A Reduction from Delayed to Immediate Feedback for Online Convex Optimization with Improved Guarantees

Alexander Ryabchenko* **Idan Attias †** **Daniel M. Roy‡**

Abstract

This paper develops a reduction-based framework for online learning with delayed feedback that recovers and improves upon existing results for first-order and bandit convex optimization. We introduce a continuous-time model that enables an observation-centric rather than prediction-centric analysis. Under this model, regret decomposes into a delay-independent learning term and a delay-induced drift term, yielding a delay-adaptive reduction that converts any online linear optimization algorithm into one that handles round-dependent delays. For bandit convex optimization, we significantly improve existing regret bounds, with delay-dependent terms matching optimal first-order rates. For first-order feedback, we recover optimal regret bounds via a simpler, unified analysis.

Quantitatively, for bandit convex optimization we obtain $O(\sqrt{d_{\text{tot}}} + T^{\frac{3}{4}}\sqrt{k})$ regret, improving the delay-dependent term from $O(\min\{\sqrt{T d_{\max}}, (T d_{\text{tot}})^{\frac{1}{3}}\})$ in prior work to $O(\sqrt{d_{\text{tot}}})$. Here, k, T, $d_{\max}$, and d_{tot} denote the dimension, time horizon, maximum and total delay, respectively. Under strong convexity, we achieve $O(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\} + (T^2 \ln T)^{\frac{1}{3}} k^{\frac{2}{3}})$, improving the delay-dependent term from $O(d_{\max} \ln T)$ to $O(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\})$, where $\sigma_{\max}$ denotes the maximum number of outstanding observations and may be considerably smaller than $d_{\max}$.

1 Introduction

Online convex optimization (OCO) is a fundamental framework for sequential decision-making under uncertainty [Hazan, 2016, Orabona, 2019]. In the classic OCO setting, a player repeatedly selects an action from a convex domain, incurs a loss specified by a convex loss function, and then receives feedback about that loss. In the full-information setting, the entire loss function is revealed. In the first-order setting, the player observes the gradient of the loss at the chosen action. In the zeroth-order (bandit) setting, the player observes only the scalar loss value.

In practice, observations are often *delayed*: feedback from round t may arrive only at round $t + d_t$, where d_t is the round-dependent delay. This forces the player to make decisions without up-to-date information. For example, consider a recommendation system that must select content instantaneously for each arriving user. Feedback needed to update the system's parameters (such as clicks or purchases) arrives only after user sessions conclude. When many users are served concurrently, new arrivals must be handled while feedback from earlier users is pending, resulting in delays. Delayed feedback has been studied extensively in many variants of sequential decision-making, including online learning, bandits, and reinforcement learning (e.g., Mesterharm, 2005, Cesa-Bianchi et al., 2016, Zimmert and Seldin, 2020, van der Hoeven et al., 2023).

*University of Toronto and Vector Institute; alex.rbch.research@gmail.com.

†Institute for Data, Econometrics, Algorithms, and Learning (IDEAL), hosted by the University of Illinois at Chicago and Toyota Technological Institute at Chicago; idanattias88@gmail.com.

‡University of Toronto and Vector Institute; daniel.roy@utoronto.ca.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

A natural question is whether delayed algorithms can be obtained as reductions from non-delayed ones. Prior work has made significant progress on this question [Weinberger and Ordentlich, 2002, Joulani et al., 2013, 2016, Flaspohler et al., 2021]. However, existing approaches either require advance knowledge of the delay process, incur memory overhead scaling with the amount of outstanding feedback, or are limited to full-information feedback. Developing a unified, delay-adaptive reduction—one that requires no advance knowledge of delays, adapts to function regularity (e.g., strong convexity), and accommodates different feedback models (first-order or bandit)—is the main goal of this work.

In this paper, we introduce a continuous-time model for delayed feedback that subsumes the standard discrete-time model and enables a clean reduction from delayed feedback to immediate feedback for both online convex optimization with first-order feedback (OCO) and bandit convex optimization (BCO). For BCO, this reduction yields improved regret rates, while for OCO it recovers optimal bounds via a simple, unified analysis. Our contributions are summarized as follows:

1. **Continuous-time model (Section 3).** We introduce a continuous-time model for delayed feedback. Instead of the usual round-based view in which each round reveals an unordered batch of observations, we place predictions and observations as events on a shared timeline, and show that the effective delay structure is induced by their interleaving.

   Theorem 3.4 characterizes this structure, identifying new invariants and equivalences among delay-related quantities that are widely used across the delayed-learning literature. Ordering the interaction by observation times provides a natural framework for analyzing algorithms that run single-instance updates upon each feedback arrival (e.g., Joulani et al. [2016]), yielding a regret decomposition into a non-delayed term and a prediction-drift term (Lemma 3.7). Our analysis provides new tools for controlling this drift, which we exploit to improve regret bounds through a unified analysis, without requiring knowledge of problem parameters such as the maximum delay $d_{\max}$, total delay d_{tot}, maximum number of outstanding observations $\sigma_{\max}$, or the horizon T.

2. **Reduction to non-delayed learning (Sections 4–5).** Building on this decomposition, we reduce delayed OCO to online linear optimization with penalties for changing predictions, a delay-free problem related to online learning with switching costs [Cesa-Bianchi et al., 2013]. We provide wrappers that transform any algorithm for this setting into a delayed OCO or BCO algorithm whose regret is controlled by the base algorithm's performance (Theorems 5.2 and 5.6). The BCO wrapper incorporates single-point gradient estimation [Flaxman et al., 2005].

3. **First-order feedback (Section 5.1).** Wrapping Proximal Follow-The-Regularized-Leader (P-FTRL) and Online Mirror Descent (OMD) update rules recovers optimal bounds for delayed OCO: $O(\sqrt{d_{\text{tot}}} + \sqrt{T})$ for convex losses (Corollary 5.3) and $O(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\} + \ln T)$ for strongly convex losses (Corollary 5.4), matching Quanrud and Khashabi [2015] and Qiu et al. [2025] respectively via an adaptive approach requiring no knowledge of delays or horizon.

4. **Bandit feedback (Section 5.2).** For delayed BCO, wrapping P-FTRL achieves $O(\sqrt{d_{\text{tot}}} + T^{3/4}\sqrt{k})$ expected regret (Corollary 5.7), where k is the dimension of the domain, improving the delay dependence over $(Td_{\text{tot}})^{1/3}$ of Bistritz et al. [2022] and $\sqrt{Td_{\max}}$ of Wan et al. [2024a]. For strongly convex losses, we obtain $O(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\} + (T^2 \ln T)^{1/3} k^{2/3})$ regret (Corollary 5.8), improving over $O(d_{\max} \ln T + (T^2 \ln T)^{1/3} k^{2/3})$ of Wan et al. [2024a]. Note that $\sigma_{\max}$ is always at most $d_{\max}$ and can be much smaller in practice.[4] See Table 1 for a summary.

5. **Two-point bandit feedback (Appendix F).** To further illustrate the applicability of our reduction, we provide the first regret bounds for delayed BCO under the two-point feedback model of Agarwal et al. [2010]. Using the two-point gradient estimator [Shamir, 2017], we obtain regret $O(\sqrt{d_{\text{tot}}} + \sqrt{Tk})$ for convex losses and $O(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\} + k \ln T)$ under strong convexity. Our bounds simultaneously recover the known lower-bound dependences for delay-free two-point BCO and delayed OCO in both the convex and strongly convex regimes [Duchi et al., 2015, Agarwal et al., 2010, Weinberger and Ordentlich, 2002], establishing rate optimality for delayed two-point BCO. See Table 2 for a summary.

6. **Skipping scheme (Section G).** We incorporate the adaptive skipping technique of Zimmert and Seldin [2020], as an external wrapper (Algorithm 4), improving the delay term from $O(\sqrt{d_{\text{tot}}})$ to $O(\min_{Q \subseteq [T]}\{|Q| + (\sum_{t \notin Q} d_t)^{\frac{1}{2}}\})$. This yields substantial gains when a few rounds have exceptionally large delays (Corollary G.2).

[4] For example, if $d_1 = \Theta(T)$ and $d_t = 0$ for all $t \geqslant 2$, then $d_{\max} = d_{\text{tot}} = \Theta(T)$ while $\sigma_{\max} = 1$.

Table 1: Notation: T is the horizon, k is the dimension, $d_{\text{tot}} = \sum_{t=1}^{T} d_t$ is the total delay, $d_{\max} = \max_{t\in[T]} d_t$ is the maximum delay, and $\sigma_{\max} \leqslant d_{\max}$ is the maximum number of outstanding observations at any round. The bounds of Wan et al. [2024a] require advance knowledge of $d_{\max}$ and T; all other algorithms are fully adaptive.

Loss type	**Regret bounds for bandit convex optimization with delays**		
	Bistritz et al. [2022]	Wan et al. [2024a]	Our work
Convex	$(Td_{\text{tot}})^{1/3}\sqrt{k} + T^{3/4}k$	$\sqrt{Td_{\max}} + T^{3/4}\sqrt{k}$	$\sqrt{d_{\text{tot}}} + T^{3/4}\sqrt{k}$
Strongly convex	N/A	$d_{\max}\ln T + (T^2\ln T)^{1/3}k^{2/3}$	$\min\{\sigma_{\max}\ln T, \sqrt{d_{\text{tot}}}\} + (T^2\ln T)^{1/3}k^{2/3}$

Table 2: Regret bounds for delayed OCO, two-point BCO, and delayed two-point BCO.

Loss type	**Regret bounds**		
	Delayed OCO	**Two-point BCO**	**Delayed two-point BCO**
	Quanrud and Khashabi [2015], Qiu et al. [2025]	Shamir [2017]	Our work
Convex	$\sqrt{d_{\text{tot}}} + \sqrt{T}$	$\sqrt{Tk}$	$\sqrt{d_{\text{tot}}} + \sqrt{Tk}$
Strongly convex	$\min\{\sigma_{\max}\ln T, \sqrt{d_{\text{tot}}}\} + \ln T$	$k\ln T$	$\min\{\sigma_{\max}\ln T, \sqrt{d_{\text{tot}}}\} + k\ln T$

Related work. Methodologically, the work closest to ours is Joulani et al. [2016], who studied delayed OCO with full-information feedback. They propose a black-box-style reduction that feeds observations to a base (non-delayed) algorithm in order of arrival, bounding regret in terms of the base algorithm's regret and prediction drift. With known d_{tot}, they achieve regret $O(\sqrt{T + d_{\text{tot}}})$. They also establish a bound of $O(\sqrt{T\tau^*})$, where τ^* is the maximum number of observations that can arrive while some earlier feedback is still outstanding, known in advance. Using the tools developed in this paper, we show that $\tau^* = \Theta(d_{\max})$, thus expressing this result in standard notation as $O(\sqrt{Td_{\max}})$.

Delayed online learning traces back to Weinberger and Ordentlich [2002], whose subsampling argument for fixed delays $d_t = d_0$ implies minimax rates of order $\Theta(\sqrt{T(d_0+1)})$ and $\Theta((d_0+1)\ln T)$ when applied to non-delayed problems with square-root and logarithmic regret, respectively. These fixed-delay rates imply $\Omega(\sqrt{d_{\text{tot}} + T})$ and $\Omega((\sigma_{\max}+1)\ln T)$ lower bounds for arbitrary delays under convex and strongly convex losses, respectively. Matching upper bounds for first-order OCO were obtained by Quanrud and Khashabi [2015] for convex losses and by Qiu et al. [2025] for strongly convex losses; earlier, Wan et al. [2022] achieved the rate $O(d_{\max}\ln T)$. Héliou et al. [2020] first extended the bandit gradient descent method of Flaxman et al. [2005] to delayed BCO. Bistritz et al. [2022] developed the first fully delay-adaptive algorithm using a delayed doubling trick, and Wan et al. [2024a] improved the bound via a blocking strategy in the regime when $d_{\max} \lesssim d_{\text{tot}}^{2/3}/T^{1/3}$.

For additional related work on OCO and online learning with delayed feedback, see Appendix A.

2 Problem Setting

Let $T \in \mathbb{N}$ denote the time horizon, and let $\mathcal{K} \subset \mathbb{R}^k$ be a non-empty, convex, and closed domain for the dimension $k \in \mathbb{N}$. We equip $\mathbb{R}^k$ with a norm $\|.\|$ and its dual norm $\|.\|_\star$.

