theta}$. Let $\tilde{\boldsymbol{u}}^* = \Pi_{\tilde{\Theta}}(\boldsymbol{u}^*)$. If

27

$\arg\min_{\boldsymbol{x}\in\mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta \; \forall t \in [T]$, then by the β-smoothness of the f_t's we have

$$\sum_{t=p}^{q} f_t(\tilde{\boldsymbol{u}}^*) - f_t(\boldsymbol{u}^*) \leq \sum_{t=p}^{q} \left[\left\langle \underbrace{\nabla f_t(\boldsymbol{u}^*)}_{=0}, \tilde{\boldsymbol{u}}^* - \boldsymbol{u}^* \right\rangle + \frac{\beta}{2} \underbrace{\|\tilde{\boldsymbol{u}}^* - \boldsymbol{u}^*\|^2}_{\leq h^2} \right] = \frac{\beta}{2}(q - p + 1)h^2. \quad (40)$$

Combining (39) with (40) yields

$$\begin{aligned} R^{\mathsf{ada}}(\mathsf{B}, T) &\leq \left(10GD + \tfrac{18}{\alpha}G^2\right) a_{p,q} b_{p,q} + \tfrac{3}{2}\beta(q - p + 1)h^2 + 4 \\ &\leq \begin{cases} C'd\sqrt{\mathsf{B}} + 4 & \text{if } h = \sqrt{d}\mathsf{B}^{-\frac{1}{4}} \\ C_1'd^2 + C_2'd + 4 & \text{if } h = r \end{cases} \\ &\leq C'\tfrac{d}{\alpha}\sqrt{\mathsf{B}} + C_1'\tfrac{1}{\alpha}d^2 + C_2'd + 4 \,, \end{aligned} \quad (41)$$

where $C', C_1', C_2' > 0$ are polylogarithmic in T and B, independent of d, defined as

$$C' = 72\left(9M_T + 5\alpha D\mathsf{B}^{-\frac{1}{4}}\right) M_T N_{T,\mathsf{B}} + \tfrac{3}{2}\beta\alpha \quad (42)$$

$$C_1' = \left(648 M_T^2 N_{T,\mathsf{B}} + \beta\alpha\right)/r^2 \quad (43)$$

$$C_2' = 360 D M_T N_{T,\mathsf{B}}/r \,. \quad (44)$$

This concludes the proof of (15). $\qquad\square$

The proof above crucially relies on Lemma 3, which we state and prove in the following section.

C.3 Upper bounds on linearized regret

Lemma 3 establishes an upper bound on the linearized regret associated with the smoothed loss $\hat{f}_t$ for any arbitrary interval $I = [p, q] \subseteq [1, T]$. This result builds on two key components: Lemma 4, which characterizes how a given arbitrary interval is covered by a sequence of GC intervals, and Lemma 5, which provides an upper bound on the linearized regret for each GC interval $I \in \mathcal{I}$. For clarity, we first present and prove Lemma 3, then proceed to detail the supporting Lemmas 4 and 5.

Lemma 3 (Linearized regret on an arbitrary interval). *For an arbitrary interval $I = [p, q] \subseteq [1, T]$, Algorithm 1 satisfies for all $\boldsymbol{u} \in \Theta$,*

$$\sum_{t=p}^{q} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \rangle \leq 10GD a_{p,q} b_{p,q} + 3G\sqrt{a_{p,q} b_{p,q}} \sqrt{\sum_{t=p}^{q} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2}, \quad (45)$$

where $a_{p,q} = \frac{1}{2} + 2\log(2q) + \frac{1}{2}\log(q - p + 1)$ and $b_{p,q} = 2\lceil \log_2(q - p + 2) \rceil$.

Proof of Lemma 3. This proof follows similar arguments to those used in proving the first part of [24, Theorem 2]. To begin, according to Lemma 4, any arbitrary interval $I = [p, q] \subseteq [1, T]$ can be covered by two sequences of consecutive and disjoint GC intervals, denoted by $I_{-m}, \ldots, I_0 \in \mathcal{I}$ and $I_1, \ldots, I_n \in \mathcal{I}$, where $n, m \in \mathbb{N}^+$ with $n \leq \lceil \log_2(q - p + 2) \rceil$ and $m + 1 \leq \lceil \log_2(q - p + 2) \rceil$. Note that negative indices correspond to GC intervals that precede I_0, while positive indices correspond to intervals that follow it. The indices indicate temporal ordering and are unrelated to the length of the intervals.

By applying the linearized regret bound from Lemma 5 to each GC interval, and noticing that $a_{r,s} \leq a_{p,q}$ for any subinterval $[r, s] \subseteq [p, q]$ (as evident from the definition of $a_{p,q}$ in (45)), we

establish for all $\boldsymbol{u} \in \Theta$,

$$\sum_{t=p}^{q} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \rangle = \sum_{i=-m}^{n} \sum_{t \in I_i} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \rangle$$

$$\leq \sum_{i=-m}^{n} \left[3G \sqrt{a_{p,q} \sum_{t \in I_i} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2} + 10GDa_{p,q} \right]$$

$$= 10GDa_{p,q}(n + m + 1) + 3G\sqrt{a_{p,q}} \sum_{i=-m}^{n} \sqrt{\sum_{t \in I_i} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2}$$

$$\leq 10GDa_{p,q}(n + m + 1) + 3G\sqrt{a_{p,q}} \sqrt{(n + m + 1) \sum_{i=-m}^{n} \sum_{t \in I_i} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2}$$

$$\leq 10GDa_{p,q}b_{p,q} + 3G\sqrt{a_{p,q}b_{p,q}} \sqrt{\sum_{t=p}^{q} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2}, \tag{46}$$

where the last step uses $n + m + 1 \leq 2\lceil \log_2(q - p + 2) \rceil =: b_{p,q}$. $\qquad\square$

We now present Lemmas 4 and 5 which we used to prove Lemma 3 above.

Lemma 4 (Covering property of GC intervals)**.** *Any arbitrary interval $I = [p, q] \subseteq \mathbb{N}^+$ can be partitioned into two finite sequences of consecutive and disjoint GC intervals, denoted by $I_{-m}, \ldots, I_0 \in \mathcal{I}$ and $I_1, \ldots, I_n \in \mathcal{I}$, where $I = \bigcup_{i=-m}^{n} I_i$, such that*

$$\frac{|I_{-i}|}{|I_{-i+1}|} \leq \tfrac{1}{2} \quad \forall i \geq 1, \quad and \quad \frac{|I_i|}{|I_{i-1}|} \leq \tfrac{1}{2}, \quad \forall i \geq 2, \tag{47}$$

with

$$n \leq \lceil \log_2(q - p + 2) \rceil, \quad and \quad m + 1 \leq \lceil \log_2(q - p + 2) \rceil. \tag{48}$$

Proof of Lemma 4. Eq. (47) directly comes from [19, Lemma 1.2]. To prove (48), suppose for contradiction $n > \lceil \log_2(q - p + 2) \rceil$, then we have

$$\sum_{i=1}^{n} |I_i| \geq \sum_{i=1}^{n} 2^{i-1} = 2^n - 1 > q - p + 1 = |I|, \tag{49}$$

contradicting the fact that $\bigcup_{i=-m}^{n} I_i = I$. By the same reasoning, we have $m + 1 \leq \lceil \log_2(q - p + 2) \rceil$. $\qquad\square$

Lemma 5 (Linearized regret on a GC interval)**.** *For any GC interval $I = [r, s] \in \mathcal{I}$, Algorithm 1 satisfies for all $\boldsymbol{u} \in \Theta$,*

$$\sum_{t=r}^{s} \langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \rangle \leq 3G \sqrt{a_{r,s} \sum_{t=r}^{s} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2} + 10GDa_{r,s}, \tag{50}$$

where $a_{r,s} = \frac{1}{2} + 2\log(2s) + \frac{1}{2}\log(s - r + 1)$.

Proof of Lemma 5. This proof is similar to that of [24, Lemma 12]. For any GC interval $I = [r, s] \in \mathcal{I}$ and learning rate $\eta \in \mathcal{S}(s - r + 1)$, we can apply the definition of surrogate loss ℓ_t^η from (11),

noticing that $\ell_t^\eta(\boldsymbol{x}_t) = 0$, to obtain for all $\boldsymbol{u} \in \Theta$,

$$\sum_{t=r}^{s} \eta \left\langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \right\rangle - \sum_{t=r}^{s} \eta^2 G^2 \|\boldsymbol{x}_t - \boldsymbol{u}\|^2$$

$$= \sum_{t=r}^{s} \mathbf{E}[-\ell_t^\eta(\boldsymbol{u}) \mid \boldsymbol{x}_t, \Lambda_T]$$

$$= \underbrace{\sum_{t=r}^{s} \mathbf{E}\left[\ell_t^\eta(\boldsymbol{x}_t) - \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) \mid \boldsymbol{x}_t, \Lambda_T\right]}_{\text{meta-regret} \leq 2\log(2s)} + \underbrace{\sum_{t=r}^{s} \mathbf{E}\left[\ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) - \ell_t^\eta(\boldsymbol{u}) \mid \boldsymbol{x}_t, \Lambda_T\right]}_{\text{expert-regret} \leq \frac{1}{2} + \frac{1}{2}\log(s-r+1)}$$

$$\leq 2\log(2s) + \tfrac{1}{2} + \tfrac{1}{2}\log(s - r + 1) =: a_{r,s}, \tag{51}$$

where the last step applies the upper bound on the expert-regret established in Lemma 6 and the upper bound on the meta-regret in Lemma 7, both of which we defer to Section C.4. Eq. (51) can be rearranged into

$$\sum_{t=r}^{s} \left\langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \right\rangle \leq \frac{a_{r,s}}{\eta} + \eta G^2 \sum_{t=r}^{s} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2. \tag{52}$$

The optimal value of η that minimizes the RHS of (52) is

$$\eta^* = \sqrt{\frac{a_{r,s}}{G^2 \sum_{t=r}^{s} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2}}. \tag{53}$$

Note that since $a_{r,s} \geq \frac{1}{2}$, $\eta^* \geq \frac{1}{GD\sqrt{2(s-r+1)}}$ for all $\boldsymbol{x} \in \Theta$. The next step is to select a value η from the set $\mathcal{S}(s - r + 1) = \left\{\frac{2^{-i}}{5GD} : i \in \left\{0, 1, \ldots, \left\lceil \frac{1}{2}\log_2(s - r + 1)\right\rceil\right\}\right\}$ that best approximates η^*. Two cases arises:

i) If $\eta^* \leq \frac{1}{5GD}$, there must exist an $\eta \in \mathcal{S}(s - r + 1)$ such that $\frac{\eta^*}{2} \leq \eta \leq \eta^*$. Substituting this choice of η into (52) gives

$$\sum_{t=r}^{s} \left\langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \right\rangle \leq \frac{2a_{r,s}}{\eta^*} + \eta^* G^2 \sum_{t=r}^{s} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2 = 3G\sqrt{a_{r,s} \sum_{t=r}^{s} \|\boldsymbol{x}_t - \boldsymbol{u}\|^2}. \tag{54}$$

ii) If $\eta^* > \frac{1}{5GD}$, then the best choice of $\eta \in \mathcal{S}(s - r + 1)$ is $\eta = \frac{1}{5GD}$, which leads to

$$\sum_{t=r}^{s} \left\langle \mathbf{E}[\boldsymbol{g}_t | \boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{u} \right\rangle \leq a_{r,s} \cdot 5GD \cdot 2 = 10GDa_{r,s}. \tag{55}$$

Combining (54)–(55) concludes the proof. $\qquad\square$

The proof of Lemma 5 above relied on the upper bounds on the expert-regret and meta-regret from Lemmas 6 and 7. We present and prove these lemmas in the following section.