The player interacts with an environment determined by an oblivious adversary over T rounds (Figure 1). Before the game begins, the adversary selects delays d_t and convex, differentiable loss functions $f_t : \mathcal{K} \to \mathbb{R}$ for each round. Every round t, the player predicts $x_t \in \mathcal{K}$, receiving the corresponding feedback (gradient $\nabla f_t(x_t)$ or value $f_t(x_t)$, depending on the feedback model) at the

Online Convex Optimization with Delays

- *Latent parameters:* number of rounds T.
- *Pre-game:* adversary selects loss functions $f_t : \mathcal{K} \to \mathbb{R}$ and delays $d_t \in \{0, \dots, T-t\}$ for $t \in [T]$.

For each round $t = 1, 2, \dots, T$:

1. The player predicts $x_t \in \mathcal{K}$, incurring loss $f_t(x_t)$.
2. For s such that $s + d_s = t$, the environment reveals $(s, \nabla f_s(x_s))$ in the first-order feedback model and $(s, f_s(x_s))$ in the bandit feedback model.

Figure 1: OCO with delays under first-order or bandit feedback.

end of round $t + d_t$. We assume $d_t \in \{0, \dots, T - t\}$ for all $t \in [T]$, since feedback arriving after round T cannot affect any prediction.

The regret against a comparator $u \in \mathcal{K}$ is defined as $R_T(u) = \sum_{t=1}^{T}(f_t(x_t) - f_t(u))$, and its expected regret as $\bar{R}_T(u) = \mathbb{E}[R_T(u)]$. Letting $x^* \in \operatorname{argmin}_{x\in\mathcal{K}} \sum_{t=1}^{T} f_t(x)$ denote the best action in hindsight, the objective is to minimize $R_T(x^*)$ almost surely or its expectation $\bar{R}_T(x^*)$.

Regularity assumptions. We impose the following standard regularity assumptions, which are necessary to obtain sublinear regret guarantees in online convex optimization.

Assumption 2.1. *The diameter of $\mathcal{K}$ is bounded by D, i.e.,* $\sup_{x,y\in\mathcal{K}} \|x - y\| \leqslant D$.

Assumption 2.2. *For all $t \in [T]$, f_t has its gradient norm bounded by G, i.e.,* $\sup_{x\in\mathcal{K}} \|\nabla f_t(x)\|_\star \leqslant G$.

Assumptions 2.1 and 2.2 hold throughout the paper. When deriving stronger guarantees, we also impose the following strong convexity condition, stated explicitly when invoked.

Assumption 2.3. *For all $t \in [T]$, f_t is λ-strongly convex for $\lambda > 0$, i.e., $f_t(y) \geqslant f_t(x) + \langle \nabla f_t(x), y - x\rangle + \frac{\lambda}{2}\|y - x\|^2$ for all $x, y \in \mathcal{K}$.*

For BCO, we take $\|\cdot\|$ to be the Euclidean norm $\|\cdot\|_2$ and impose two additional assumptions.

Assumption 2.4. *For all $t \in [T]$, absolute values of f_t are bounded by M, i.e.,* $\sup_{x\in\mathcal{K}} |f_t(x)| \leqslant M$.

Assumption 2.5. *The domain satisfies $r\mathbb{B}^k \subseteq \mathcal{K} \subseteq R\mathbb{B}^k$ for some radii $0 < r < R$.*

Steady algorithms. An algorithm in this setting is a (possibly randomized) mapping from the history of received feedback to predictions. The class of *steady algorithms* is formally defined below.

Definition 2.6 (Steady Algorithm)**.** An algorithm for online learning with delays is *steady* if it does not change its prediction after rounds without feedback: $x_t = x_{t+1}$ whenever $\{s : s + d_s = t\} = \varnothing$. It is *steady in expectation* if $\mathbb{E}[x_t] = \mathbb{E}[x_{t+1}]$ whenever $\{s : s + d_s = t\} = \varnothing$.

Notation. For $n \in \mathbb{N}$, we let $[n] = \{1, \dots, n\}$. For any sequence $\{\mathrm{a}_t\}_{t=1}^{T}$, we write $\mathrm{a}_{\mathrm{tot}} = \sum_{t=1}^{T} \mathrm{a}_t$ and $\mathrm{a}_{\max} = \max_{t\in[T]} \mathrm{a}_t$. We denote by $\mathbb{B}^k$ and $\mathbb{S}^{k-1}$ the unit Euclidean ball and sphere centered at the origin in $\mathbb{R}^k$, respectively. For any $S \subseteq \mathbb{R}^k$ and $c \in \mathbb{R}$, we let $cS = \{cx : x \in S\}$.

3 Continuous Time Model for Delays

The classical formulation of online learning with delays assumes a discrete, round-based time structure: a player makes a prediction in round t and receives feedback in round $t + d_t$. Real-world systems, however, generate predictions and observations asynchronously. Rather than forcing these events into discrete rounds, we adopt a continuous time model that places them on a timeline $\mathbb{R}_+$.

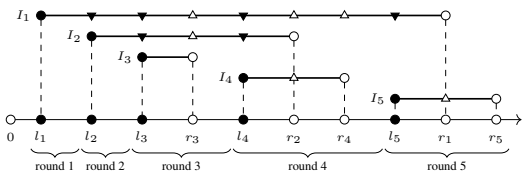
Definition 3.1 (Continuous Time Model (CTM))**.** The timing of predictions and observations is determined by latency-intervals $\{I_t\}_{t=1}^{T}$, where $I_t = (l_t, r_t) \subset \mathbb{R}_+$. For every $t \in [T]$, the environment requests a prediction at time l_t and reveals the corresponding observation at time r_t. All timestamps $\{l_t, r_t\}_{t=1}^{T}$ are distinct, and intervals are indexed in increasing order of prediction times: $0 < l_1 < \ldots < l_T$.

In the round-based model, the delay sequence $\{d_t\}_{t=1}^{T}$ fully specifies the time structure. In particular, it determines the number of outstanding observations in each round, $\sigma_t = \sum_{s=1}^{t-1} \mathbb{I}(s + d_s \geqslant t)$, which is commonly used in the literature; for brevity, we refer to σ_t as the *backlog* in this paper. Under the CTM, both delays and backlogs arise directly from the geometry of the latency intervals.

Section 3.1 formalizes this structure. Building on it, Section 3.2 introduces an observation-centric view of the problem that complements the standard prediction-centric perspective. We show that steady algorithms admit a natural regret decomposition: non-delayed regret on observation-ordered losses plus delay-induced prediction drift (Lemma 3.7).

3.1 Delays, Backlogs, and Their Duals

Under the CTM, rounds of the delayed problem (Fig. 1) naturally correspond to time-intervals: round t spans the interval $[l_t, l_{t+1})$, during which the t-th prediction is made at time l_t, and observations



	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
d_t	4	2	0	0	0
σ_t	0	1	2	2	1
$d_t^\star$	3	1	0	1	1
$\sigma_t^\star$	1	2	2	1	0
$\beta(t)$	4	2	1	3	5

Figure 2: Example of latency-intervals for $T=5$. The table provides $d_t, \sigma_t, d_t^\star, \sigma_t^\star$ and $\beta : [T] \to [T]$ from Theorem 3.4.

arrive at times $\{r_s : r_s \in (l_t, l_{t+1})\}$. Therefore, the discrete *delay* d_t and *backlog* σ_t arise as

$$d_t = |\{s : l_s \in I_t\}| \quad \text{and} \quad \sigma_t = |\{s : l_t \in I_s\}|, \tag{1}$$

where d_t counts how many predictions are made between the t-th prediction and observing its feedback, and σ_t counts how many previous predictions still await feedback immediately before making the t-th prediction.

The next lemma and remark show that these definitions recover the classical round-based formulation.

Lemma 3.2. *For quantities in* (1)*, it holds that:*

$$r_t \in (l_{t+d_t}, l_{t+d_t+1}) \quad \textit{and} \quad \sigma_t = |\{s : s < t \leqslant s + d_s\}|.$$

Proof. Exactly $|\{s : l_s < r_t\}| = |\{s : l_s \leqslant l_t\}| + |\{s : l_t < l_s < r_t\}| = t + d_t$ prediction times precede time r_t, so $r_t \in (l_{t+d_t}, l_{t+d_t+1})$. Then, the identity $\sigma_t = |\{s : l_s < l_t < r_s\}|$ follows immediately. □

Remark 3.3 (Consistency and Realizability in the CTM). Lemma 3.2 confirms that our definitions are consistent with round-based protocols: d_t equals the number of round-intervals $[l_s, l_{s+1})$ spanned by I_t, placing observation time r_t in round $t + d_t$; σ_t matches the classical definition as the number of outstanding delays. Conversely, any round-based delayed-feedback protocol can be represented within the CTM. Specifically, given a delay sequence $\{d_t\}_{t=1}^T$, we can set $(l_t, r_t) = (t, t + d_t + 1 - 2^{-t})$ so that round t spans interval $[t, t+1)$ and observation time r_t clearly falls within round $t + d_t$ spanned by interval $[t + d_t, t + d_t + 1)$.

To characterize the delay structure in the CTM, we introduce *dual-delay* and *dual-backlog* sequences

$$d_t^\star = |\{s : r_s \in I_t\}| \quad \text{and} \quad \sigma_t^\star = |\{s : r_t \in I_s\}|. \tag{2}$$

Here, dual-delay $d_t^\star$ counts observations between the t-th prediction and its observation, while $\sigma_t^\star$ counts how many predictions still await feedback when receiving observation for the t-th prediction.

The following theorem establishes the fundamental properties and relations among the four sequences.

Theorem 3.4. *For quantities in* (1) *and* (2)*, it holds that:*

(a) $\sum_{t=1}^T d_t = \sum_{t=1}^T \sigma_t = \sum_{t=1}^T d_t^\star = \sum_{t=1}^T \sigma_t^\star$,
(b) $\sigma_{\max} = \sigma_{\max}^\star, \quad \frac{1}{2} d_{\max} \leqslant d_{\max}^\star \leqslant 2 d_{\max}$,
(c) $d_t^\star = \sigma_t + \beta(t) - t, \quad \sigma_t^\star = d_t + t - \beta(t)$,

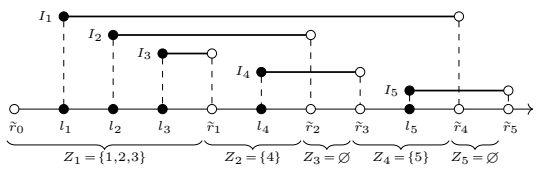
where $\beta : [T] \to [T]$ *is the permutation* $\beta(t) = |\{s : r_s \leqslant r_t\}|$.

In particular, all four sequences sum to the total delay d_{tot}. Figure 2 illustrates this in a concrete example: delays d_t (dual-delays $d_t^\star$) correspond to black (white) triangles along intervals I_t, while backlogs σ_t (dual-backlogs $\sigma_t^\star$) correspond to black (white) triangles at timestamps l_t (r_t), with the total number of triangles of each color $d_{\text{tot}} = 6$.

See Appendix C for proofs of the results in this section, as well as additional properties of the CTM.

3.2 Observation-Ordering and Steady Algorithms

The CTM naturally suggests an observation-centric perspective, in which we order quantities by observation times rather than by prediction times. This viewpoint complements the standard prediction-centric view and is essential for our analysis of steady algorithms. To facilitate this, we introduce the notational convention for *observation-ordering*, which we adopt throughout the paper.



	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
$\widetilde{d}_n$	0	2	0	4	0
$\widetilde{\sigma}_n$	2	1	2	0	1
$\widetilde{d}^\star_n$	0	1	1	3	1
$\widetilde{\sigma}^\star_n$	2	2	1	1	0
$\rho(n)$	3	2	4	1	5

Figure 3: Partition of rounds induced by the latency intervals in Fig. 2: if $t \in Z_n$ (equivalently, $l_t \in [\widetilde{r}_{n-1}, \widetilde{r}_n)$), then a steady algorithm outputs $x_t = z_n$. The table provides corresponding values $\widetilde{d}_n, \widetilde{\sigma}_n, \widetilde{d}^\star_n, \widetilde{\sigma}^\star_n$ and permutation $\rho : [T] \to [T]$.