C.4 Upper bounds on expert-regret and meta-regret

Lemma 6 (Expert-regret). *For any GC interval $I = [r, s] \in \mathcal{I}$ and learning rate $\eta \in \mathcal{S}(s - r + 1)$, Algorithm 1 satisfies for all $\boldsymbol{u} \in \tilde{\Theta}$,*

$$\sum_{t=r}^{s} \mathbf{E}\left[\ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) - \ell_t^\eta(\boldsymbol{u}) \mid \boldsymbol{x}_t, \boldsymbol{x}_{t,I}^\eta, \Lambda_T\right] \leq \tfrac{1}{2} + \tfrac{1}{2}\log(s - r + 1). \tag{56}$$

Proof of Lemma 6. The proof follows standard convergence analysis of projected online gradient descent for strongly convex objective functions, see e.g., [74, Theorem 1]. For any time step $t \in I$, the surrogate loss ℓ_t^η associated with the expert with learning rate η and lifetime $I = [r, s]$ serves as our strongly-convex objective function. By applying the definition of ℓ_t^η, we have for all $\boldsymbol{x} \in \Theta$,

$$\mathbf{E}\left[\|\nabla \ell_t^\eta(\boldsymbol{x})\|^2 \mid \boldsymbol{x}_t, \Lambda_T\right] = \mathbf{E}\left[\|\eta \boldsymbol{g}_t + 2\eta^2 G^2 (\boldsymbol{x} - \boldsymbol{x}_t)\|^2 \mid \boldsymbol{x}_t, \Lambda_T\right]$$

$$\leq \mathbf{E}\left[\left(\|\eta \boldsymbol{g}_t\| + \|2\eta^2 G^2 (\boldsymbol{x} - \boldsymbol{x}_t)\|\right)^2 \mid \boldsymbol{x}_t, \Lambda_T\right] \qquad \leq (G')^2, \qquad (57)$$

where we introduced $G' = \eta G + 2\eta^2 G^2 D$. By the update rule of our projected online gradient descent with step size μ_t (line 4 of Algorithm 2), we have for all $\boldsymbol{u} \in \tilde{\Theta}$,

$$\|\boldsymbol{x}_{t+1,I}^\eta - \boldsymbol{u}\|^2 = \left\|\Pi_{\tilde{\Theta}}\left(\boldsymbol{x}_{t,I}^\eta - \mu_t \nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta)\right) - \boldsymbol{u}\right\|^2$$

$$\leq \left\|\boldsymbol{x}_{t,I}^\eta - \mu_t \nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) - \boldsymbol{u}\right\|^2$$

$$= \|\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u}\|^2 + \mu_t^2 \|\nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta)\|^2 - 2\mu_t (\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u})^\top \nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta),$$

which can be rearranged into

$$2(\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u})^\top \nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) \leq \frac{\|\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u}\|^2 - \|\boldsymbol{x}_{t+1,I}^\eta - \boldsymbol{u}\|^2}{\mu_t} + \mu_t \|\nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta)\|^2. \qquad (58)$$

Define shorthand $\lambda \equiv 2\eta^2 G^2$ and recall $\mu_t = 1/(\lambda(t - r + 1))$, then Eq. (58) implies that

$$2\sum_{t=r}^s (\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u})^\top \nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) - \lambda \sum_{t=r}^s \|\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u}\|^2$$

$$\leq \sum_{t=r}^s \frac{\|\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u}\|^2 - \|\boldsymbol{x}_{t+1,I}^\eta - \boldsymbol{u}\|^2}{\mu_t} + \sum_{t=r}^s \mu_t \|\nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta)\|^2 - \lambda \sum_{t=r}^s \|\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u}\|^2$$

$$= \sum_{t=r+1}^s \|\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u}\|^2 \underbrace{\left(\frac{1}{\mu_t} - \frac{1}{\mu_{t-1}} - \lambda\right)}_{=0} + \|\boldsymbol{x}_{r,I}^\eta - \boldsymbol{u}\|^2 \underbrace{\left(\frac{1}{\mu_r} - \lambda\right)}_{=0} - \underbrace{\frac{\|\boldsymbol{x}_{s+1,I}^\eta - \boldsymbol{u}\|^2}{\mu_s}}_{\geq 0}$$

$$+ \sum_{t=r}^s \mu_t \|\nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta)\|^2 \leq \sum_{t=r}^s \mu_t \|\nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta)\|^2. \qquad (59)$$

Noticing that with any given $\boldsymbol{x}_t \in \mathbb{R}^d$, ℓ_t^η is λ-strongly-convex, we apply (59) to obtain that for all $\boldsymbol{u} \in \tilde{\Theta}$,

$$2\sum_{t=r}^s \mathbf{E}\left[\ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) - \ell_t^\eta(\boldsymbol{u}) \mid \boldsymbol{x}_t, \boldsymbol{x}_{t,I}^\eta, \Lambda_T\right]$$

$$\leq \mathbf{E}\left[2\sum_{t=r}^s (\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u})^\top \nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) - \lambda \sum_{t=r}^s \|\boldsymbol{x}_{t,I}^\eta - \boldsymbol{u}\|^2 \,\bigg|\, \{\boldsymbol{x}_t, \boldsymbol{x}_{t,I}^\eta\}_{t=r}^s, \Lambda_T\right]$$

$$\leq \sum_{t=r}^s \mu_t \mathbf{E}\left[\|\nabla \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta)\|^2 \mid \boldsymbol{x}_t, \boldsymbol{x}_{t,I}^\eta, \Lambda_T\right]$$

$$\overset{(i)}{\leq} (G')^2 \sum_{t=r}^s \mu_t \overset{(ii)}{\leq} \frac{(G')^2}{\lambda}\left(1 + \log(s - r + 1)\right) \overset{(iii)}{\leq} 1 + \log(s - r + 1), \qquad (60)$$

where (i) is a result of (57), (ii) uses the bound $\sum_{k=1}^n \frac{1}{k} \leq 1 + \log n$ for any $n \in \mathbb{N}^+$, and (iii) uses the fact that given $\eta \leq \frac{1}{5GD}$ it holds that

$$(G')^2 = \left(\eta G + 2\eta^2 G^2 D\right)^2 = \eta^2 G^2 + 4\eta^3 G^3 D + 4\eta^4 G^4 D^2 \leq \eta^2 G^2 + \tfrac{4}{5}\eta^2 G^2 + \tfrac{4}{25}\eta^2 G^2 \leq \lambda.$$

$\square$

Lemma 7 (Meta-regret). *For any GC interval $I = [r, s] \in \mathcal{I}$ and learning rate $\eta \in \mathcal{S}(s - r + 1)$, Algorithm 1 satisfies*

$$\sum_{t=r}^s \mathbf{E}\left[\ell_t^\eta(\boldsymbol{x}_t) - \ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) \mid \boldsymbol{x}_t, \boldsymbol{x}_{t,I}^\eta, \Lambda_T\right] = -\sum_{t=r}^s \mathbf{E}\left[\ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) \mid \boldsymbol{x}_t, \boldsymbol{x}_{t,I}^\eta, \Lambda_T\right] \leq 2\log(2s). \quad (61)$$

Proof of Lemma 7. The proof is similar to that of [24, Lemma 6]. By Jensen's inequality and the convexity of norms, we have for all $\boldsymbol{x} \in \Theta$,

$$
\begin{aligned}
|\langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}\rangle| &\leq \|\mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T]\| \|\boldsymbol{x}_t - \boldsymbol{x}\| \\
&\leq \mathbf{E}\left[\|\boldsymbol{g}_t\| \mid \boldsymbol{x}_t, \Lambda_T\right] \|\boldsymbol{x}_t - \boldsymbol{x}\| \leq GD,
\end{aligned} \tag{62}
$$

which, given $\eta \leq \frac{1}{5GD}$, implies that

$$
\eta \langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}\rangle \geq -\eta GD \geq -\frac{1}{5}. \tag{63}
$$

Using (62)–(63) and applying the inequality $\ln(1 + z) \geq z - z^2$ for any $z \geq -\frac{2}{3}$ with $z = \eta \langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}\rangle$, we obtain for all $\boldsymbol{x} \in \Theta$,

$$
\begin{aligned}
\exp\left(-\mathbf{E}[\ell_t^\eta(\boldsymbol{x}) \mid \boldsymbol{x}_t, \Lambda_T]\right) &= \exp\left(\eta \langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}\rangle - \eta^2 G^2 \|\boldsymbol{x}_t - \boldsymbol{x}\|^2\right) \\
&\leq \exp\left(\eta \langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}\rangle - \eta^2 \langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}\rangle^2\right) \\
&\leq 1 + \eta \langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}\rangle.
\end{aligned} \tag{64}
$$

Define shorthand $\mathbb{F}_{t,I}^\eta = \{\boldsymbol{x}_t, \boldsymbol{x}_{t,I}^\eta, \Lambda_T\}$, and $\mathbb{H}_{t,I}^\eta = \cup_{\tau \in [t]} \mathbb{F}_{\tau,I}^\eta$ for $t \in [T]$. Using (64), we can write for every $t \in [T]$,

$$
\begin{aligned}
\sum_{E(I,\eta) \in \mathcal{A}_t} &\exp\left(-\mathbf{E}[L_{t,I}^\eta \mid \mathbb{H}_{t,I}^\eta]\right) \\
&= \sum_{E(I,\eta) \in \mathcal{A}_t} \exp\left(-\mathbf{E}[L_{t-1,I}^\eta \mid \mathbb{H}_{t-1,I}^\eta]\right) \exp\left(-\mathbf{E}[\ell_t^\eta(\boldsymbol{x}_{t,I}^\eta) \mid \mathbb{F}_{t,I}^\eta]\right) \\
&\leq \sum_{E(I,\eta) \in \mathcal{A}_t} \exp\left(-\mathbf{E}[L_{t-1,I}^\eta \mid \mathbb{H}_{t-1,I}^\eta]\right) \left[1 + \eta \left\langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}_{t,I}^\eta\right\rangle\right].
\end{aligned} \tag{65}
$$

The second term on the RHS can be bounded as follows:

$$
\begin{aligned}
\sum_{E(I,\eta) \in \mathcal{A}_t} &\exp(-\mathbf{E}[L_{t-1,I}^\eta \mid \mathbb{H}_{t-1,I}^\eta]) \left[\eta \left\langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \boldsymbol{x}_t - \boldsymbol{x}_{t,I}^\eta\right\rangle\right] \\
&= \left\langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \sum_{E(I,\eta) \in \mathcal{A}_t} \eta \exp(-\mathbf{E}[L_{t-1,I}^\eta \mid \mathbb{H}_{t-1,I}^\eta])(\boldsymbol{x}_t - \boldsymbol{x}_{t,I}^\eta)\right\rangle \\
&\overset{(i)}{\leq} \left\langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \sum_{E(I,\eta) \in \mathcal{A}_t} \eta \mathbf{E}\left[\exp(-L_{t-1,I}^\eta) \mid \mathbb{H}_{t-1,I}^\eta\right](\boldsymbol{x}_t - \boldsymbol{x}_{t,I}^\eta)\right\rangle \\
&= \left\langle \mathbf{E}[\boldsymbol{g}_t|\boldsymbol{x}_t, \Lambda_T], \mathbf{E}\Big[\underbrace{\sum_{E(I,\eta) \in \mathcal{A}_t} \eta \exp(-L_{t-1,I}^\eta)(\boldsymbol{x}_t - \boldsymbol{x}_{t,I}^\eta)}_{=0} \Big| \mathbb{H}_{t-1,I}^\eta\Big]\right\rangle \overset{(ii)}{=} 0,
\end{aligned} \tag{66}
$$

where (i) applies Jensen's inequality, and (ii) is due to the update rule of $\boldsymbol{x}_t$ in (27). Combining (65)–(66) yields

$$
\sum_{E(I,\eta) \in \mathcal{A}_t} \exp\left(-\mathbf{E}[L_{t,I}^\eta \mid \mathbb{H}_{t,I}^\eta]\right) \leq \sum_{E(I,\eta) \in \mathcal{A}_t} \exp\left(-\mathbf{E}[L_{t-1,I}^\eta \mid \mathbb{H}_{t-1,I}^\eta]\right). \tag{67}
$$

By summing both sides of (67) over $t = 1, \ldots, s$ and rewriting, we obtain

$$
\begin{aligned}
&\sum_{E(I,\eta) \in \mathcal{A}_s} \exp\left(-\mathbf{E}[L_{s,I}^\eta \mid \mathbb{H}_{s,I}^\eta]\right) + \sum_{t=1}^{s-1} \sum_{E(I,\eta) \in \mathcal{A}_t \setminus \mathcal{A}_{t+1}} \exp\left(-\mathbf{E}[L_{t,I}^\eta \mid \mathbb{H}_{t,I}^\eta]\right) + \\
&\sum_{t=1}^{s-1} \sum_{E(I,\eta) \in \mathcal{A}_t \cap \mathcal{A}_{t+1}} \exp\left(-\mathbf{E}[L_{t,I}^\eta \mid \mathbb{H}_{t,I}^\eta]\right) \\
&\leq \sum_{E(I,\eta) \in \mathcal{A}_1} \exp\left(-\mathbf{E}[L_{0,I}^\eta]\right) + \sum_{t=2}^{s} \sum_{E(I,\eta) \in \mathcal{A}_t \setminus \mathcal{A}_{t-1}} \exp\left(-\mathbf{E}[L_{t-1,I}^\eta \mid \mathbb{H}_{t-1,I}^\eta]\right) + \\
&\sum_{t=2}^{s} \sum_{E(I,\eta) \in \mathcal{A}_t \cap \mathcal{A}_{t-1}} \exp\left(-\mathbf{E}[L_{t-1,I}^\eta \mid \mathbb{H}_{t-1,I}^\eta]\right).
\end{aligned} \tag{68}
$$

Canceling the equivalent last terms on both sides of (68) and noting that $L_{\tau,I}^{\eta} = 0$ for $\tau = \min\{t : t \in I\} - 1$ by construction (see line 4 of Algorithm 1), we obtain for $s \geq 1$,

$$
\sum_{E(I,\eta)\in\mathcal{A}_s} \exp\left(-\mathbf{E}[L_{s,I}^{\eta} \mid \mathbb{H}_{s,I}^{\eta}]\right) + \sum_{t=1}^{s-1}\left[\sum_{E(I,\eta)\in\mathcal{A}_t\setminus\mathcal{A}_{t+1}} \exp\left(-\mathbf{E}[L_{t,I}^{\eta} \mid \mathbb{H}_{t,I}^{\eta}]\right)\right]
$$

$$
\leq \sum_{E(I,\eta)\in\mathcal{A}_1} \exp\left(\mathbf{E}[\underbrace{L_{0,I}^{\eta}}_{=0}]\right) + \sum_{t=2}^{s}\left[\sum_{E(I,\eta)\in\mathcal{A}_t\setminus\mathcal{A}_{t-1}} \exp\left(-\mathbf{E}[\underbrace{L_{t-1,I}^{\eta}}_{=0} \mid \mathbb{H}_{t-1,I}^{\eta}]\right)\right]
$$

$$
= \sum_{E(I,\eta)\in\mathcal{A}_1} \exp(0) + \sum_{t=2}^{s}\sum_{E(I,\eta)\in\mathcal{A}_t\setminus\mathcal{A}_{t-1}} \exp(0)
$$

$$
= |\mathcal{A}_1| + \sum_{t=2}^{s}|\mathcal{A}_t\setminus\mathcal{A}_{t-1}| \leq \sum_{t=1}^{s}|\mathcal{A}_t|
$$

$$
\overset{(i)}{\leq} \sum_{t=1}^{s}(1+\lfloor\log_2 t\rfloor)\cdot\left(1+\lceil\tfrac{1}{2}\log_2 t\rceil\right) \leq \sum_{t=1}^{s}(1+\log_2 t)^2 \overset{(ii)}{\leq} 4s^2\,, \tag{69}
$$

where (i) applies (26), and (ii) is due to $1 + \log_2 t \leq 2\sqrt{t} \ \forall t \geq 1$. Since $\exp(x) > 0$ for $x \in \mathbb{R}$, Eq. (69) implies that for any GC interval $I = [r, s] \in \mathcal{I}$ and learning rate $\eta \in \mathcal{S}(|I|)$,

$$
\exp\left(-\mathbf{E}[L_{s,I}^{\eta} \mid \mathbb{H}_{s,I}^{\eta}]\right) = \exp\left(-\sum_{t=r}^{s}\mathbf{E}[\ell_t^{\eta}(\boldsymbol{x}_{t,I}^{\eta}) \mid \mathbb{F}_{t,I}^{\eta}]\right) \leq 4s^2. \tag{70}
$$

Taking the logarithm of both sides completes the proof. □

C.5   Proof of Corollary 1

We first restate Corollary 1 and then provide the proof. Recall that for clarity we drop the $\lceil\cdot\rceil$ operators from the expressions for B and assume without loss of generality the expressions take integer values.

Corollary 1. *Consider any horizon $T \in \mathbb{N}^+$ and assume that, for all $t \in [T]$, the loss f_t is convex, or strongly-convex with $\arg\min_{\boldsymbol{x}\in\mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$. We refer to the second scenario as the strongly-convex (SC) case. Then, Algorithm 1 tuned with parameter B satisfies the following regret guarantees:*

Switching. $\mathsf{B} = \frac{T}{S} \Longrightarrow R^{\mathsf{swi}}(T, S) \lesssim \begin{cases} \sqrt{d}S^{\frac{1}{4}}T^{\frac{3}{4}} + d^2 S \\ d\sqrt{ST} + d^2 S \quad (SC) \end{cases}$

Dynamic. $\begin{cases} \mathsf{B} = \frac{T}{S} \vee \left(\frac{\sqrt{d}T}{\Delta}\right)^{\frac{4}{5}} \Rightarrow R^{\mathsf{dyn}}(T, \Delta, S) \lesssim R^{\mathsf{swi}}(T, S) \wedge (d^{\frac{2}{5}}\Delta^{\frac{1}{5}}T^{\frac{4}{5}} + d^{\frac{8}{5}}\Delta^{\frac{4}{5}}T^{\frac{1}{5}}) \\ \mathsf{B} = \frac{T}{S} \vee \left(\frac{dT}{\Delta}\right)^{\frac{2}{3}} \Rightarrow R^{\mathsf{dyn}}(T, \Delta, S) \lesssim R^{\mathsf{swi}}(T, S) \wedge (d^{\frac{2}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}} + d^{\frac{4}{3}}\Delta^{\frac{2}{3}}T^{\frac{1}{3}}) \quad (SC) \end{cases}$

Path-length. $\begin{cases} \mathsf{B} = \left(\frac{r\sqrt{d}T}{P}\right)^{\frac{4}{5}} \Rightarrow R^{\mathsf{path}}(T, P) \lesssim r^{-\frac{1}{5}}d^{\frac{2}{5}}P^{\frac{1}{5}}T^{\frac{4}{5}} + r^{-\frac{4}{5}}d^{\frac{8}{5}}P^{\frac{4}{5}}T^{\frac{1}{5}} \\ \mathsf{B} = \left(\frac{rdT}{P}\right)^{\frac{2}{3}} \Rightarrow R^{\mathsf{path}}(T, P) \lesssim r^{-\frac{1}{3}}d^{\frac{2}{3}}P^{\frac{1}{3}}T^{\frac{2}{3}} + r^{-\frac{2}{3}}d^{\frac{4}{3}}P^{\frac{2}{3}}T^{\frac{1}{3}} \quad (SC). \end{cases}$