Definition 3.5 (Observation-Ordering)**.** Under the CTM, let $\rho\colon[T]\to[T]$ be the permutation such that $r_{\rho(1)}<\ldots<r_{\rho(T)}$. For any sequence $\{\mathrm{a}_t\}_{t=1}^T$ indexed by $t\in[T]$, define its observation-ordering $\{\widetilde{\mathrm{a}}_n\}_{n=1}^T$ by $\widetilde{\mathrm{a}}_n=\mathrm{a}_{\rho(n)}$ for $n\in[T]$.

For ease of notation, we use subscripts $n \in [T]$ for observation-ordered quantities and $t \in [T]$ for prediction-ordered quantities throughout the paper.

Under observation-ordering, the n-th observation arrives at time $\widetilde{r}_n$ for the prediction $\widetilde{x}_n$ made at time $\widetilde{l}_n$ and evaluated on loss function $\widetilde{f}_n$. Setting $\widetilde{r}_0 = 0$, we obtain a chronological sequence of observation times $0 = \widetilde{r}_0 < \widetilde{r}_1 < \ldots < \widetilde{r}_T$ such that all timestamps lie within $[\widetilde{r}_0, \widetilde{r}_T]$. Since no two timestamps coincide, each prediction time $\widetilde{l}_n$ lies strictly between two consecutive observation times $\{\widetilde{r}_n\}_{n=0}^T$. The following lemma, the observation-centric analogue of Lemma 3.2, characterizes this interleaving and expresses the dual-backlog $\widetilde{\sigma}^\star_n$ as the number of outstanding dual-delays.

Lemma 3.6. *For quantities in* (2)*, it holds that:*

$$\widetilde{l}_n\in(\widetilde{r}_{n-\widetilde{d}^\star_n-1},\widetilde{r}_{n-\widetilde{d}^\star_n}) \quad \textit{and} \quad \widetilde{\sigma}^\star_n=|\{m:m-\widetilde{d}^\star_m\leqslant n<m\}|.$$

Proof. Exactly $|\{m:\widetilde{r}_m<\widetilde{l}_n\}|=|\{m:\widetilde{r}_m<\widetilde{r}_n\}|-|\{m:\widetilde{l}_n<\widetilde{r}_m<\widetilde{r}_n\}| = n-1-\widetilde{d}^\star_n$ observations precede time $\widetilde{l}_n$, so $\widetilde{l}_n \in (\widetilde{r}_{n-\widetilde{d}^\star_n-1}, \widetilde{r}_{n-\widetilde{d}^\star_n})$, and $\widetilde{\sigma}^\star_n = |\{m : \widetilde{l}_m < \widetilde{r}_n < \widetilde{r}_m\}|$ follows. □

The observation-centric perspective is particularly useful for analyzing steady algorithms (Definition 2.6). Since a steady algorithm does not change its prediction in rounds without feedback, it admits a *base-prediction sequence* $\{z_n\}_{n=1}^T \subset \mathcal{K}$ such that the prediction remains constant between consecutive observation times: for each $n \in [T]$ and every prediction time $l_t \in (\widetilde{r}_{n-1}, \widetilde{r}_n)$, the algorithm outputs $x_t = z_n$. Equivalently, the rounds $[T]$ partition into sets $Z_n = \{t : l_t \in [\widetilde{r}_{n-1}, \widetilde{r}_n)\}$, with $x_t = z_n$ whenever $t \in Z_n$. Figure 3 illustrates this for the example in Figure 2.

Lemma 3.6 links the observation-ordered predictions to the base-predictions through the dual-delays: $\widetilde{x}_n = z_{n-\widetilde{d}^\star_n}$. Substituting this identity, we can write the regret of any steady algorithm entirely in terms of the base sequence z_n evaluated on the observation-ordered losses $\widetilde{f}_n$. For G-Lipschitz losses, this yields the following decomposition:

Lemma 3.7 (Regret decomposition)**.** *Let $f_t : \mathcal{K} \to \mathbb{R}$ be G-Lipschitz for every $t \in [T]$, with no convexity or smoothness required. For any steady algorithm with base-prediction sequence $\{z_n\}_{n=1}^T \subset \mathcal{K}$ and every comparator $u \in \mathcal{K}$,*

$$R_T(u) \leqslant \underbrace{\textstyle\sum_{n=1}^T\big(\widetilde{f}_n(z_n)-\widetilde{f}_n(u)\big)}_{\text{Non-Delayed Regret}} + G\underbrace{\textstyle\sum_{n=1}^T\|z_n-z_{n-\widetilde{d}^\star_n}\|}_{\text{Prediction Drift}}.$$

Proof. Since ρ is a permutation and $\widetilde{x}_n = z_{n-\widetilde{d}^\star_n}$, reindexing the losses gives

$$\begin{aligned} R_T(u) &= \textstyle\sum_{n=1}^T\big(\widetilde{f}_n(z_{n-\widetilde{d}^\star_n}) - \widetilde{f}_n(u)\big) \\ &= \textstyle\sum_{n=1}^T\big(\widetilde{f}_n(z_n) - \widetilde{f}_n(u)\big) + \sum_{n=1}^T\big(\widetilde{f}_n(z_{n-\widetilde{d}^\star_n}) - \widetilde{f}_n(z_n)\big). \end{aligned}$$

By G-Lipschitzness, each term in the second sum is at most $G\|z_{n-\widetilde{d}^\star_n} - z_n\|$, proving the bound. □

Here, the *non-delayed regret* is the regret of the sequence $\{z_n\}_{n=1}^T$ on the observation-ordered losses $\{\widetilde{f}_n\}_{n=1}^T$ and the *prediction drift* captures the cost of the played predictions $\widetilde{x}_n = z_{n-\widetilde{d}_n^\star}$ lagging behind base-predictions z_n. This decomposition underlies our reduction from online learning with delays to online learning with drift-penalization, which we define in the following section.

4 Online Learning with Drift Penalization

This section introduces online linear optimization with drift penalization (Figure 4), a game that serves as the foundation for reductions in subsequent sections. Crucially, this formulation involves no delays; the connection to delayed feedback emerges only through the reductions.

In this game, a player selects predictions from $\mathcal{K}$ (the same domain as in OCO with delays) over N rounds, incurring in each round a linear loss and a penalty for moving too far from past predictions, measured against one of them. Specifically, in round n, the player chooses z_n and incurs the loss $\langle c_n, z_n \rangle$ plus the drift penalty $\|z_n - z_{n-\lambda_n}\|$, specified by the lag λ_n. The loss vectors and lags $\{c_n, \lambda_n\}_{n=1}^N$ are fixed by an adversary before the game begins. To update his strategy, after each round the player observes (c_n, λ_n) as well as the number of outstanding lags $\nu_n = |\{m : m - \lambda_m \leqslant n < m\}|$, which summarizes how many future penalties will involve predictions up to round n.

Online Linear Optimization with Drift Penalization
- *Latent parameters:* number of rounds N.
- *Pre-game:* adversary picks loss vectors $c_n \in \mathbb{R}^k$ and lags $\lambda_n \in \{0, \dots, n-1\}$ for $n \in [N]$.

For each round $n = 1, 2, \dots, N$:

1. The player predicts $z_n \in \mathcal{K}$, incurring loss $\langle c_n, z_n \rangle$ and drift penalty $\|z_{n-\lambda_n} - z_n\|$.
2. The environment reveals a tuple (c_n, λ_n, ν_n), where $\nu_n = |\{m : m - \lambda_m \leqslant n < m\}|$ denotes the number of outstanding lags.

Figure 4: Online learning with drift penalization for linear losses.

To measure performance, we define the drift-penalized regret with weight $W \geqslant 0$ against $u \in \mathcal{K}$ as

$$\mathfrak{R}_N^{\text{drift}}(u; W) = \mathfrak{R}_N(u) + W\mathfrak{D}_N, \tag{3}$$

where $\mathfrak{R}_N(u) = \sum_{n=1}^N \langle c_n, z_n - u \rangle$ is the standard regret and $\mathfrak{D}_N = \sum_{n=1}^N \|z_{n-\lambda_n} - z_n\|$ is the total drift penalty. Any online linear optimization algorithm can be applied to this setting, with its standard regret bounding $\mathfrak{R}_N(u)$; moreover, stable algorithms naturally incur smaller drift penalty $\mathfrak{D}_N$.

P-FTRL and OMD Updates. We analyze how Proximal Follow-The-Regularized-Leader (P-FTRL) and Online Mirror Descent (OMD) perform in online learning with drift penalization. Given a non-increasing learning rate sequence $\{\eta_n\}_{n=1}^N \subset (0, \infty)$ and initial point $z_1 \in \mathcal{K}$, these algorithms select predictions sequentially according to the following respective update rules:

$$\text{P-FTRL}\colon\ z_{n+1} = \operatorname{argmin}_{z \in \mathcal{K}} \textstyle\sum_{m=1}^n (\langle c_m, z \rangle + \frac{\alpha_m \|z_m - z\|^2}{2}), \tag{4}$$

$$\text{OMD}\colon\ z_{n+1} = \operatorname{argmin}_{z \in \mathcal{K}} (\langle c_n, z \rangle + \frac{\|z_n - z\|^2}{2\eta_n}), \tag{5}$$

where $\{\alpha_n\}_{n=1}^N$ is given by $\alpha_1 = \frac{1}{\eta_1}, \alpha_{n+1} = \frac{1}{\eta_{n+1}} - \frac{1}{\eta_n}$. The next theorem bounds regret and drift.

Theorem 4.1. *For every $u \in \mathcal{K}$, predictions $\{z_n\}_{n=1}^N$ generated by P-FTRL (4) or OMD (5) satisfy*

$$\mathfrak{R}_N(u) \leqslant \textstyle\sum_{n=1}^N \frac{1/\eta_n - 1/\eta_{n-1}}{2} \|z_n - u\|^2 + \frac{1}{2} \sum_{n=1}^N \eta_n \|c_n\|_\star^2 \quad \textit{and} \quad \mathfrak{D}_N \leqslant \textstyle\sum_{n=1}^N \eta_n \nu_n \|c_n\|_\star.$$

The regret bound is standard (e.g., Orabona [2019]). The drift bound arises by decomposing drift into single-step terms satisfying $\|z_{n+1} - z_n\| \leqslant \eta_n \|c_n\|_\star$; the dependence on ν_n is the key to our delay-adaptive analysis. We also establish a novel drift guarantee for P-FTRL, which is essential for our BCO results.

Lemma 4.2. *P-FTRL (4) generates $\{z_n\}_{n=1}^N$ such that $\mathfrak{D}_N \leqslant \sum_{n=1}^N \eta_n \| \sum_{m=n-\lambda_n}^{n-1} c_m \|_\star + D\, H_\eta$, where $H_\eta = \sum_{n=1}^N (1 - \frac{\eta_n}{\eta_{n-\lambda_n}})$ and D is an upper bound on the diameter of $\mathcal{K}$ (Assumption 2.1).*

5 Blackbox Reductions and Main Results

Here, we develop algorithms for delayed OCO and BCO via reduction to drift-penalized online linear optimization. The key construction is a wrapper $\mathcal{W}_{\text{OCO}}$ (Algorithm 1) that takes any algorithm $\mathcal{B}$ for drift-penalized OLO as a blackbox and produces a new algorithm $\mathcal{W}_{\text{OCO}}(\mathcal{B})$ for delayed OCO. The wrapper feeds observation-ordered gradients $\widetilde{g}_n = \nabla \widetilde{f}_n(\widetilde{x}_n)$ as loss vectors to $\mathcal{B}$, with dual-delays $\widetilde{d}_n^\star$ as lags and dual-backlogs $\widetilde{\sigma}_n^\star$ as numbers of outstanding lags. The base algorithm $\mathcal{B}$ produces base-predictions that the wrapper outputs as its own predictions. An analogous wrapper $\mathcal{W}_{\text{BCO}}$ (Algorithm 2) handles bandit feedback by using single-point gradient estimates from the classical work by Flaxman et al. [2005].