Proof of Corollary 1. We begin by applying the first result in Proposition 1 with the adaptive regret guarantees in Theorem 1 to obtain switching regret guarantees. For known S, Algorithm 1 with parameter $\mathsf{B} = \frac{T}{S}$ achieves in the general convex case,

$$
R^{\mathsf{swi}}(T, S) \leq 2C\sqrt{d}S^{\frac{1}{4}}T^{\frac{1}{4}} + 2(C_1 + \tfrac{C_2}{d} + \tfrac{4}{d^2})d^2 S\,, \tag{71}
$$

and in the case where f_t is α-strongly-convex and $\arg\min_{\boldsymbol{x}\in\mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$ for all $t \in [T]$,

$$
R^{\mathsf{swi}}(T, S) \leq 2C'd\sqrt{ST} + 2(C_1' + \tfrac{C_2'}{d} + \tfrac{4}{d^2})d^2 S\,, \tag{72}
$$

where $C, C_1, C_2, C', C_1', C_2' > 0$ are the terms defined in (35)–(37) and (42)–(44) which are polylog-arithmic in T and B. When S and Δ are both known, we use (71)–(72) and apply the second result in Proposition 1 to bound $R^{\mathsf{dyn}}(T, \Delta, S)$. Specifically, for general convex losses, Algorithm 1 with $\mathsf{B} = \frac{T}{S} \vee \left(\frac{\sqrt{d}T}{\Delta}\right)^{\frac{4}{5}}$ yields

$$
R^{\mathsf{dyn}}(T, \Delta, S) \leq R^{\mathsf{swi}}(T, S) \wedge F^{\mathsf{dyn}}(T, \Delta)\,,
$$

where $F^{\mathsf{dyn}}(T,\Delta) := (2C+1)d^{\frac{2}{5}}\Delta^{\frac{1}{5}}T^{\frac{4}{5}} + 2(C_1 + \frac{C_2}{d} + \frac{4}{d^2})d^{\frac{8}{5}}\Delta^{\frac{4}{5}}T^{\frac{1}{5}}$. For strongly-convex losses with minimizers inside Θ, Algorithm 1 with $\mathsf{B} = \frac{T}{S} \vee \left(\frac{dT}{\Delta}\right)^{\frac{2}{3}}$ gives

$$R^{\mathsf{dyn}}(T,\Delta,S) \leq R^{\mathsf{swi}}(T,S) \wedge F^{\mathsf{dyn}}_{\mathsf{sc}}(T,\Delta)\,.$$

where $F^{\mathsf{dyn}}_{\mathsf{sc}}(T,\Delta) := (2C'+1)d^{\frac{2}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}} + 2(C'_1 + \frac{C'_2}{d} + \frac{4}{d^2})d^{\frac{4}{3}}\Delta^{\frac{2}{3}}T^{\frac{1}{3}}$. Finally, for known P we use (71)–(72) and apply the third result in Proposition 1 to bound $R^{\mathsf{path}}(T,P)$. For the general convex case, taking $\mathsf{B} = \left(\frac{r\sqrt{d}T}{P}\right)^{\frac{4}{5}}$ gives

$$R^{\mathsf{path}}(T,P) \leq (2C+1)r^{-\frac{1}{5}}d^{\frac{2}{5}}P^{\frac{1}{5}}T^{\frac{4}{5}} + 2(C_1 + \frac{C_2}{d} + \frac{4}{d^2})r^{-\frac{4}{5}}d^{\frac{8}{5}}P^{\frac{4}{5}}T^{\frac{1}{5}}\,.$$

For strongly-convex losses with minimizers inside Θ, taking $\mathsf{B} = \left(\frac{rdT}{P}\right)^{\frac{2}{3}}$ yields

$$R^{\mathsf{path}}(T,P) \leq (2C'+1)r^{-\frac{1}{3}}d^{\frac{2}{3}}P^{\frac{1}{3}}T^{\frac{2}{3}} + 2(C'_1 + \frac{C'_2}{d} + \frac{4}{d^2})r^{-\frac{2}{3}}d^{\frac{4}{3}}P^{\frac{2}{3}}T^{\frac{1}{3}}\,.$$

$\square$

C.6 Parameter-free upper bounds

Corollary 1 presents the optimal choice of parameter B for TEWA-SE when S, Δ and P are known. When the non-stationarity measures are unknown, the optimal B cannot be directly computed, and we therefore employ the Bandit-over-Bandit (BoB) framework from [27] to adaptively select B from a prespecified set $\mathcal{B} = \{2^i : i = 0, 1, \ldots, \lfloor \log_2 T \rfloor\}$. BoB has been used in [66] in a similar fashion to obtain parameter-free algorithms. Specifically, BoB divides the time horizon into $E = \lceil T/L \rceil$ epochs each with length L, denoted by $(I_e)_{e=1}^E$ (where the last epoch may be shorter than L). In the first epoch, it runs TEWA-SE with $\mathsf{B} = \mathsf{B}_1$ which is randomly selected from $\mathcal{B}$. For subsequent epochs, it uses the cumulative empirical loss on the current epoch $e-1$ to select $\mathsf{B}_e \in \mathcal{B}$ for the next epoch via EXP3 [59]. That is, BoB computes

$$p_{e,i} = (1-\gamma)\frac{s_{e,i}}{\sum_{i' \in [|\mathcal{B}|]} s_{e,i'}} + \frac{\gamma}{|\mathcal{B}|} \quad \forall i \in [|\mathcal{B}|], \quad \text{with} \quad \gamma = 1 \wedge \sqrt{\frac{|\mathsf{B}|\ln(|\mathcal{B}|)}{(\mathrm{e}-1)E}}\,, \quad (73)$$

where e denotes the base of the exponential function, and then samples $i_e = i$ with probability $p_{e,i}$ yielding $\mathsf{B}_e = 2^{i_e-1}$.[5] For $i \in [|\mathcal{B}|]$, initialized with $s_{0,i} = 1$, the quantity $s_{e,i}$ for $e \in \mathbb{N}^+$ is updated by computing

$$s_{e+1,i} = s_{e,i}\exp\left(\frac{\gamma}{|\mathcal{B}|}\widehat{r}_{e,i}\right), \quad (74)$$

where with $M_T = 1 + 2\sigma\sqrt{\log(T+1)}$, the importance-weighted reward $\widehat{r}_{e,i}$ takes the form

$$\widehat{r}_{e,i} = \begin{cases} \left(\frac{1}{2} + \frac{1}{2LM_T}\sum_{t \in I_e}(1-y_t)\right)/p_{e,i} & \text{if } i = i_e \\ 0 & \text{otherwise}\,. \end{cases} \quad (75)$$

Note that conditioned on the event $\Lambda_T = \left\{|\xi_t| \leq 2\sigma\sqrt{\log(T+1)},\ \forall t \in [T]\right\}$ defined above (10), the absolute total reward in each epoch is bounded by $Q := \max_{e \in [E]}\left|\sum_{t \in I_e}(1-y_t)\right| \leq LM_T$, which ensures the rescaled reward $\frac{1}{2} + \frac{1}{2LM_T}\sum_{t \in I_e}(1-y_t)$ in (75) remains bounded within $[0,1]$. The pseudo-code for TEWA-SE equipped with BoB is provided in Algorithm 4, with theoretical guarantees detailed in Corollary 3.

Corollary 3 (TEWA-SE with BoB). *Consider any horizon $T \in \mathbb{N}^+$ and assume that, for all $t \in [T]$, the loss f_t is convex, or strongly-convex with $\arg\min_{\boldsymbol{x} \in \mathbb{R}^d} f_t(\boldsymbol{x}) \in \Theta$ (referred to as the strongly-convex (SC) case).*

Then, for the general convex case, Algorithm 4 with epoch size $L = (dT)^{\frac{2}{3}}$ attains all the regret bounds from Corollary 1 plus an additional term of $d^{\frac{1}{3}}T^{\frac{5}{6}} + d^{\frac{4}{3}}T^{\frac{1}{3}}$. For the SC case, Algorithm 4 with epoch size $L = d\sqrt{T}$ satisfies all the regret bounds from Corollary 1 plus an additional term of $d^{\frac{1}{2}}T^{\frac{3}{4}} + d\sqrt{T}$. Both results omitted polylogarithmic factors.

---

[5]We adopt clipping (by γ) following [27, 59], though $\gamma = 0$ suffices as discussed in [77] and [9, Section 11.6].

Algorithm 4 TEWA equipped with Bandit-over-Bandit (BoB)

Input: $d, T, L, E = \lceil T/L \rceil, (I_e)_{e=1}^E, \mathcal{B} = \{2^i : i = 0, 1, \ldots, \lfloor \log_2 T \rfloor\}$, and $\gamma \in (0, 1)$ as defined in (73)
Initialize: $s_{0,i} = 1 \; \forall i \in [|\mathcal{B}|]$
 1: **for** $e = 1, 2, \ldots, E$ **do**
 2:     Compute $p_{e,i}$ according to (73) $\forall i \in [|\mathcal{B}|]$
 3:     Sample $i_e = i$ with probability $p_{e,i}$, and select $\mathsf{B}_e = 2^{i_e - 1} \in \mathcal{B}$
 4:     **for** $t \in I_e$ **do**
 5:         Run TEWA-SE with $\mathsf{B} = \mathsf{B}_e$ to select action $\boldsymbol{z}_t$ and observe losses $y_t = f_t(\boldsymbol{z}_t) + \xi_t$
 6:     **end for**
 7:     Update $s_{e+1,i}$ according to (74) $\forall i \in [|\mathcal{B}|]$
 8: **end for**

Proof of Corollary 3. For brevity, we suppress terms that are polylogarithmic in T using $\lesssim$ in this proof. For all $\mathsf{B}^\dagger \in \mathcal{B}$, we have

$$
\begin{aligned}
R^{\mathsf{dyn}}(T) &= \sum_{t=1}^T \mathbf{E}\left[f_t(\boldsymbol{z}_t) - \min_{\boldsymbol{z} \in \Theta} f_t(\boldsymbol{z})\right] \\
&= \sum_{t=1}^T \mathbf{E}\left[f_t(\boldsymbol{z}_t) - f_t\left(\boldsymbol{z}_t(\mathsf{B}^\dagger)\right)\right] + \sum_{t=1}^T \mathbf{E}\left[f_t\left(\boldsymbol{z}_t(\mathsf{B}^\dagger)\right) - \min_{\boldsymbol{z} \in \Theta} f_t(\boldsymbol{z})\right] \\
&= \underbrace{\sum_{e=1}^E \sum_{t \in I_e} \mathbf{E}\left[f_t\left(\boldsymbol{z}_t(\mathsf{B}_e)\right) - f_t\left(\boldsymbol{z}_t(\mathsf{B}^\dagger)\right)\right]}_{\text{term I}} + \underbrace{\sum_{e=1}^E \sum_{t \in I_e} \mathbf{E}\left[f_t\left(\boldsymbol{z}_t(\mathsf{B}^\dagger)\right) - \min_{\boldsymbol{z} \in \Theta} f_t(\boldsymbol{z})\right]}_{\text{term II}},
\end{aligned}
\tag{76}
$$

where $\boldsymbol{z}_t(\mathsf{B}_e)$ represents the actual action taken by TEWA-SE in round t of epoch e, and $\boldsymbol{z}_t(\mathsf{B}^\dagger)$ denotes the hypothetical action that TEWA-SE would have chosen had its B parameter been set to $\mathsf{B}^\dagger$. Term I in (76) can be bounded by applying the classical analysis of EXP3 from [59, Corollary 3.2], combined with (28), as follows

$$
\begin{aligned}
\forall \mathsf{B}^\dagger \in \mathcal{B}: \quad \text{term I} &\le 4\sqrt{e-1}\sqrt{E|\mathcal{B}|\log|\mathcal{B}|} \cdot \mathbf{E}[Q] \\
&\lesssim \sqrt{E} \cdot (\mathbf{P}(\Lambda_T)\mathbf{E}[Q \mid \Lambda_T] + \mathbf{P}(\Lambda_T^c)\mathbf{E}[Q \mid \Lambda_T^c]) \\
&\lesssim \sqrt{T/L} \cdot (LM_T + \tfrac{2}{T} \cdot L) \lesssim \sqrt{TL}.
\end{aligned}
\tag{77}
$$

To bound term II, we introduce shorthand $F^{\mathsf{ada}}(\mathsf{B}, T)$ to refer to the upper bound on $R^{\mathsf{ada}}(\mathsf{B}, T)$ in Theorem 1, and $F^{\mathsf{swi}}(T, S), F^{\mathsf{dyn}}(T, \Delta, S)$ and $F^{\mathsf{path}}(T, P)$ to refer to the upper bounds on $R^{\mathsf{swi}}(T, S), R^{\mathsf{dyn}}(T, \Delta, S)$ and $R^{\mathsf{path}}(T, P)$ in Corollary 1 for known S, Δ and P. We also use $S_e = 1 + \sum_{t \in I_e} \mathbf{1}(f_t \ne f_{t-1})$. By choosing $\mathsf{B}^\dagger = 2^{i^\dagger}$ in the analysis with $i^\dagger = \lfloor \log_2 \frac{T}{S} \rfloor \wedge \lfloor \log_2 L \rfloor$, term II can be bounded in terms of the number of switches S by

$$
\begin{aligned}
\text{term II} &\le \sum_{e=1}^E \left(\lceil \tfrac{L}{\mathsf{B}^\dagger} \rceil + S_e\right) R^{\mathsf{ada}}(\mathsf{B}^\dagger, L) \le \left(\tfrac{T}{\mathsf{B}^\dagger} + S + E\right) R^{\mathsf{ada}}(\mathsf{B}^\dagger, L) \\
&\le F^{\mathsf{swi}}(T, S) + \tfrac{T}{L} F^{\mathsf{ada}}(L, L).
\end{aligned}
\tag{78}
$$

Combining (77) and (78), we obtain

$$
\begin{aligned}
R^{\mathsf{swi}}(T, S) &\lesssim F^{\mathsf{swi}}(T, S) + \left[\tfrac{T}{L} F^{\mathsf{ada}}(L, L) + \sqrt{TL}\right] \\
&\lesssim F^{\mathsf{swi}}(T, S) + \begin{cases} d^{\frac{1}{3}} T^{\frac{5}{6}} + d^{\frac{4}{3}} T^{\frac{1}{3}} \\ d^{\frac{1}{2}} T^{\frac{3}{4}} + d\sqrt{T} \quad (\text{SC}), \end{cases}
\end{aligned}
\tag{79}
$$

where we used $L = (dT)^{\frac{2}{3}}$ for the general convex case, and $L = d\sqrt{T}$ for the strongly-convex case. Following similar steps, by choosing $\mathsf{B}^\dagger = 2^{i^\dagger}$ in the analysis with $i^\dagger = (\lfloor \log_2 \frac{T}{S} \rfloor \vee \lfloor \log_2(\mathsf{B}_\Delta) \rfloor) \wedge$

35

$\lfloor \log_2 L \rfloor$ where $\mathsf{B}_\Delta = \left(\frac{\sqrt{d} T}{\Delta} \right)^{\frac{4}{5}}$ for the general convex case or $\mathsf{B}_\Delta = \left(\frac{dT}{\Delta} \right)^{\frac{2}{3}}$ for the strongly-convex case, we obtain

$$R^{\mathsf{dyn}}(T, \Delta, S) \lesssim F^{\mathsf{dyn}}(T, \Delta, S) + \begin{cases} d^{\frac{1}{3}} T^{\frac{5}{6}} + d^{\frac{4}{3}} T^{\frac{1}{3}} \\ d^{\frac{1}{2}} T^{\frac{3}{4}} + d\sqrt{T} \quad \text{(SC)}. \end{cases} \tag{80}$$

The bound on $R^{\mathsf{path}}(T, P)$ can be established analogously. □

D   Proofs of lower bounds

We call $\pi = \{z_t\}_{t=1}^\infty$ a randomized procedure if $z_t = \Phi_t(\{z_k\}_{k=1}^{t-1}, \{y_k\}_{k=1}^{t-1})$ where Φ_t are Borel functions, and $z_1 \in \mathbb{R}^d$ is deterministic. We emphasize that, throughout this section, we assume the noise variables $\{\xi_t\}_{t=1}^T$ are independent with cumulative distribution function F satisfying the condition

$$\int \log \left(\mathrm{d}F(u) / \mathrm{d}F(u+v) \right) \mathrm{d}F(u) \leq I_0 v^2, \qquad |v| < v_0, \tag{81}$$

for some $0 < I_0 < \infty, 0 < v_0 \leq \infty$. This condition holds, for instance, if F has a sufficiently smooth density with finite Fisher information. In the special case where F is Gaussian, the inequality (81) holds with $v_0 = \infty$. Note that Gaussian noise also satisfies our sub-Gaussian noise assumption in Section 1, which is used in the proof of the upper bounds.

We first restate and prove Theorem 2, which establishes a lower bound on $R^{\mathsf{dyn}}(T, \Delta, S)$, and then present and prove Theorem 4, which establishes a lower bound on $R^{\mathsf{path}}(T, P)$.

Theorem 2. *Let $\Theta = \mathbb{B}^d$. For $\alpha > 0$ denote by $\mathcal{F}_\alpha$ the class of α-strongly convex and smooth functions. Let $\pi = \{z_t\}_{t=1}^T$ be any randomized algorithm (see Appendix D for a definition). Then there exists $T_0 > 0$ such that for all $T \geq T_0$ it holds that*

$$\sup_{f_1, \ldots, f_T \in \mathcal{F}_\alpha} R^{\mathsf{dyn}}(T, \Delta, S) \geq c_1 \cdot \left(d\sqrt{ST} \wedge d^{\frac{2}{3}} \Delta^{\frac{1}{3}} T^{\frac{2}{3}} \right), \tag{16}$$

where $c_1 > 0$ is a constant independent of d, T, S and Δ.

Proof of Theorem 2. Let $\eta_0 : \mathbb{R} \to \mathbb{R}$ be an infinitely many times differentiable function that satisfies

$$\eta_0(x) \begin{cases} = 1 & \text{if } |x| \leq 1/4, \\ \in (0, 1) & \text{if } 1/4 < |x| < 1, \\ = 0 & \text{if } |x| \geq 1. \end{cases}$$

Denote by $\Omega = \{-1, 1\}^d$ the set of binary sequences of length d, and let $\eta(x) = \int_{-\infty}^x \eta_0(u) \, \mathrm{d}u$. Consider the set of functions $f_{\boldsymbol{\omega}} : \mathbb{R}^d \to \mathbb{R}$ with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d) \in \{-1, 1\}^d$ such that:

$$f_{\boldsymbol{\omega}}(\boldsymbol{x}) = \alpha \|\boldsymbol{x}\|^2 + \iota h^2 \left(\sum_{i=1}^d \omega_i \eta \left(\frac{x_i}{h} \right) \right), \qquad \boldsymbol{x} = (x_1, \ldots, x_d), \tag{82}$$

where $h = \min \left(d^{-\frac{1}{2}}, \left(\frac{T}{S} \right)^{-\frac{1}{4}}, \left(\frac{dT}{\Delta} \right)^{-\frac{1}{6}} \right)$, and $\iota > 0$ is to be assigned later. Let $L' = \max_{x \in \mathbb{R}} |\eta''(x)|$. By [78, Lemma 10] we have that if $\iota \leq \min(1/2\eta(1), \alpha/L')$ then $f_{\boldsymbol{\omega}} \in \mathcal{F}_\alpha$. Moreover, if $\iota \leq \alpha/2$, the equation $\nabla f_{\boldsymbol{\omega}}(\boldsymbol{x}) = 0$ has the solution

$$\boldsymbol{x}^*(\boldsymbol{\omega}) = (x_1^*(\boldsymbol{\omega}), \ldots, x_d^*(\boldsymbol{\omega})), \quad \text{with} \quad x_i^*(\boldsymbol{\omega}) = -\frac{h\iota\omega_i}{2\alpha} \quad \text{for } 1 \leq i \leq d. \tag{83}$$

This is the unique minimizer of $f_{\boldsymbol{\omega}}$ and belongs to $\boldsymbol{x}^*(\boldsymbol{\omega}) \in \Theta = \mathbb{B}^d$ because

$$\|\boldsymbol{x}^*(\boldsymbol{\omega})\|^2 \leq \frac{h^2 \iota^2 d}{4\alpha^2} \leq \frac{1}{16}.$$