Remark 5.1 (Feasibility). This reduction is feasible because (i) the dual-backlog $\widetilde{\sigma}_n^\star$ equals the number of outstanding dual-delays by Lemma 3.6, and (ii) the wrapper runs updates on $\mathcal{B}$ only at observation times $\widetilde{r}_n$, when all required quantities are available. Specifically, at time $\widetilde{r}_n$, the gradient $\widetilde{g}_n = \nabla \widetilde{f}_n(\widetilde{x}_n)$ (or the value $\widetilde{f}_n(\widetilde{x}_n)$ in the case of bandit feedback) is revealed, while the dual-delay $\widetilde{d}_n^\star = |\{m : \widetilde{l}_n < \widetilde{r}_m < \widetilde{r}_n\}|$ and the dual-backlog $\widetilde{\sigma}_n^\star = |\{m : \widetilde{l}_m < \widetilde{r}_n < \widetilde{r}_m\}|$ can be computed from the observed history up to time $\widetilde{r}_n$.

5.1 Delayed OCO via Drift-Penalized OLO

We now present the wrapper $\mathcal{W}_{\text{OCO}}$ for delayed OCO with first-order feedback. The wrapper maintains a current base-prediction $\bar{z}$ initialized by $\mathcal{B}$, outputs $\bar{z}$ whenever a prediction is requested, and updates $\bar{z}$ by forwarding each received gradient to $\mathcal{B}$ along with the dual-delay and dual-backlog.

Algorithm 1: $\mathcal{W}_{\text{OCO}}(\mathcal{B})$

Access : Algorithm $\mathcal{B}$ for OLO with drift penalization.

Initialize $\bar{z}$ with the initial output of $\mathcal{B}$.

For round $t = 1, 2, \ldots, T$:

1. Predict $x_t = \bar{z}$.
2. **For** each s such that $s + d_s = t$, in observation order:
   - Receive $(s, \nabla f_s(x_s))$, send $(\nabla f_s(x_s), d_s^\star, \sigma_s^\star)$ to $\mathcal{B}$, and set $\bar{z}$ to the new output of $\mathcal{B}$.

This algorithm is steady. It updates only upon receiving feedback and outputs the current base-prediction otherwise. Effectively, the algorithm updates $\mathcal{B}$ sequentially on observation-ordered tuples $(\widetilde{g}_n, \widetilde{d}_n^\star, \widetilde{\sigma}_n^\star)$, where $\widetilde{g}_n = \nabla \widetilde{f}_n(\widetilde{x}_n)$ (i.e., $g_t = \nabla f_t(x_t)$), producing base-predictions z_n at observation times $\widetilde{r}_{n-1}$ that yield predictions $\widetilde{x}_m = z_{m-\widetilde{d}_m^\star}$ at prediction times $\widetilde{l}_m \in (\widetilde{r}_{m-\widetilde{d}_m^\star-1}, \widetilde{r}_{m-\widetilde{d}_m^\star})$.

The next theorem shows that the regret of $\mathcal{W}_{\text{OCO}}(\mathcal{B})$ is controlled by the drift-penalized regret of $\mathcal{B}$.

Theorem 5.2 (Regret of Algorithm 1)**.** *For any base algorithm $\mathcal{B}$ and $u \in \mathcal{K}$, $\mathcal{W}_{\text{OCO}}(\mathcal{B})$ guarantees*

$$R_T(u) \leqslant \mathfrak{R}_T^{\text{drift}}(u; G),$$

where $\mathfrak{R}_T^{\text{drift}}(u; W)$ is the drift-penalized regret (3) of $\mathcal{B}$ for loss vectors $c_n = \widetilde{g}_n$ and lags $\lambda_n = \widetilde{d}_n^\star$.

Under λ-strong convexity (2.3), it further holds that $R_T(u) \leqslant \mathfrak{R}_T^{\text{drift}}(u; 3G) - \frac{\lambda}{2}\sum_{n=1}^T \|z_n - u\|^2$.

Instantiating $\mathcal{B}$ with (4) or (5) with appropriate learning rates yields the following guarantees.

Corollary 5.3. *$\mathcal{W}_{\text{OCO}}(\mathcal{B})$ with $\mathcal{B}$ as P-FTRL (4) or OMD (5) using $\eta_n = \frac{D/G}{\sqrt{n+\sum_{m=1}^n \widetilde{\sigma}_m^\star}}$, guarantees*

$$R_T(x^*) = O(GD\,[\sqrt{d_{\text{tot}}} + \sqrt{T}]).$$

Corollary 5.4. *Under λ-strong convexity (2.3), $\mathcal{W}_{\text{OCO}}(\mathcal{B})$, with $\mathcal{B}$ as P-FTRL (4) using $\eta_n = \frac{1}{n\lambda}$, guarantees $R_T(x^*) = O(\frac{G^2}{\lambda}\left[\min\left\{\sigma_{\max}\ln T,\, \sqrt{d_{\text{tot}}}\right\} + \ln T\right])$. Running OMD (5) with $\eta_n = \frac{1}{n\lambda}$ instead guarantees $R_T(x^*) = O(\frac{G^2}{\lambda}(\sigma_{\max}+1)\ln T)$.*

In both corollaries, the learning rates form a deterministic sequence that can be computed online: η_n depends only on n and the dual-backlogs $\widetilde{\sigma}_1^\star, \ldots, \widetilde{\sigma}_n^\star$, all of which are available at the n-th update. The proofs apply Theorem 4.1 and Lemma 4.2 with our learning rates. Proofs are in Appendix E.

5.2 Delayed BCO via Drift-Penalized OLO

We adapt the wrapper approach to bandit convex optimization, where only scalar loss values are observed, by using single-point gradient estimation from Flaxman et al. [2005].

Under Assumption 2.5, i.e., $r\mathbb{B}^k \subseteq \mathcal{K} \subseteq R\mathbb{B}^k$, we have $(1-\delta/r)\,\mathcal{K} + \delta\,\mathbb{B}^k \subseteq \mathcal{K}$ for all $\delta \in (0, r]$. This allows us to define, for any $\delta \in (0, r]$ and integrable $f : \mathcal{K} \to \mathbb{R}$, the δ-smoothing of f as $f^\delta : (1-\delta/r)\mathcal{K} \to \mathbb{R}$ given by $f^\delta(x) = \mathbb{E}_{v\sim\mathrm{Unif}(\mathbb{B}^k)}[f(x+\delta v)]$. The key properties of this construction are summarized in the following theorem.

Theorem 5.5 (Flaxman et al. [2005])**.** *For any $\delta \in (0, r]$ and integrable $f : \mathcal{K} \to \mathbb{R}$:*

1. *The δ-smoothing f^δ is differentiable with gradients $\nabla f^\delta(x) = \frac{k}{\delta}\mathbb{E}_{u\sim\mathrm{Unif}(\mathbb{S}^{k-1})}[f(x+\delta u)\,u]$.*
2. *If f is convex (λ-strongly convex), so is f^δ.*
3. *If f is G-Lipschitz, then $|f^\delta(x) - f(x)| \leqslant G\delta$ and $\|\nabla f^\delta(x)\| \leqslant G$ for $x \in (1-\delta/r)\mathcal{K}$.*

The wrapper $\mathcal{W}_{\mathrm{BCO}}$ (Algorithm 2) follows the structure of $\mathcal{W}_{\mathrm{OCO}}$, but replaces true gradients with single-point estimates. The algorithm takes as input a non-increasing sequence of smoothing parameters $(\delta_t)_{t\geqslant 1} \subset (0, r]$, that can be specified independently of horizon T when it's unknown. In round t, given the current base-prediction $y_t = \bar{z}$ from $\mathcal{B}$, the algorithm samples an exploration direction $u_t \sim \mathrm{Unif}(\mathbb{S}^{k-1})$, predicts $x_t = (1-\delta_t/r)y_t + \delta_t u_t$, and upon receiving loss $f_s(x_s)$ constructs the gradient estimate $\widehat{g}_s^\delta = \frac{k}{\delta_s} f_s(x_s)u_s$, which it forwards to $\mathcal{B}$ along with dual-delay $d_s^\star$ and dual-backlog $\sigma_s^\star$ to update $\bar{z}$. By Theorem 5.5, we have $\mathbb{E}[\widehat{g}_t^\delta \mid y_t] = \nabla f_t^{\delta_t}((1-\delta_t/r)y_t)$.

Algorithm 2: $\mathcal{W}_{\mathrm{BCO}}(\mathcal{B})$

Access : Algorithm $\mathcal{B}$ for OLO with drift penalization.
Parameters : Non-increasing sequence $(\delta_t)_{t\geqslant 1} \subset (0, r]$.

Initialize $\bar{z}$ with initial output of $\mathcal{B}$.

For round $t = 1, 2, \ldots, T$:

1. Sample $u_t \sim \mathrm{Unif}(\mathbb{S}^{k-1})$; predict $x_t = (1-\frac{\delta_t}{r})\bar{z} + \delta_t u_t$.
2. **For** each s such that $s + d_s = t$, in observation order:
   - Receive $(s, f_s(x_s))$, send $(\frac{k}{\delta_s} f_s(x_s)u_s, d_s^\star, \sigma_s^\star)$ to $\mathcal{B}$, and set $\bar{z}$ to the new output of $\mathcal{B}$.

Unlike the OCO case, the resulting algorithm is not steady due to random exploration, but is steady in expectation for a fixed schedule $\delta_t = \delta$. Nevertheless, since the current base-prediction is updated only upon receiving feedback, the same analysis applies. The algorithm updates $\mathcal{B}$ sequentially on tuples $(\widetilde{\widehat{g}}_n^\delta, \widetilde{d}_n^\star, \widetilde{\sigma}_n^\star)$, where $\widetilde{\widehat{g}}_n^\delta = \frac{k}{\widetilde{\delta}_n}\widetilde{f}_n(\widetilde{x}_n)\widetilde{u}_n$, producing base-predictions z_n that yield predictions $\widetilde{x}_n = (1-\widetilde{\delta}_n/r)\widetilde{y}_n + \widetilde{\delta}_n\widetilde{u}_n$ with $\widetilde{y}_n = z_{n-\widetilde{d}_n^\star}$.

The following theorem bounds the expected regret of $\mathcal{W}_{\mathrm{BCO}}(\mathcal{B})$ in terms of the drift-penalized regret of $\mathcal{B}$, with an additional bias of order $\delta_{\mathrm{tot}} = \sum_{t=1}^T \delta_t$ due to smoothing.

Theorem 5.6 (Regret of Algorithm 2)**.** *For any base-algorithm $\mathcal{B}$, non-increasing sequence $(\delta_t)_{t\geqslant 1} \subset (0, r]$, and comparator $u \in \mathcal{K}$, $\mathcal{W}_{\mathrm{BCO}}(\mathcal{B})$ guarantees*

$$\bar{R}_T(u) \leqslant \mathbb{E}\left[\mathfrak{R}_T^{\mathrm{drift}}(u; G)\right] + \tfrac{6GR\delta_{\mathrm{tot}}}{r},$$

where $\mathfrak{R}_T^{\mathrm{drift}}(u; W)$ is the drift-penalized regret (3) of $\mathcal{B}$ for loss vectors $c_n = \widetilde{\widehat{g}}_n^\delta$ and lags $\lambda_n = \widetilde{d}_n^\star$. Moreover, under λ-strong convexity (2.3), $\bar{R}_T(u) \leqslant \mathbb{E}\left[\mathfrak{R}_T^{\mathrm{drift}}(u; 3G) - \frac{\lambda}{2}\sum_{n=1}^T \|z_n - u\|^2\right] + \frac{10GR\delta_{\mathrm{tot}}}{r}$.

Selecting a smoothing sequence $(\delta_t)_{t\geqslant 1}$ and instantiating $\mathcal{B}$ with P-FTRL (4) yields our guarantees. To state them, we introduce the *dimensionless ratio* $\nu = \frac{M}{Gr}$, where M is the absolute value bound (Assumption 2.4). Since the n-th update occurs in round $\rho(n) + d_{\rho(n)}$, the observation round of feedback from round $\rho(n)$, the smoothing parameter $\delta'_n = \delta_{\rho(n)+d_{\rho(n)}}$ is available and can be incorporated into the rate η_n.

Corollary 5.7. *$\mathcal{W}_{\mathrm{BCO}}(\mathcal{B})$, where $(\delta_t)_{t\geqslant 1}$ is set to $\delta_t = r\,\min\{1, \frac{\sqrt{\nu k}}{t^{1/4}}\}$ and $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{D/G}{\sqrt{n + \sum_{m=1}^n (\widetilde{\sigma}_m^\star + (\nu k r/\delta'_m)^2)}}$, guarantees $\bar{R}_T(x^*) = O\left(GD\left[\sqrt{d_{\mathrm{tot}}} + T^{3/4}\sqrt{\nu k}\right]\right)$.*

Corollary 5.8. *Under λ-strong convexity (2.3), $\mathcal{W}_{\mathrm{BCO}}(\mathcal{B})$, where $(\delta_t)_{t\geqslant 1}$ is set to $\delta_t = r\,\min\{1, (\frac{\nu^2 k^2 \ln(et)}{t})^{1/3}\}$ and $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{1}{n\lambda}$, guarantees*

$$\bar{R}_T(x^*) = O\Big(\tfrac{G^2}{\lambda}\Big[\min\{\sigma_{\max}\ln T,\, \sqrt{d_{\mathrm{tot}}}\} + T^{2/3}(\ln T)^{1/3}(\nu k)^{2/3}\Big]\Big)\,.$$

The proofs of both corollaries combine Theorem 4.1 and Lemma 4.2, using a concentration bound for $\|\sum_{m=n-\tilde{d}_n^\star}^{n-1} \widetilde{\widehat{g}}_m^{\delta}\|$ to control the drift term in Appendix E. These bounds can also be proved directly, without invoking the reduction, by extending the delayed-FTRL analysis of Flaspohler et al. [2021] from OCO to single-point bandit gradient estimates and using the drift-control techniques developed in this paper, namely, the FTRL drift bound in Lemma 4.2 and the bound on sums of gradient estimates in Lemma E.4. Such a direct analysis is given in Ryabchenko et al. [2026, Theorem 3.2] for delayed and weighted bandit feedback, with the unweighted setting recovered by taking $w_t = 1$ for all t. The reduction-based proof is stronger, however, since it separates the effect of delayed feedback from the underlying learning algorithm and yields a reusable delayed-to-immediate framework, rather than a proof specialized to a particular bandit update.

6 Discussion and Future Work

We introduced a continuous-time model for delayed feedback that yields a black-box reduction to drift-penalized online linear optimization, where any base algorithm with regret and drift guarantees produces a delayed learner whose regret is controlled by its drift-penalized regret under the induced observation ordering. The framework extends beyond OCO to any online learning protocol where prediction drift can be quantified.

Several directions remain open. First, while our first-order guarantees extend to adaptive adversaries, extending our BCO guarantees to settings where losses and delays depend on past interactions requires additional analysis. Such settings are relevant to congestion control and strategic environments. Second, delayed feedback arises naturally in distributed optimization with asynchronous communication, and our model may provide a useful abstraction for multi-agent online learning with heterogeneous, time-varying delays, including federated learning systems. Finally, developing guarantees that adapt to temporal variation in the losses is a promising direction, requiring a refined analysis of both the base learner's regret and prediction drift.

Acknowledgments and Disclosure of Funding

Idan Attias is supported by the National Science Foundation under Grant ECCS-2217023, through the Institute for Data, Econometrics, Algorithms, and Learning (IDEAL). Daniel M. Roy is supported by the funding through NSERC Discovery Grant and Canada CIFAR AI Chair at the Vector Institute. The authors declare no competing interests.

References

Jacob Abernethy, Peter L Bartlett, Alexander Rakhlin, and Ambuj Tewari. Optimal strategies and minimax lower bounds for online convex games. In *Proceedings of the 21st Annual Conference on Learning Theory*, pages 415–424, 2008.

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the 23rd Annual Conference on Learning Theory*, pages 28–40, 2010.

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

Ilai Bistritz, Zhengyuan Zhou, Xi Chen, Nicholas Bambos, and Jose Blanchet. No weighted-regret learning in adversarial bandits with delays. *Journal of Machine Learning Research*, 23(139):1–43, 2022.

Yann Bouteiller, Simon Ramstedt, Giovanni Beltrame, Christopher Pal, and Jonathan Binas. Reinforcement learning with random delays. In *International Conference on Learning Representations*, 2021.

Sébastien Bubeck and Ronen Eldan. Multi-scale exploration of convex functions and bandit convex optimization. In *Conference on Learning Theory*, pages 583–589. PMLR, 2016.

Nicolò Cesa-Bianchi, Ofer Dekel, and Ohad Shamir. Online learning with switching costs and other adaptive adversaries. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Nicolò Cesa-Bianchi, Claudio Gentile, Yishay Mansour, and Alberto Minora. Delay and cooperation in nonstochastic bandits. In *Conference on Learning Theory*, pages 605–622. PMLR, 2016.

John C Duchi, Michael I Jordan, Martin J Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015.

Emmanuel Esposito, Saeed Masoudian, Hao Qiu, Dirk Van Der Hoeven, Nicolò Cesa-Bianchi, and Yevgeny Seldin. Delayed bandits: When do intermediate observations help? In *Proceedings of the 40th International Conference on Machine Learning*, pages 9374–9395, 2023.

Genevieve E Flaspohler, Francesco Orabona, Judah Cohen, Soukayna Mouatadid, Miruna Oprescu, Paulo Orenstein, and Lester Mackey. Online learning with optimism and delay. In *International Conference on Machine Learning*, pages 3363–3373. PMLR, 2021.

Abraham D Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 385–394, 2005.

G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 2 edition, 1952.

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

Elad Hazan and Yuanzhi Li. An optimal algorithm for bandit convex optimization. *arXiv preprint arXiv:1603.04350*, 2016.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

Amélie Héliou, Panayotis Mertikopoulos, and Zhengyuan Zhou. Gradient-free online learning in continuous games with delayed rewards. In *International Conference on Machine Learning*, pages 4172–4181. PMLR, 2020.

Tiancheng Jin, Tal Lancewicki, Haipeng Luo, Yishay Mansour, and Aviv Rosenberg. Near-optimal regret for adversarial MDP with delayed bandit feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 33469–33481, 2022.

Pooria Joulani, András György, and Csaba Szepesvári. Online learning under delayed feedback. In *International Conference on Machine Learning*, pages 1453–1461. PMLR, 2013.

Pooria Joulani, András György, and Csaba Szepesvári. Delay-tolerant online convex optimization: Unified analysis and adaptive-gradient algorithms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Tal Lancewicki, Aviv Rosenberg, and Yishay Mansour. Cooperative online learning in stochastic and adversarial MDPs. In *International Conference on Machine Learning*, pages 11918–11968. PMLR, 2022.

John Langford, Alexander J. Smola, and Martin Zinkevich. Slow learners are fast. In *Advances in Neural Information Processing Systems*, volume 22, 2009.

Tor Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. *Mathematical Statistics and Learning*, 2(3-4):311–334, 2019.

Tor Lattimore. *Bandit convex optimisation*. Cambridge University Press, 2026. doi: 10.1017/9781009607551.

Orin Levy, Liad Erez, Alon Cohen, and Yishay Mansour. Regret bounds for adversarial contextual bandits with general function approximation and delayed feedback. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Saeed Masoudian, Julian Zimmert, and Yevgeny Seldin. A best-of-both-worlds algorithm for bandits with delayed feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 11752–11762, 2022.

Saeed Masoudian, Julian Zimmert, and Yevgeny Seldin. A best-of-both-worlds algorithm for bandits with delayed feedback with robustness to excessive delays. In *Advances in Neural Information Processing Systems*, volume 37, pages 141071–141102, 2024.

H. Brendan McMahan and Matthew Streeter. Delay-tolerant algorithms for asynchronous distributed online learning. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

Chris Mesterharm. On-line learning with delayed label feedback. In *International Conference on Algorithmic Learning Theory*, pages 399–413. Springer, 2005.

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

Hao Qiu, Emmanuel Esposito, and Mengxiao Zhang. Exploiting curvature in online convex optimization with delayed feedback. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 50448–50479, 2025.

Hao Qiu, Andrew Jacobsen, Emmanuel Esposito, and Mengxiao Zhang. Parameter-free dynamic regret: Time-varying movement costs, delayed feedback, and memory. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*, pages 102119–102146. PMLR, 2026.

Kent Quanrud and Daniel Khashabi. Online learning with adversarial delays. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Alexander Ryabchenko, Idan Attias, and Daniel M. Roy. Capacity-constrained online learning with delays: Scheduling frameworks and regret trade-offs. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 4973–5014. PMLR, 2025.

Alexander Ryabchenko, Idan Attias, and Daniel M. Roy. Capacity-Constrained Online Convex Optimization with Delayed Feedback. In *Advances in Neural Information Processing Systems*, 2026.

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

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017.

Tobias Sommer Thune, Nicolò Cesa-Bianchi, and Yevgeny Seldin. Nonstochastic multiarmed bandits with unrestricted delays. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Dirk van der Hoeven, Lukas Zierahn, Tal Lancewicki, Aviv Rosenberg, and Nicolò Cesa-Bianchi. A unified analysis of nonstochastic delayed feedback for combinatorial semi-bandits, linear bandits, and MDPs. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 1285–1321. PMLR, 2023.

Yuanyu Wan, Wei-Wei Tu, and Lijun Zhang. Online strongly convex optimization with unknown delays. *Machine Learning*, 111(3):871–893, 2022.

Yuanyu Wan, Chang Yao, Mingli Song, and Lijun Zhang. Improved regret for bandit convex optimization with delayed feedback. In *Advances in Neural Information Processing Systems*, volume 37, pages 169–196, 2024a.

Yuanyu Wan, Chang Yao, Mingli Song, and Lijun Zhang. Non-stationary online convex optimization with arbitrary delays. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 49991–50011, 2024b.

Marcelo J Weinberger and Erik Ordentlich. On delayed prediction of individual sequences. *IEEE Transactions on Information Theory*, 48(7):1959–1976, 2002.

Ping Wu, Heyan Huang, and Zhengyang Liu. Online sequential decision-making with unknown delays. In *Proceedings of the ACM Web Conference 2024*, pages 4028–4036, 2024.

Julian Zimmert and Yevgeny Seldin. An optimal algorithm for adversarial bandits with arbitrary delays. In *International Conference on Artificial Intelligence and Statistics*, pages 3285–3294. PMLR, 2020.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 928–936, 2003.

A Additional Related Work

Online convex optimization. Online convex optimization (OCO) and its feedback models have been studied extensively. Early work includes Zinkevich [2003], who introduced online convex programming and established $O(\sqrt{T})$-type guarantees for online gradient methods. For first-order feedback, subsequent work established the standard $\Theta(\sqrt{T})$ minimax scaling for general convex losses (see, e.g., Abernethy et al., 2008). For λ-strongly convex losses, logarithmic regret is achievable; in particular, Hazan et al. [2007] gives $O(\ln T)$-type guarantees (matching the $\Omega(\ln T)$ lower bound of Abernethy et al., 2008). Background and further developments can be found in Hazan [2016], Orabona [2019].

Bandit convex optimization. Bandit convex optimization (BCO) was initiated by Flaxman et al. [2005], who introduced the classical single-point gradient estimator and obtained regret of order $O(T^{3/4}k^{1/2})$. Subsequently, Agarwal et al. [2010] developed optimal algorithms for multi-point bandit feedback and sharpened guarantees: in particular, for two-point feedback they obtain $\widetilde{O}(T^{1/2}k^2)$ regret for convex losses and $O(k^2 \ln T)$ regret under strong convexity; for single-point feedback, they also note improved rates under additional curvature, including a $T^{2/3}$-type guarantee for strongly convex losses (more precisely, $O((T^2 \ln T)^{1/3}k^{2/3})$ up to problem-dependent constants). Complementary results for stochastic convex optimization under bandit feedback appear in Agarwal et al. [2011]. For two-point feedback, Shamir [2017] gives an optimal algorithm with improved dimension dependence with regret $O(T^{1/2}k^{1/2})$. A line of work designs (nearly) minimax-optimal algorithms for general BCO, approaching the $\Omega(k\sqrt{T})$ lower bound of Shamir [2013]; see, e.g., Bubeck and Eldan [2016], Hazan and Li [2016], Lattimore [2019]. These methods typically incur prohibitive computational costs (exponential or high-degree polynomial dependence on k and/or T), and we refer to Lattimore [2026] for a detailed treatment and further references.