We consider the following adversarial protocol. At the beginning of the game, the adversary selects $N_c = \min(S, (T\Delta^2/d^2)^{\frac{1}{3}})$ points from Ω, sampled uniformly at random with replacement. Here

without loss of generality we assumed that $(T\Delta^2)^{\frac{1}{3}}$ is a positive integer. Denote these points by $\{\boldsymbol{\omega}_k\}_{k=1}^{N_c}$, and then for each $k = 1, 2, \ldots, N_c$, let

$$f_{(k-1)T/N_c+1} = \cdots = f_{kT/N_c} = f_{\boldsymbol{\omega}_k}.$$

For any $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega$ let $\rho(\boldsymbol{\omega}, \boldsymbol{\omega}') = \sum_{i=1}^{d} \mathbb{1}(\omega_i \neq \omega_i')$ be the Hamming distance between $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$, with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$ and $\boldsymbol{\omega}' = (\omega_1', \ldots, \omega_d')$. By construction, $N_c \leq S$ and

$$\sum_{k=2}^{N_c} \max_{\boldsymbol{x} \in \Theta} |f_{\boldsymbol{\omega}_{k-1}}(\boldsymbol{x}) - f_{\boldsymbol{\omega}_k}(\boldsymbol{x})| \leq 2\iota h^2 \eta(1) \sum_{k=2}^{N_c} \rho(\boldsymbol{\omega}_{k-1}, \boldsymbol{\omega}_k) \leq \Delta.$$

For any fixed $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_{N_c} \in \Omega$, and $1 \leq t \leq T$, denote $\Gamma = [\boldsymbol{\omega}_1 | \ldots | \boldsymbol{\omega}_{N_c}]$ as the matrix whose columns are the $\boldsymbol{\omega}_k$'s. Denote by $\mathbf{P}_{\Gamma,t}$ the probability measure corresponding to the joint distribution of $\{\boldsymbol{z}_k, y_k\}_{k=1}^{t}$ where $y_k = f_k(\boldsymbol{z}_k) + \xi_k$ with independent identically distributed ξ_k's such that (81) holds and $\boldsymbol{z}_k$'s are chosen by the algorithm π. We have

$$d\mathbf{P}_{\Gamma,t}(\boldsymbol{z}_{1:t}, y_{1:t}) = dF(y_1 - f_1(\boldsymbol{z}_1)) \prod_{\tau=2}^{t} dF(y_\tau - f_\tau(\Phi_\tau(\boldsymbol{z}_{1:\tau-1}, y_{1:\tau-1})))$$

$$= dF(y_1 - f_{\boldsymbol{\omega}_1}(\boldsymbol{z}_1)) \prod_{\tau=2}^{t} dF(y_\tau - f_{\boldsymbol{\omega}_{k_\tau}}(\Phi_\tau(\boldsymbol{z}_{1:\tau-1}, y_{1:\tau-1}))), \quad (84)$$

where $k_\tau = \lfloor (\tau-1)N_c/T \rfloor + 1$. (We omit explicit mention of the dependence of $\mathbf{P}_{\Gamma,t}$ and Φ_τ on $\boldsymbol{z}_2, \ldots, \boldsymbol{z}_{\tau-1}$, since $\boldsymbol{z}_\tau$ for $\tau \geq 2$ is a Borel function of $\boldsymbol{z}_1, y_1, \ldots, \boldsymbol{z}_{\tau-1}, y_{\tau-1}$.) Let $\mathbf{E}_{\Gamma,t}$ denote the expectation w.r.t. $\mathbf{P}_{\Gamma,t}$.

Note that by α-strong convexity of $f_{\boldsymbol{\omega}}$ and the fact that $\boldsymbol{x}^*(\boldsymbol{\omega}) \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^d} f_{\boldsymbol{\omega}}(\boldsymbol{x})$ from (83), we have

$$\sum_{t=1}^{T} \mathbf{E}_{\Gamma,t} \left[f_{\boldsymbol{\omega}_{k_t}}(\boldsymbol{z}_t) - \min_{\boldsymbol{x} \in \Theta} f_{\boldsymbol{\omega}_{k_t}}(\boldsymbol{x}) \right] \geq \frac{\alpha}{2} \sum_{t=1}^{T} \mathbf{E}_{\Gamma,t} \left[\|\boldsymbol{z}_t - \boldsymbol{x}^*(\boldsymbol{\omega}_{k_t})\|^2 \right]. \quad (85)$$

Define the nearest-neighbour estimator

$$\hat{\boldsymbol{\omega}}_t \in \arg\min_{\boldsymbol{\omega} \in \Omega} \|\boldsymbol{z}_t - \boldsymbol{x}^*(\boldsymbol{\omega})\|.$$

Using this combined with the triangle inequality, we have $\|\boldsymbol{x}^*(\hat{\boldsymbol{\omega}}_t) - \boldsymbol{x}^*(\boldsymbol{\omega}_{k_t})\| \leq \|\boldsymbol{z}_t - \boldsymbol{x}^*(\hat{\boldsymbol{\omega}}_t)\| + \|\boldsymbol{z}_t - \boldsymbol{x}^*(\boldsymbol{\omega}_{k_t})\| \leq 2\|\boldsymbol{z}_t - \boldsymbol{x}^*(\boldsymbol{\omega}_{k_t})\|$. Together with (83) this implies that

$$\mathbf{E}_{\Gamma,t} \left[\|\boldsymbol{z}_t - \boldsymbol{x}^*(\boldsymbol{\omega}_{k_t})\|^2 \right] \geq \frac{1}{4} \mathbf{E}_{\Gamma,t} \left[\|\boldsymbol{x}^*(\hat{\boldsymbol{\omega}}_t) - \boldsymbol{x}^*(\boldsymbol{\omega}_{k_t})\|^2 \right] = \frac{h^2 \iota^2}{4\alpha^2} \mathbf{E}_{\Gamma,t} \left[\rho(\hat{\boldsymbol{\omega}}_t, \boldsymbol{\omega}_{k_t}) \right].$$

Summing over $1, \ldots, T$, then taking the maximum over $\Gamma = [\boldsymbol{\omega}_1 | \ldots | \boldsymbol{\omega}_{N_c}]$ and then the minimum over all estimators $\hat{\boldsymbol{\omega}}_1, \ldots, \hat{\boldsymbol{\omega}}_T$ with values in Ω, we get

$$\max_{\Gamma \in \Omega^{N_c}} \sum_{t=1}^{T} \mathbf{E}_{\Gamma,t} \left[\|\boldsymbol{z}_t - \boldsymbol{x}^*(\boldsymbol{\omega}_{k_t})\|^2 \right] \geq \frac{h^2 \iota^2}{4\alpha^2} \underbrace{\min_{\hat{\boldsymbol{\omega}}_1, \ldots, \hat{\boldsymbol{\omega}}_T \in \Omega} \max_{\Gamma \in \Omega^{N_c}} \sum_{t=1}^{T} \sum_{i=1}^{d} \mathbf{E}_{\Gamma,t} \left[\mathbb{1}(\hat{\omega}_{t,i} \neq \omega_{k_t,i}) \right]}_{\text{term I}}.$$

$$(86)$$

For term I, lower bounding the maximum with the average we can write

$$\text{term I} \geq 2^{-dN_c} \min_{\hat{\boldsymbol{\omega}}_1, \ldots, \hat{\boldsymbol{\omega}}_T \in \Omega} \sum_{t=1}^{T} \sum_{\Gamma \in \Omega^{N_c}} \sum_{i=1}^{d} \mathbf{E}_{\Gamma,t} \left[\mathbb{1}(\hat{\omega}_{t,i} \neq \omega_{k_t,i}) \right]$$

$$\geq 2^{-dN_c} \sum_{t=1}^{T} \sum_{\Gamma \in \Omega^{N_c}} \sum_{i=1}^{d} \min_{\hat{\omega}_{t,i} \in \{-1,1\}} \mathbf{E}_{\Gamma,t} \left[\mathbb{1}(\hat{\omega}_{t,i} \neq \omega_{k_t,i}) \right].$$

37

Next, for each $i = 1, \ldots, d$, define $\Gamma_i^{k_t} = \{[\boldsymbol{\omega}_1 | \ldots | \boldsymbol{\omega}_{N_c}] : \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_{N_c} \in \Omega, \omega_{k_t,i} = 1\}$. Given any $\Gamma \in \Gamma_i^{k_t}$, let $\bar{\Gamma} = [\bar{\boldsymbol{\omega}}_1 | \ldots | \bar{\boldsymbol{\omega}}_{N_c}]$ such that $\bar{\omega}_{k,j} = \omega_{k,j}$ for any $k \neq k_t$, and let $\bar{\omega}_{k_t,i} = -1$ and $\bar{\omega}_{k_t,j} = \omega_{k_t,j}$ for $j \neq i$. Hence,

$$
\begin{aligned}
\text{term I} &\geq 2^{-dN_c} \sum_{t=1}^{T} \sum_{\Gamma \in \Omega^{N_c}} \sum_{i=1}^{d} \min_{\hat{\omega}_{t,i} \in \{-1,1\}} \left(\mathbf{E}_{\Gamma,t} \left[\mathbb{1} \left(\hat{\omega}_{t,i} \neq 1 \right) \right] + \mathbf{E}_{\bar{\Gamma},t} \left[\mathbb{1} \left(\hat{\omega}_{t,i} \neq -1 \right) \right] \right) \\
&\geq \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{d} \min_{\Gamma \in \Gamma_i^{k_t}} \min_{\hat{\omega}_{t,i} \in \{-1,1\}} \left(\mathbf{E}_{\Gamma,t} \left[\mathbb{1} \left(\hat{\omega}_{t,i} \neq 1 \right) \right] + \mathbf{E}_{\bar{\Gamma},t} \left[\mathbb{1} \left(\hat{\omega}_{t,i} \neq -1 \right) \right] \right).
\end{aligned}
$$