Online convex optimization with delayed feedback. Early work on online learning with delayed feedback includes Weinberger and Ordentlich [2002]. A systematic treatment of delayed-feedback online learning and reduction-style approaches appears in Joulani et al. [2013], and Joulani et al. [2016] develops a black-box-style reduction that processes feedback in order of arrival. Beyond black-box techniques, there is substantial work on algorithms tailored to delayed OCO (e.g., Langford et al., 2009, McMahan and Streeter, 2014, Quanrud and Khashabi, 2015, Flaspohler et al., 2021, Wan et al., 2022, Wu et al., 2024, Wan et al., 2024b, Qiu et al., 2025). Delayed BCO has also received focused attention, starting from delayed extensions of bandit gradient methods [Héliou et al., 2020] and followed by delay-adaptive or improved-delay-dependence results [Bistritz et al., 2022, Wan et al., 2024a]. Concurrently, Qiu et al. [2026] reduce delayed feedback to online learning with time-varying movement costs in the dynamic regret setting, which is conceptually related to our drift-penalization framework (Section 4).

Delays in related sequential decision problems. Delayed observations have been studied extensively in multi-armed bandits and related models (see, e.g., Cesa-Bianchi et al., 2016, Thune et al., 2019, Bistritz et al., 2022, Esposito et al., 2023, Masoudian et al., 2022, 2024). Delays also arise naturally under resource limitations such as capacity constraints [Ryabchenko et al., 2025]. Delayed feedback has also been studied in adversarial contextual bandits with general function approximation [Levy et al., 2025]. Delays also appear in reinforcement learning / MDP settings with delayed or asynchronous feedback (e.g., Bouteiller et al., 2021, Lancewicki et al., 2022, Jin et al., 2022, van der Hoeven et al., 2023).

B General Facts

Fact B.1. *Under Assumptions 2.2 and 2.3, the diameter of $\mathcal{K}$ is bounded by $\frac{2G}{\lambda}$, i.e.,* $\sup_{x,y\in\mathcal{K}} \|x-y\| \leqslant \frac{2G}{\lambda}$.

Proof. Consider $f = f_t$ for arbitrary $t \in [T]$. From the λ-strong convexity of f (e.g., see Orabona [2019]), it holds that for all $x, y \in \mathcal{K}$, $\lambda \|x-y\|^2 \leqslant \langle \nabla f(x) - \nabla f(y), x-y\rangle \leqslant 2G\|x-y\|$. Hence, $\sup_{x,y\in\mathcal{K}} \|x-y\| \leqslant \frac{2G}{\lambda} < \infty$. □

Fact B.2 (Rearrangement Inequality)**.** *For every choice of $N \in \mathbb{N}$, real numbers $x_1 \leqslant x_2 \leqslant ... \leqslant x_N$, $y_1 \leqslant y_2 \leqslant ... \leqslant y_N$, and index permutation $\rho : [N] \to [N]$, it holds that*

$$x_1 y_N + ... + x_N y_1 \leqslant x_1 y_{\rho(1)} + ...x_N y_{\rho(N)} \leqslant x_1 y_1 + ... + x_N y_N.$$

Proof. This fact restates Theorem 368 in Hardy et al. [1952]. □

Fact B.3 (Logarithmic Telescoping Inequality)**.** *Let $(X_n)_{n\geqslant 1}$ be non-negative reals such that $X_1 > 0$. Set $S_n = \sum_{m=1}^{n} X_m$. Then, for every $N \geqslant 1$, we have*

$$\textstyle\sum_{n=1}^{N} \frac{X_n}{S_n} \leqslant \ln(eS_N/X_1).$$

Proof. Since $X_1 > 0$, all $S_n > 0$. For $n \geqslant 2$, since $S_n = S_{n-1} + X_n$,

$$\ln S_n - \ln S_{n-1} = \ln\Big(1 + \tfrac{X_n}{S_{n-1}}\Big) \;\geqslant\; \frac{\frac{X_n}{S_{n-1}}}{1 + \frac{X_n}{S_{n-1}}} \;=\; \tfrac{X_n}{S_n},$$

using $\ln(1+u) \geqslant \frac{u}{1+u}$ for $u \geqslant 0$. Summing from $n = 1$ to N gives

$$\textstyle\sum_{n=1}^{N} \frac{X_n}{S_n} \leqslant 1 + \sum_{n=2}^{N}(\ln S_n - \ln S_{n-1}) = 1 + \ln S_N - \ln S_1 = \ln(eS_N/X_1).$$

This concludes the proof. □

Fact B.4 (Square-Root Telescoping Inequality)**.** *Let $(X_n)_{n\geqslant 1}$ be non-negative reals and $X_1 > 0$. Set $S_n = \sum_{m=1}^{n} X_m$. Then, for every $N \geqslant 1$, we have*

$$\textstyle\sum_{n=1}^{N} \frac{X_n}{\sqrt{S_n}} \leqslant 2\sqrt{S_N} - \sqrt{S_1}.$$

Proof. Since $X_1 > 0$, all $S_n > 0$. For $n \geqslant 2$, since $S_n = S_{n-1} + X_n$, it holds that

$$\sqrt{S_n} - \sqrt{S_{n-1}} = \frac{S_n - S_{n-1}}{\sqrt{S_n} + \sqrt{S_{n-1}}} = \frac{X_n}{\sqrt{S_n} + \sqrt{S_{n-1}}} \geqslant \frac{X_n}{2\sqrt{S_n}}.$$

Summing over $n = 1, \ldots, N$, multiplying by 2, and telescoping,

$$\textstyle\sum_{n=1}^{N} \frac{X_n}{\sqrt{S_n}} \leqslant \sqrt{S_1} + 2\sum_{n=2}^{N}\big(\sqrt{S_n} - \sqrt{S_{n-1}}\big) = 2\sqrt{S_N} - \sqrt{S_1}.$$

This concludes the proof. □

C Technical Results for the Continuous Time Model: Proofs

This section establishes the technical results concerning the CTM (Definition 3.1). Recall the observation-ordering convention (Definition 3.5): $\widetilde{\mathrm{a}}_n = \mathrm{a}_{\rho(n)}$, where ρ is the permutation of $[T]$ satisfying $r_{\rho(1)} \leqslant \cdots \leqslant r_{\rho(T)}$. Fact C.1 serves as a central tool throughout this section, connecting delays and backlogs with their duals. Facts C.2 and C.3 are used in the proof of Corollary 5.7, while Fact C.4 plays a key role in the analysis of delay-adaptive learning rates.

Fact C.1. *Define $\beta(t) = |\{s : r_s \leqslant r_t\}|$. For all $t \in [T]$, $\sigma_t^\star = d_t + t - \beta(t)$ and $d_t^\star = \sigma_t + \beta(t) - t$. Furthermore, $\beta = \rho^{-1}$ as permutations of $[T]$, and for all $n \in [T]$, $\widetilde{\sigma}_n^\star = \widetilde{d}_n + \rho(n) - n$ and $\widetilde{d}_n^\star = \widetilde{\sigma}_n + n - \rho(n)$.*

Proof. From the definition of delay and backlog (1) as induced by the CTM, we have

$$t + d_t = |\{s : l_s < r_t\}| = |\{s : r_s \leqslant r_t\}| + |\{s : l_s < r_t < r_s\}| = \beta(t) + \sigma_t^\star,$$
$$t - \sigma_t = |\{s : r_s < l_t\} \cup \{t\}| = |\{s : r_s \leqslant r_t\}| - |\{s : l_t < r_s < r_t\}| = \beta(t) - d_t^\star.$$

Note that $\beta = \rho^{-1}$ as permutations of $[T]$, since $\beta(\rho(n)) = |\{s : r_s \leqslant r_{\rho(n)}\}| = n$ for all $n \in [T]$. The remaining identities follow by substituting $t = \rho(n)$. □

Theorem 3.4 (Restated)**.** *For quantities in (1) and (2), it holds that:*

(a) $\sum_{t=1}^{T} d_t = \sum_{t=1}^{T} \sigma_t = \sum_{t=1}^{T} d_t^\star = \sum_{t=1}^{T} \sigma_t^\star$,
(b) $\sigma_{\max} = \sigma_{\max}^\star, \quad \frac{1}{2} d_{\max} \leqslant d_{\max}^\star \leqslant 2 d_{\max}$,
(c) $d_t^\star = \sigma_t + \beta(t) - t, \quad \sigma_t^\star = d_t + t - \beta(t)$,
where $\beta : [T] \to [T]$ *is the permutation* $\beta(t) = |\{s : r_s \leqslant r_t\}|$.

Proof. We address each claim separately.

(a) To show the classical result $\sum_{t=1}^{T} d_t = \sum_{t=1}^{T} \sigma_t$ simply swap the order of summation:

$$\sum_{t=1}^{T} d_t = \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{I}(l_s \in I_t) = \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{I}(l_s \in I_t) = \sum_{s=1}^{T} \sigma_s.$$

The identities $\sum_{t=1}^{T} \sigma_t^\star = \sum_{t=1}^{T} d_t$ and $\sum_{t=1}^{T} d_t^\star = \sum_{t=1}^{T} \sigma_t$ follow immediately from Fact C.1, because β is a permutation of $[T]$ and $\sum_{t=1}^{T} \beta(t) = \sum_{t=1}^{T} t$.

(b) To show that $\sigma_{\max}^\star \leqslant \sigma_{\max}$, note that $r_t \in (l_{t+d_t}, l_{t+d_t+1})$ by Lemma 3.2, and so

$$\{s : r_t \in I_s\} \subseteq (\{t + d_t\} \cup \{s : l_{t+d_t} \in I_s\}) \backslash \{t\}.$$

Hence, $\sigma_t^\star \leqslant \sigma_{t+d_t} + 1 - 1 = \sigma_{t+d_t}$ for every $t \in [T]$, which means $\sigma_{\max}^\star \leqslant \sigma_{\max}$. To show that $\sigma_{\max}^\star \geqslant \sigma_{\max}$, note that $\widetilde{l}_n \in (\widetilde{r}_{n-\tilde{d}_n^\star - 1}, \widetilde{r}_{n-\tilde{d}_n^\star})$ by Lemma 3.6, and so

$$\{m : \widetilde{l}_n \in \widetilde{I}_m\} \subseteq (\{n - \tilde{d}_n^\star\} \cup \{m : \widetilde{r}_{n-\tilde{d}_n^\star} \in \widetilde{I}_m\}) \backslash \{n\}.$$

Hence, $\widetilde{\sigma}_n \leqslant \widetilde{\sigma}^\star_{n-\tilde{d}_n^\star} + 1 - 1 = \widetilde{\sigma}^\star_{n-\tilde{d}_n^\star}$ for every $n \in [T]$, which means $\sigma_{\max} \leqslant \sigma_{\max}^\star$. To show that $d_{\max}^\star \leqslant 2 d_{\max}$, note that for each $t \in [T]$, $\{s : r_s \in I_t\} \subseteq \{s : l_s \in I_t\} \cup \{s : l_t \in I_s\}$. Hence, $d_{\max}^\star \leqslant \max_t(\sigma_t + d_t) \leqslant \sigma_{\max} + d_{\max} \leqslant 2 d_{\max}$, where the result $\sigma_{\max} \leqslant d_{\max}$ can be inferred from Lemma 3.2. To show that $d_{\max} \leqslant 2 d_{\max}^\star$, note that for each $t \in [T]$, $\{s : l_s \in I_t\} \subseteq \{s : r_s \in I_t\} \cup \{s : r_t \in I_s\}$. Hence, $d_{\max} \leqslant \max_t(d_t^\star + \sigma_t^\star) \leqslant d_{\max}^\star + \sigma_{\max}^\star \leqslant 2 d_{\max}^\star$, where the result $\sigma_{\max}^\star \leqslant d_{\max}^\star$ can be inferred from Lemma 3.6.

(c) These identities are established in Fact C.1. □

Fact C.2. *For every* $N \in [T]$*, it holds that*

$$\sum_{n=1}^{N} \tilde{d}_n^\star \leqslant \sum_{n=1}^{N} \tilde{\sigma}_n^\star.$$

Proof. For every $N \in [T]$, we can write

$$\begin{aligned}
\sum_{n=1}^{N} (\tilde{\sigma}_n^\star - \tilde{d}_n^\star) &= \sum_{n=1}^{N} \sum_{m=1}^{T} (\mathbb{I}(\widetilde{r}_n \in \widetilde{I}_m) - \mathbb{I}(\widetilde{r}_m \in \widetilde{I}_n)) \\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} (\mathbb{I}(\widetilde{r}_n \in \widetilde{I}_m) - \mathbb{I}(\widetilde{r}_m \in \widetilde{I}_n)) \\
&\quad + \sum_{n=1}^{N} \sum_{m=N+1}^{T} (\mathbb{I}(\widetilde{r}_n \in \widetilde{I}_m) - \mathbb{I}(\widetilde{r}_m \in \widetilde{I}_n)) \\
&= 0 + \sum_{n=1}^{N} \sum_{m=N+1}^{T} \mathbb{I}(\widetilde{r}_n \in \widetilde{I}_m),
\end{aligned}$$

where the final equality used symmetry and the fact that $\widetilde{r}_m \notin \widetilde{I}_n$ for $m > N \geqslant n$. □

Fact C.3. *For every* $t \in [T-1]$ *and* $n \in [T-1]$*, it holds that*

$$\sigma_{t+1} \leqslant \sigma_t + 1, \quad \tilde{\sigma}_{n+1}^\star \geqslant \tilde{\sigma}_n^\star - 1.$$

Proof. Using the definition of backlog and dual-backlog, we write

$$\begin{aligned}
\sigma_{t+1} &= |\{s : l_{t+1} \in I_s\}| = |\{s \leqslant t : r_s > l_{t+1}\}| \leqslant |\{s \leqslant t : r_s > l_t\}| = \sigma_t + 1, \\
\tilde{\sigma}_n^\star &= |\{m : \widetilde{r}_n \in \widetilde{I}_m\}| = |\{m > n : \widetilde{l}_m < \widetilde{r}_n\}| \leqslant |\{m > n : \widetilde{l}_m < \widetilde{r}_{n+1}\}| = \tilde{\sigma}_{n+1}^\star + 1.
\end{aligned}$$ □

Fact C.4. *The following two inequalities hold*

$$\sum_{n=1}^{T} \frac{\tilde{\sigma}_n^\star}{n} \leqslant \sigma_{\max} \ln(eT), \quad \sum_{n=1}^{T} \frac{\tilde{d}_n^\star}{n} \leqslant \min\left\{\sigma_{\max} \ln(eT),\, 2\sqrt{d_{\text{tot}}}\right\}.$$

Proof. To begin, observe the following: $\sigma^\star_{\max} = \sigma_{\max}$ by Theorem 3.4, $\sum_{n=1}^T \frac{1}{n} \leqslant \ln(eT)$ by Fact B.3, and $\sum_{n=1}^T \frac{\rho(n)}{n} \geqslant T$ by Fact B.2. We use these observations freely throughout the following proof. To prove the first inequality, simply write

$$\sum_{n=1}^T \frac{\widetilde{\sigma}^\star_n}{n} \leqslant \max_{n\in[T]} \widetilde{\sigma}^\star_n \cdot \sum_{n=1}^T \frac{1}{n} \leqslant \sigma_{\max} \ln(eT).$$

We split the proof of the second inequality into two parts.

- To show that $\sum_{n=1}^T \frac{\widetilde{d}^\star_n}{n} \leqslant \sigma_{\max} \ln(eT)$, use Fact C.1 to write

$$\sum_{n=1}^T \frac{\widetilde{d}^\star_n}{n} = \sum_{n=1}^T \frac{\widetilde{\sigma}_n + n - \rho(n)}{n} \leqslant \sigma_{\max} \ln(eT) + T - \sum_{n=1}^T \frac{\rho(n)}{n} \leqslant \sigma_{\max} \ln(eT).$$

- To show that $\sum_{n=1}^T \frac{\widetilde{d}^\star_n}{n} \leqslant 2\sqrt{d_{\text{tot}}}$, observe that for all $n \in [T]$ it holds that $n \geqslant \widetilde{d}^\star_n + 1$ as the total number of observations up to and including time $\widetilde{r}_n$ is equal to n. Hence, for all $n \in [T]$,

$$n^2 \geqslant \sum_{m=1}^n m \geqslant \sum_{m=1}^n (\widetilde{d}^\star_m + 1) \geqslant 1 + \sum_{m=1}^n \widetilde{d}^\star_m.$$

  By Fact B.4, we have

$$\sum_{n=1}^T \frac{\widetilde{d}^\star_n}{n} \leqslant \left(1 + \sum_{n=1}^T \frac{\widetilde{d}^\star_n}{\sqrt{1+\sum_{m=1}^n \widetilde{d}^\star_m}}\right) - 1 \leqslant 2\sqrt{1 + \sum_{m=1}^T \widetilde{d}^\star_m} - 2 \leqslant 2\sqrt{d_{\text{tot}}}.$$

This concludes the proof of both inequalities. □

D Online Linear Optimization with Drift Penalization: Proofs

This section proves the results for drift-penalized OLO from Section 4. Theorem 4.1 and Lemma 4.2 establish how P-FTRL and OMD control the standard regret $\mathfrak{R}_N(u) = \sum_{n=1}^N \langle c_n, z_n - u\rangle$ and the prediction drift $\mathfrak{D}_N = \sum_{n=1}^N \|z_{n-\lambda_n} - z_n\|$. Given a non-increasing learning rate sequence $\{\eta_n\}_{n=1}^N \subset (0,\infty)$ and initial point $z_1 \in \mathcal{K}$, we consider the update rules

$$\text{P-FTRL}: \quad z_{n+1} = \operatorname{argmin}_{z\in\mathcal{K}} \langle \textstyle\sum_{m=1}^n c_m,\, z\rangle + \sum_{m=1}^n \frac{1/\eta_m - 1/\eta_{m-1}}{2} \|z_m - z\|^2,$$

$$\text{OMD}: \quad z_{n+1} = \operatorname{argmin}_{z\in\mathcal{K}} \langle c_n,\, z\rangle + \frac{1/\eta_n}{2} \|z_n - z\|^2,$$

with the convention $1/\eta_0 = 0$.

D.1 General results for FTRL and OMD

The following results are standard in the analysis of FTRL and OMD.

Theorem D.1 (Orabona [2019], adapted from Lemmas 7.1 and 7.8 for linear losses)**.** *Let $\mathcal{K} \subset \mathbb{R}^k$ be a convex compact set, let $\{\eta_n\}_{n=1}^N \subset (0,\infty)$ be a non-increasing sequence, and let $\psi_1, \dots, \psi_N : \mathbb{R}^k \to \mathbb{R}$ be differentiable functions such that each ψ_n is $(1/\eta_n)$-strongly convex with respect to $\|\cdot\|$ on $\mathcal{K}$. Given loss vectors $\{c_n\}_{n=1}^N \subset \mathbb{R}^k$ and an initial point $z_1 \in \mathcal{K}$, we can iteratively select unique minimizers $z_{n+1} = \operatorname{argmin}_{z\in\mathcal{K}} \{\psi_n(z) + \sum_{m=1}^n \langle c_m, z\rangle\}$. For proximal regularizers ψ_n such that $z_n \in \operatorname{argmin}_{z\in\mathcal{K}} \{\psi_n(z) - \psi_{n-1}(z)\}$, for all $u \in \mathcal{K}$, it holds that*

$$\sum_{n=1}^N \langle c_n, z_n - u\rangle \leqslant \psi_N(u) + \sum_{n=1}^N \left(\tfrac{1}{2}\eta_n \|c_n\|_\star^2 + \psi_{n-1}(z_n) - \psi_n(z_n)\right).$$

Lemma D.2 (Stability lemma, modified version of Lemma A.2 in Qiu et al. [2025])**.** *Let $\mathcal{K} \subset \mathbb{R}^k$ be a convex compact set. For $i \in \{0,1\}$, let function $\phi_i : \mathbb{R}^k \to \mathbb{R}$ be differentiable and λ_i-strongly convex with respect to $\|.\|$ on $\mathcal{K}$. Consider $z_i \in \operatorname{argmin}_{x\in\mathcal{K}} \langle w_i,\, x\rangle + \phi_i(x)$. It holds that*

$$\tfrac{\lambda_0+\lambda_1}{2} \|z_0 - z_1\|^2 \leqslant \langle w_0 - w_1,\, z_1 - z_0\rangle + (\phi_1(z_0) - \phi_0(z_0)) - (\phi_1(z_1) - \phi_0(z_1)).$$

Proof. For $i \in \{0,1\}$, let $h_i(x) = \langle w_i,\, x\rangle + \phi_i(x)$ denote the λ_i-strongly convex function minimized at point z_i. Therefore, by the strong convexity of h_i, it holds that

$$h_i(z_{1-i}) - h_i(z_i) \geqslant \tfrac{\lambda_i}{2} \|z_{1-i} - z_i\|^2.$$

By summing the above inequality for both $i \in \{0,1\}$, we conclude

$$\begin{aligned}
&\langle w_0 - w_1,\, z_1 - z_0\rangle + (\phi_1(z_0) - \phi_0(z_0)) - (\phi_1(z_1) - \phi_0(z_1)) \\
&= h_0(z_1) - h_0(z_0) + h_1(z_0) - h_1(z_1) \\
&\geqslant \tfrac{\lambda_0+\lambda_1}{2} \|z_0 - z_1\|^2.
\end{aligned}$$

□

Lemma D.3 (Orabona [2019], adapted from Lemma 6.10 for linear losses)**.** *Let $\mathcal{K} \subset \mathbb{R}^k$ be a convex compact set. Consider function $\psi : \mathbb{R}^k \to \mathbb{R}$ that is proper, closed, differentiable, 1-strongly convex function with respect to $\|.\|$ in $\mathcal{K}$, with $B_\psi : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ denoting its Bregman divergence, i.e., $B_\psi(x; y) = \psi(x) - \psi(y) - \langle \nabla\psi(y), x - y\rangle$. Given $\{c_n\}_{n=1}^N \subset \mathbb{R}^k$, non-increasing sequence $\{\eta_n\}_{n=1}^N \subset (0, \infty)$, and $z_1 \in \mathcal{K}$, we can iteratively select unique minimizers $z_{n+1} = \operatorname{argmin}_{z\in\mathcal{K}}\langle z, c_n\rangle + \frac{1}{\eta_n} B_\psi(z; z_n)$. Moreover, for all $u \in \mathcal{K}$ and $n \in [N]$:*

$$\eta_n\langle c_n, z_n - u\rangle \leqslant B_\psi(u; z_n) - B_\psi(u; z_{n+1}) + \tfrac{\eta_n^2}{2}\|c_n\|_\star^2.$$

Theorem D.4 (OMD: Regret and Drift)**.** *Under the assumptions of Lemma D.3, for all $u \in \mathcal{K}$, it holds that*

$$\textstyle\sum_{n=1}^N\langle c_n, z_n - u\rangle \leqslant \sum_{n=1}^N(\frac{1}{\eta_n} - \frac{1}{\eta_{n-1}})B_\psi(u; z_n) + \frac{1}{2}\sum_{n=1}^N \eta_n\|c_n\|_\star^2.$$

Moreover, for all $n \in [N-1]$, it holds that $\|z_{n+1} - z_n\| \leqslant \eta_n\|c_n\|_\star$.