Thus, we can write

$$
\begin{aligned}
\text{KL}\left(\mathbf{P}_{\Gamma,t} || \mathbf{P}_{\bar{\Gamma},t} \right) &= \int \log \left(\frac{d\mathbf{P}_{\Gamma,t}}{d\mathbf{P}_{\bar{\Gamma},t}} \right) d\mathbf{P}_{\Gamma,t} \\
&= \int \left[\log \left(\frac{dF(y_1 - f_{\boldsymbol{\omega}_1}(\boldsymbol{z}_1))}{dF(y_1 - f_{\bar{\boldsymbol{\omega}}_1}(\boldsymbol{z}_1))} \right) + \right. \\
&\qquad \left. + \sum_{\tau=2}^{t} \log \left(\frac{dF(y_\tau - f_{\boldsymbol{\omega}_{k_\tau}}(\Phi_\tau(\boldsymbol{z}_{1:\tau-1}, y_{1:\tau-1})))}{dF(y_\tau - f_{\bar{\boldsymbol{\omega}}_{k_\tau}}(\Phi_\tau(\boldsymbol{z}_{1:\tau-1}, y_{1:\tau-1})))} \right) \right] \\
&\qquad dF\left(y_1 - f_{\boldsymbol{\omega}_1}(\boldsymbol{z}_1) \right) \prod_{\tau=2}^{t} dF\left(y_\tau - f_{\boldsymbol{\omega}_{k_\tau}}(\Phi_\tau(\boldsymbol{z}_{1:\tau-1}, y_{1:\tau-1})) \right) \\
&\leq I_0 \sum_{\tau=1}^{t} \max_{\boldsymbol{x} \in \Theta} |f_{\boldsymbol{\omega}_{k_\tau}}(\boldsymbol{x}) - f_{\bar{\boldsymbol{\omega}}_{k_\tau}}(\boldsymbol{x})|^2 \leq 4TN_c^{-1} I_0 \iota^2 h^4 \eta^2(1).
\end{aligned}
$$

Since $h \leq \min\left(\left(\frac{S}{T} \right)^{\frac{1}{4}}, \left(\frac{\Delta}{dT} \right)^{\frac{1}{6}} \right)$, and by choosing $\iota \leq \sqrt{\log(2)/(4I_0 \eta^2(1))}$, we have $\text{KL}(\mathbf{P}_{\Gamma,t} || \mathbf{P}_{\bar{\Gamma},t}) \leq \log(2)$. Hence, Theorem 2.12 of [79] gives

$$
\text{term I} \geq \frac{Td}{4} \exp(-\log(2)) = \frac{Td}{8}.
$$

Substituting this into (86) and our overall bound (85) yields

$$
\max_{\Gamma \in \Omega^{N_c}} \sum_{t=1}^{T} \mathbf{E}_{\Gamma,t} \left[f_{\boldsymbol{\omega}_{k_t}}(\boldsymbol{z}_t) - \min_{\boldsymbol{x} \in \Theta} f_{\boldsymbol{\omega}_{k_t}}(\boldsymbol{x}) \right] \geq \frac{\alpha}{2} \cdot \frac{h^2 \iota^2}{4\alpha^2} \cdot \frac{Td}{8} = \frac{h^2 \iota^2 Td}{64\alpha}.
$$

Finally, substituting the definition of h and noting that ι is independent of d, T, S and Δ completes the proof. $\qquad \square$

Theorem 4. *Let $\Theta = \mathbb{B}^d$. For $\alpha > 0$ denote by $\mathcal{F}_\alpha$ the class of α-strongly convex and smooth functions. Let $\pi = \{\boldsymbol{z}_t\}_{t=1}^{T}$ be any randomized algorithm. Then there exists $T_0 > 0$ such that for all $T \geq T_0$ it holds that*

$$
\sup_{f_1, \ldots, f_T \in \mathcal{F}_\alpha} R^{\text{path}}(T, P) \geq c_2 \cdot (d^2 P)^{\frac{2}{5}} T^{\frac{3}{5}},
$$

where $c_2 > 0$ is a constant indepedent of d, T and P.

Proof of Theorem 4. The proof uses the same notation and follows the same steps as in the proof of Theorem 2, but with different choices for the parameters h and N_c. Define the set of functions $f_{\boldsymbol{\omega}} : \mathbb{R}^d \to \mathbb{R}$ with $\boldsymbol{\omega} \in \{-1, 1\}^d$ as they are defined in (82), and choose $h = \min(d^{-\frac{1}{2}}, \frac{P}{N_c \sqrt{d}})$ and $N_c = \lfloor P^{\frac{4}{5}} T^{\frac{1}{5}} d^{-\frac{2}{5}} \rfloor$. Then we have that

$$
\sum_{k=2}^{N_c} \|\boldsymbol{x}^*(\boldsymbol{\omega}_{k-1}) - \boldsymbol{x}^*(\boldsymbol{\omega}_k)\| = \frac{h\iota}{\alpha} \sum_{k=2}^{N_c} \sqrt{\rho(\boldsymbol{\omega}_{k-1}, \boldsymbol{\omega}_k)} \leq \frac{h\iota}{\alpha} \sqrt{d} N_c \leq P, \tag{87}
$$

for any $\iota \leq \frac{\alpha}{2}$. Following similar steps as in the proof of Theorem 2 for large enough T (when $h = \frac{P}{N_c\sqrt{d}}$) we get

$$\max_{\Gamma \in \Omega^{N_c}} \sum_{t=1}^{T} \mathbf{E}_{\Gamma,t} \left[f_{\boldsymbol{\omega}_{k_t}}(\boldsymbol{z}_t) - \min_{\boldsymbol{x} \in \Theta} f_{\boldsymbol{\omega}_{k_t}}(\boldsymbol{x}) \right] \geq \frac{h^2 \iota^2 T d}{64\alpha} \geq c_2 (d^2 P)^{\frac{2}{5}} T^{\frac{3}{5}} \,,$$

where $c_2 > 0$ is independent of d, T and P. $\qquad\square$

E Proofs for clipped Exploration by Optimization

We restate and prove Theorem 3 which establishes an adaptive regret guarantee for cExO. In this section, we use $\langle \boldsymbol{p}, f_t \rangle = \mathbf{E}_{\boldsymbol{z} \sim \boldsymbol{p}}[f_t(\boldsymbol{z})]$ where $\boldsymbol{p}$ belongs to a probability simplex.

Theorem 3. *For* $T \in \mathbb{N}^+$ *and* $\mathsf{B} \in [T]$, *Algorithm 3 calibrated with* $\varepsilon = \frac{1}{T}$, $\gamma = \frac{1}{T|\mathcal{C}|}$, $\eta = \sqrt{\log(\gamma^{-1})/(d^4 \log(dT)\mathsf{B})}$ *and* $\log|\mathcal{C}| = \mathcal{O}(d\log(dT^2))$ *satisfies*

$$R^{ada}(\mathsf{B}, T) \lesssim d^{\frac{5}{2}}\sqrt{\mathsf{B}} \,. \tag{19}$$

Proof of Theorem 3. Consider an arbitrary interval $[a, b]$ of length $b - a + 1 \leq \mathsf{B}$, and notice that for any $\boldsymbol{q}^\star \in \tilde{\Delta}$,

$$\max_{\boldsymbol{u} \in \Theta} \sum_{t=a}^{b} \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})] = \underbrace{\sum_{t=a}^{b} \langle \boldsymbol{p}_t - \boldsymbol{q}^\star, f_t \rangle}_{\text{term I}} + \underbrace{\sum_{t=a}^{b} \mathbf{E}_{\boldsymbol{z} \sim \boldsymbol{q}^\star}[f_t(\boldsymbol{z})] - \min_{\boldsymbol{u} \in \Theta} \sum_{t=a}^{b} f_t(\boldsymbol{u})}_{\text{term II}} \,. \tag{88}$$

In what follows, we choose a suitable $\boldsymbol{q}^\star$ and bound term I and term II separately.

Recall that the covering set $\mathcal{C}$ is assumed in Section 3 to have a discretization error of ε, implying that there exists a $\boldsymbol{u}_\mathcal{C} \in \mathcal{C}$ such that $\sum_{t=a}^{b} f_t(\boldsymbol{u}_\mathcal{C}) - \min_{\boldsymbol{u} \in \Theta} \sum_{t=a}^{b} f_t(\boldsymbol{u}) \leq \varepsilon\mathsf{B}$. Define $\boldsymbol{q}^\star \in \tilde{\Delta}$ to be the distribution with probability mass given by

$$q^\star(\boldsymbol{z}) = \begin{cases} 1 - \gamma(|\mathcal{C}| - 1) & \text{if } \boldsymbol{z} = \boldsymbol{u}_\mathcal{C} \\ \gamma & \text{otherwise} \,. \end{cases} \tag{89}$$

This construction ensures that

$$\text{term II} \leq (\varepsilon + 2\gamma|\mathcal{C}|)\mathsf{B} \,. \tag{90}$$

To bound term I, we first apply Lemma 8 to the sequence of Online Mirror Descent (OMD) updates $\boldsymbol{q}_t \in \tilde{\Delta}$ and the sequence of loss estimates $\widehat{\boldsymbol{s}}_t$ to obtain

$$\sum_{t=a}^{b} \langle \boldsymbol{q}_t - \boldsymbol{q}^\star, \widehat{\boldsymbol{s}}_t \rangle \leq \frac{1}{\eta} \left(\text{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_a) + \sum_{t=a}^{b} S_{\boldsymbol{q}_t}(\eta\widehat{\boldsymbol{s}}_t) \right) \,, \tag{91}$$

where by the definition of $q^\star(\cdot)$ in (89), we have

$$\begin{aligned}
\text{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_a) &= \sum_{\boldsymbol{z} \in \mathcal{C}} q^\star(\boldsymbol{z}) \log\left(\frac{q^\star(\boldsymbol{z})}{q_a(\boldsymbol{z})} \right) \\
&= (1 - \gamma(|\mathcal{C}| - 1)) \log\left(\frac{1 - \gamma(|\mathcal{C}| - 1)}{q_a(\boldsymbol{u}_\mathcal{C})} \right) + \sum_{\boldsymbol{z} \in \mathcal{C} \setminus \{\boldsymbol{u}_\mathcal{C}\}} \gamma \log\left(\frac{\gamma}{q_a(\boldsymbol{z})} \right) \\
&\leq \log(\gamma^{-1}) \,.
\end{aligned} \tag{92}$$

Then applying (91) and (92), we have

$$\text{term I} = \sum_{t=a}^{b} [\langle \boldsymbol{q}_t - \boldsymbol{q}^\star, \widehat{\boldsymbol{s}}_t \rangle + \langle \boldsymbol{p}_t - \boldsymbol{q}^\star, f_t \rangle + \langle \boldsymbol{q}^\star - \boldsymbol{q}_t, \widehat{\boldsymbol{s}}_t \rangle]$$

$$\leq \frac{\log(\gamma^{-1})}{\eta} + \sum_{t=a}^{b} \left[\langle \boldsymbol{p}_t - \boldsymbol{q}^\star, f_t \rangle + \langle \boldsymbol{q}^\star - \boldsymbol{q}_t, \widehat{\boldsymbol{s}}_t \rangle + \frac{1}{\eta} S_{\boldsymbol{q}_t}(\eta \widehat{\boldsymbol{s}}_t) \right]$$

$$\overset{(i)}{\leq} \frac{\log(\gamma^{-1})}{\eta} + \mathsf{B} \left(\inf_{\substack{\boldsymbol{p} \in \Delta(\mathcal{C}), \\ E \in \mathcal{E}}} \Lambda_\eta(\boldsymbol{q}_t, \boldsymbol{p}, E) + \eta d \right)$$

$$\overset{(ii)}{\leq} \frac{\log(\gamma^{-1})}{\eta} + \mathsf{B} \left(\eta \kappa d^4 \log(dT) + \eta d \right) \,, \tag{93}$$

where (i) follows from the update rule (18) and the precision level assumed for solving the minimization problem (18) (see line 3 of Algorithm 3), and (ii) uses [1, Theorems 8.19 and 8.21] which establish that there exists a universal constant κ such that

$$\sup_{\boldsymbol{q} \in \tilde{\Delta}} \inf_{\substack{\boldsymbol{p} \in \Delta(\mathcal{C}), \\ E \in \mathcal{E}}} \frac{1}{\eta} \Lambda_\eta(\boldsymbol{q}, \boldsymbol{p}, E) \leq \kappa d^4 \log(dT) \,.$$

Finally, combining (90) and (93) we obtain

$$\max_{\boldsymbol{u} \in \Theta} \sum_{t=a}^{b} \mathbf{E}[f_t(\boldsymbol{z}_t) - f_t(\boldsymbol{u})] = \sum_{t=a}^{b} \langle \boldsymbol{p}_t - \boldsymbol{q}^\star, f_t \rangle + \sum_{t=a}^{b} \mathbf{E}_{\boldsymbol{z} \sim \boldsymbol{q}^\star}[f_t(\boldsymbol{z})] - \min_{\boldsymbol{u} \in \Theta} \sum_{t=a}^{b} f_t(\boldsymbol{u})$$

$$\leq (\varepsilon + 2\gamma|\mathcal{C}|)\mathsf{B} + \frac{\log(\gamma^{-1})}{\eta} + \mathsf{B} \left(\eta \kappa d^4 \log(dT) + \eta d \right)$$

$$\lesssim \frac{\mathsf{B}}{T} + \sqrt{\mathsf{B} d^4 \log(dT) \log(T|\mathcal{C}|)} \lesssim d^{\frac{5}{2}} \sqrt{\mathsf{B}} \,,$$

where (i) applies $\varepsilon = \frac{1}{T}$, $\gamma = \frac{1}{T|\mathcal{C}|}$ and $\eta = \sqrt{\log(\gamma^{-1})/(d^4 \log(dT)\mathsf{B})}$, and (ii) is by selecting the covering set $\mathcal{C}$ such that $\log|\mathcal{C}| \leq d \log(1 + 16dT^2)$ (existence given by [1, Definition 8.12 and Exercise 8.13]). $\qquad\square$

The proof of Theorem 3 above relied on Lemma 8, which we present and prove below.

Lemma 8. *Consider Online Mirror Descent (OMD) with KL divergence regularization and fixed learning rate $\eta > 0$ applied to a sequence of loss estimates $\boldsymbol{s}_t \in \mathbb{R}^n$ for $t \in \mathbb{N}^+$. When run over a convex and complete domain $\tilde{\Delta} \subseteq \Delta^{n-1}$, the algorithm produces a sequence of updates $\boldsymbol{q}_t \in \tilde{\Delta}$ for $t \in \mathbb{N}^+$. For any comparator in $\boldsymbol{q}^\star \in \tilde{\Delta}$ and time interval $\{t \in \mathbb{N}^+ : a \leq t \leq b\}$, it holds that*

$$\sum_{t=a}^{b} \langle \boldsymbol{q}_t - \boldsymbol{q}^\star, \boldsymbol{s}_t \rangle \leq \frac{1}{\eta} \left(KL(\boldsymbol{q}^\star || \boldsymbol{q}_a) + \sum_{t=a}^{b} S_{\boldsymbol{q}_t}(\eta \boldsymbol{s}_t) \right) \,,$$

where $S_{\boldsymbol{q}}(\eta \boldsymbol{s}) = \max_{\boldsymbol{q}' \in \Delta(\mathcal{C})} \langle \boldsymbol{q} - \boldsymbol{q}', \eta \boldsymbol{s} \rangle - KL(\boldsymbol{q}' || \boldsymbol{q})$.

Proof of Lemma 8. The proof is standard and included for completeness. Let F denote the neg-entropy $F(\boldsymbol{q}) = \sum_{i=1}^{n} q_i \log(q_i)$ for $\boldsymbol{q} \in \Delta^{n-1}$, and note that

$$KL(\boldsymbol{p} || \boldsymbol{q}) = F(\boldsymbol{p}) - \langle \boldsymbol{p} - \boldsymbol{q}, \nabla F(\boldsymbol{q}) \rangle - F(\boldsymbol{q}) \quad \forall \, \boldsymbol{p}, \boldsymbol{q} \in \Delta^{n-1} \,. \tag{94}$$

Consider the update rule of the OMD defined in the lemma:

$$\boldsymbol{q}_{t+1} = \underset{\boldsymbol{q} \in \tilde{\Delta}}{\arg\min} \, \langle \boldsymbol{q}, \eta \boldsymbol{s}_t \rangle + KL(\boldsymbol{q} || \boldsymbol{q}_t) = \underset{\boldsymbol{q} \in \tilde{\Delta}}{\arg\min} \, \langle \boldsymbol{q}, \eta \boldsymbol{s}_t \rangle + F(\boldsymbol{q}) - \boldsymbol{q} \nabla F(\boldsymbol{q}_t) \,,$$

which implies by the first order optimality condition [9, Proposition 26.14] that, for any $\boldsymbol{q}^\star \in \tilde{\Delta}$ and time t,

$$\langle \boldsymbol{q}^\star - \boldsymbol{q}_{t+1}, \eta \boldsymbol{s}_t + \nabla F(\boldsymbol{q}_{t+1}) - \nabla F(\boldsymbol{q}_t) \rangle \geq 0 \,. \tag{95}$$

Rearranging (95) and applying (94) we obtain

$$\langle \boldsymbol{q}_{t+1} - \boldsymbol{q}^\star, \boldsymbol{s}_t \rangle \le \frac{1}{\eta} \langle \boldsymbol{q}^\star - \boldsymbol{q}_{t+1}, \nabla F(\boldsymbol{q}_{t+1}) - \nabla F(\boldsymbol{q}_t) \rangle$$

$$= \frac{1}{\eta} \left(\mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_t) - \mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_{t+1}) - \mathrm{KL}(\boldsymbol{q}_{t+1} \| \boldsymbol{q}_t) \right)$$

$$\le -\langle \boldsymbol{q}_t - \boldsymbol{q}_{t+1}, \boldsymbol{s}_t \rangle + \frac{1}{\eta} S_{\boldsymbol{q}_t}(\eta \boldsymbol{s}_t) + \frac{1}{\eta} \left(\mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_t) - \mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_{t+1}) \right). \quad (96)$$

Rearranging (96) and summing over $t \in [a, b]$ yields

$$\sum_{t=a}^{b} \langle \boldsymbol{q}_t - \boldsymbol{q}^\star, \boldsymbol{s}_t \rangle = \sum_{t=a}^{b} \left(\langle \boldsymbol{q}_{t+1} - \boldsymbol{q}^\star, \boldsymbol{s}_t \rangle + \langle \boldsymbol{q}_t - \boldsymbol{q}_{t+1}, \boldsymbol{s}_t \rangle \right)$$

$$\le \frac{1}{\eta} \sum_{t=a}^{b} \left(\mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_t) - \mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_{t+1}) + S_{\boldsymbol{q}_t}(\eta \boldsymbol{s}_t) \right)$$

$$= \frac{1}{\eta} \left(\mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_a) - \mathrm{KL}(\boldsymbol{q}^\star \| \boldsymbol{q}_{b+1}) + \sum_{t=a}^{b} S_{\boldsymbol{q}_t}(\eta \boldsymbol{s}_t) \right),$$

which combined with non-negativity of the KL divergence completes the proof. $\square$

Finally, we apply Theorem 3 to prove the bounds on $R^{\mathsf{swi}}(T, S)$, $R^{\mathsf{dyn}}(T, \Delta, S)$ and $R^{\mathsf{path}}(T, P)$ in Corollary 2, as well as the parameter-free guarantees in Corollary 4.

Corollary 2. *For any horizon $T \in \mathbb{N}^+$, Algorithm 3 calibrated as in Theorem 3 and tuned with interval size B (which determines η) satisfies the following regret guarantees:*

> ***Switching:*** $\quad \mathsf{B} = \frac{T}{S} \implies R^{\mathsf{swi}}(T, S) \lesssim d^{\frac{5}{2}} \sqrt{ST}$,
>
> ***Dynamic:*** $\quad \mathsf{B} = \frac{T}{S} \vee (d^{\frac{5}{2}} T / \Delta)^{\frac{2}{3}} \implies R^{\mathsf{dyn}}(T, \Delta, S) \lesssim R^{\mathsf{swi}}(T, S) \wedge d^{\frac{5}{3}} \Delta^{\frac{1}{3}} T^{\frac{2}{3}}$,
>
> ***Path-length:*** $\quad \mathsf{B} = (r d^{\frac{5}{2}} T / P)^{\frac{2}{3}} \implies R^{\mathsf{path}}(T, P) \lesssim r^{-\frac{1}{3}} d^{\frac{5}{3}} P^{\frac{1}{3}} T^{\frac{2}{3}}$.

Proof of Corollary 2. We prove these results by applying the adaptive regret guarantee from Theorem 3 and the conversions results from Proposition 1, similarly to the proof of Corollary 1. $\square$

Corollary 4 (cExO with BoB)**.** *Let $T \in \mathbb{N}^+$. By partitioning the time horizon $[T]$ into epochs of length $L = d^{\frac{5}{2}} \sqrt{T}$, and employing Bandit-over-Bandit to select cExO's parameter B for each epoch from the set $\mathcal{B} = \{2^i : i = 0, 1, \dots, \lfloor \log_2 T \rfloor\}$, this algorithm achieves all regret bounds in Corollary 2 with an additional term of $d^{\frac{5}{4}} T^{\frac{3}{4}}$ (up to polylogarithmic factors).*

Proof of Corollary 4. The proof is similar to that of Corollary 3 and is therefore omitted. $\square$