Proof. For all $u \in \mathcal{K}$, using Lemma D.3, we write

$$\begin{aligned}\textstyle\sum_{n=1}^N\langle c_n, z_n - u\rangle &\leqslant \textstyle\sum_{n=1}^N \left(\frac{1}{\eta_n}B_\psi(u; z_n) - \frac{1}{\eta_n}B_\psi(u; z_{n+1})\right) + \frac{1}{2}\sum_{n=1}^N \eta_n\|c_n\|_\star^2\\ &= \textstyle\sum_{n=1}^N(\frac{1}{\eta_n} - \frac{1}{\eta_{n-1}})B_\psi(u; z_n) - \frac{1}{\eta_N}B_\psi(u; z_{N+1}) + \frac{1}{2}\sum_{n=1}^N \eta_n\|c_n\|_\star^2,\end{aligned}$$

where $B_\psi(u; z_{N+1}) \geqslant 0$ for Bregman divergence. This concludes the proof of the regret bound. Fix arbitrary $n \in [N]$ and let $\phi_0(z) = \frac{1}{\eta_n}B_\psi(z; z_n)$. Let $g_0 = \nabla\phi_0(z_n)$ and $g_1 = \nabla\phi_0(z_{n+1})$. By the first-order optimality of $z_n = \operatorname{argmin}_{z\in\mathcal{K}}\{\phi_0(z)\}$ and $z_{n+1} = \operatorname{argmin}_{z\in\mathcal{K}}\{\langle z, c_n\rangle + \phi_0(z)\}$, we have

$$\langle g_0,\, z_n - z_{n+1}\rangle \leqslant 0 \quad \text{and} \quad \langle c_n + g_1,\, z_n - z_{n+1}\rangle \geqslant 0.$$

The monotonicity inequality for $\frac{1}{\eta_n}$-strong convex ϕ_0 yields $\langle g_0 - g_1,\, z_n - z_{n+1}\rangle \geqslant \frac{1}{\eta_n}\|z_n - z_{n+1}\|^2$.
Combining these inequalities, we have

$$\begin{aligned}\tfrac{1}{\eta_n}\|z_n - z_{n+1}\|^2 &\leqslant \langle g_0 - g_1,\, z_n - z_{n+1}\rangle\\ &= \langle c_n,\, z_n - z_{n+1}\rangle + \langle g_0,\, z_n - z_{n+1}\rangle - \langle c_n + g_1,\, z_n - z_{n+1}\rangle\\ &\leqslant \|c_n\|_\star\|z_n - z_{n+1}\| + 0 + 0.\end{aligned}$$

Consequently, $\|z_n - z_{n+1}\| \leqslant \eta_n\|c_n\|_\star$. □

D.2 Auxiliary Results for Drift Control of P-FTRL

Lemma D.5. *Let $\{z_n\}_{n=1}^T$ denote predictions of P-FTRL (4). For all $n \in [N]$ and $\lambda \in [N-n]$, it holds that*

$$\|z_{n+\lambda} - z_n\| \leqslant \eta_{n+\lambda}\,\|\textstyle\sum_{m=n}^{n+\lambda-1} c_m\|_\star + (1 - \eta_{n+\lambda}/\eta_n)D.$$

Moreover, it holds that $\|z_{n+1} - z_n\| \leqslant \eta_n\,\|c_n\|_\star$ for $n \in [N-1]$.

Proof. Let $\alpha_1 = 1/\eta_1$ and $\alpha_{n+1} = 1/\eta_{n+1} - 1/\eta_n$ for $n \in [N-1]$. To apply Lemma D.2, consider the following vectors and functions

$$\begin{aligned} w_0 &= \textstyle\sum_{m=1}^{n-1} c_m, \quad \phi_0(z) = \sum_{m=1}^{n} \frac{\alpha_m}{2}\|z - z_m\|^2,\\ w_1 &= \textstyle\sum_{m=1}^{n+\lambda-1} c_m, \quad \phi_1(z) = \sum_{m=1}^{n+\lambda} \frac{\alpha_m}{2}\|z - z_m\|^2,\end{aligned}$$

so that $z_n = \operatorname{argmin}_{z\in\mathcal{K}}\langle z, w_0\rangle + \phi_0(z)$ and $z_{n+\lambda} = \operatorname{argmin}_{z\in\mathcal{K}}\langle z, w_1\rangle + \phi_1(z)$, while ϕ_0 and ϕ_1 are $1/\eta_n$ and $1/\eta_{n+\lambda}$-strongly convex, respectively. Applying Lemma D.2, we have

$$\begin{aligned}&\tfrac{\frac{1}{\eta_{n+\lambda}} + \frac{1}{\eta_n}}{2}\|z_n - z_{n+\lambda}\|^2\\ &\leqslant \langle \textstyle\sum_{m=n}^{n+\lambda-1} c_m,\, z_n - z_{n+\lambda}\rangle + \sum_{m=n+1}^{n+\lambda} \frac{\alpha_m}{2}\left(\|z_n - z_m\|^2 - \|z_{n+\lambda} - z_m\|^2\right)\\ &= \langle \textstyle\sum_{m=n}^{n+\lambda-1} c_m,\, z_n - z_{n+\lambda}\rangle + \langle \sum_{m=n+1}^{n+\lambda} \frac{\alpha_m}{2}(z_n + z_{n+\lambda} - 2z_m),\, z_n - z_{n+\lambda}\rangle\\ &= \langle \textstyle\sum_{m=n}^{n+\lambda-1} c_m,\, z_n - z_{n+\lambda}\rangle + \langle \sum_{m=n+1}^{n+\lambda} \alpha_m(z_n - z_m),\, z_n - z_{n+\lambda}\rangle\\ &- \tfrac{\frac{1}{\eta_{n+\lambda}} - \frac{1}{\eta_n}}{2}\|z_n - z_{n+\lambda}\|^2.\end{aligned}$$

Grouping terms containing $\|z_n - z_{n+\lambda}\|^2$ on the left side and using Assumptions 2.1, we write

$$\begin{aligned}\tfrac{1}{\eta_{n+\lambda}}\|z_n - z_{n+\lambda}\|^2 &\leqslant \langle \textstyle\sum_{m=n}^{n+\lambda-1} c_m,\, z_n - z_{n+\lambda}\rangle + \langle \textstyle\sum_{m=n+1}^{n+\lambda} \alpha_m(z_n - z_m),\, z_n - z_{n+\lambda}\rangle \\ &\leqslant \left(\|\textstyle\sum_{m=n}^{n+\lambda-1} c_m\|_\star + (1/\eta_{n+\lambda} - 1/\eta_n)D\right)\|z_n - z_{n+\lambda}\|.\end{aligned}$$

The first inequality immediately follows. Alternatively, for $\lambda = 1$, we have

$$\begin{aligned}\tfrac{1}{\eta_n}\|z_n - z_{n+1}\|^2 &= (\tfrac{1}{\eta_{n+1}} - \alpha_{n+1})\|z_n - z_{n+1}\|^2 \\ &\leqslant \langle c_n,\, z_n - z_{n+1}\rangle + \langle \alpha_{n+1}(z_n - z_{n+1}),\, z_n - z_{n+1}\rangle - \alpha_{n+1}\|z_n - z_{n+1}\|^2 \\ &= \langle c_n,\, z_n - z_{n+1}\rangle \\ &\leqslant \|c_n\|_\star \|z_n - z_{n+1}\|.\end{aligned}$$

The second result follows from this, concluding the proof. □

Lemma D.6. *For learning rates $\{\eta_n\}_{n=1}^N$ of P-FTRL (4), let $H_\eta = \sum_{n=1}^N (1 - \frac{\eta_n}{\eta_{n-\lambda_n}})$. Then,*

$$H_\eta \leqslant (\max_{n\in[N]} \nu_n + 1)\ln\left(\tfrac{e\eta_1}{\eta_N}\right) \qquad \text{and} \qquad H_\eta \leqslant \tfrac{\max_{n\in[N]}\sum_{m=n}^{n+\nu_n}\eta_m}{\eta_N}.$$

Proof. Let $\alpha_1 = 1/\eta_1 > 0$ and $\alpha_{n+1} = 1/\eta_{n+1} - 1/\eta_n \geqslant 0$ for $n \in [N-1]$. Then, we can write

$$\begin{aligned}H_\eta = \textstyle\sum_{n=1}^N \eta_n\left(\frac{1}{\eta_n} - \frac{1}{\eta_{n-\lambda_n}}\right) &= \textstyle\sum_{n=1}^N \eta_n \sum_{m=1}^N \alpha_m\, \mathbb{I}(n - \lambda_n < m \leqslant n) \\ &= \textstyle\sum_{m=1}^N \alpha_m \sum_{n=1}^N \eta_n\, \mathbb{I}(n - \lambda_n < m \leqslant n).\end{aligned}$$

Since the sequence $\{\eta_n\}_{n=1}^N$ is non-increasing and by Lemma 3.6,

$$\textstyle\sum_{n=1}^N \mathbb{I}(n - \lambda_n < m \leqslant n) \leqslant \sum_{n=1}^N \mathbb{I}(n - \lambda_n \leqslant m < n) + 1 = \nu_m + 1,$$

we can further write, with $\nu_{\max} = \max_{n\in[N]}\nu_n$:

$$\begin{aligned}H_\eta &\leqslant \textstyle\sum_{m=1}^N \alpha_m\eta_m\,(\nu_m + 1) \leqslant (\nu_{\max}+1)\sum_{m=1}^N \alpha_m\eta_m \\ &= (\nu_{\max}+1)\textstyle\sum_{m=1}^N \frac{\alpha_m}{\sum_{k=1}^m \alpha_k} \leqslant (\nu_{\max}+1)\ln\left(\tfrac{e\eta_1}{\eta_N}\right),\end{aligned}$$

where the final inequality follows from Fact B.3 with $X_n = \alpha_n$ and the identity $\eta_m = 1/(\sum_{k=1}^m \alpha_k)$. Using a different approach, we can also write

$$H_\eta \leqslant \textstyle\sum_{m=1}^N \alpha_m \sum_{n=m}^{m+\nu_m} \eta_n \leqslant \left(\sum_{m=1}^N \alpha_m\right)\max_{n\in[N]}\sum_{m=n}^{n+\nu_n}\eta_m = \frac{\max_{n\in[N]}\sum_{m=n}^{n+\nu_n}\eta_m}{\eta_N}.$$

This concludes the proof. □

D.3 Proofs of Theorem 4.1 and Lemma 4.2

Theorem 4.1 (Restated)**.** *Predictions $\{z_n\}_{n=1}^N$ generated by either P-FTRL (4) or OMD (5) satisfy*

$$\mathfrak{R}_N(u) \leqslant \textstyle\sum_{n=1}^N \frac{1/\eta_n - 1/\eta_{n-1}}{2}\|z_n - u\|^2 + \frac{1}{2}\sum_{n=1}^N \eta_n\|c_n\|_\star^2 \quad \text{and} \quad \mathfrak{D}_N \leqslant \textstyle\sum_{n=1}^N \eta_n\nu_n\|c_n\|_\star.$$

Proof. The regret bound for P-FTRL follows from Theorem D.1 with regularizer functions $\psi_n(z) = \sum_{m=1}^n \frac{1/\eta_m - 1/\eta_{m-1}}{2}\|z_m - z\|^2$. The regret bound for OMD follows from Theorem D.4 with $\psi(z) = \frac{1}{2}\|z\|^2$ so that $B_\psi(x;z) = \frac{1}{2}\|x - z\|^2$. For both algorithms, it holds that $\|z_{n+1} - z_n\| \leqslant \eta_n\|c_n\|_\star$ for all $n \in [N-1]$, as shown in Theorem D.1 and Lemma D.5. Using triangle inequality and the fact that $\nu_m = |\{n : n - \lambda_n \leqslant m < n\}|$, we can write

$$\begin{aligned}\mathfrak{D}_N &= \textstyle\sum_{n=1}^N \|z_n - z_{n-\lambda_n}\| \\ &\leqslant \textstyle\sum_{n=1}^N \sum_{m=1}^N \mathbb{I}(n - \lambda_n \leqslant m < n)\|z_{m+1} - z_m\| \\ &= \textstyle\sum_{m=1}^N (\sum_{n=1}^N \mathbb{I}(n - \lambda_n \leqslant m < n))\|z_{m+1} - z_m\| \\ &= \textstyle\sum_{m=1}^N \nu_m\|z_{m+1} - z_m\| \\ &\leqslant \textstyle\sum_{m=1}^N \eta_m\nu_m\|c_m\|_\star,\end{aligned}$$

where the final inequality plugs-in $\|z_{m+1} - z_m\| \leqslant \eta_m\|c_m\|_\star$ for single-step drifts. □

Lemma 4.2 (Restated). *Predictions $\{z_n\}_{n=1}^N$ generated by P-FTRL (4) satisfy*

$$\mathfrak{D}_N \leqslant \sum_{n=1}^N \eta_n \| \sum_{m=n-\lambda_n}^{n-1} c_m \|_\star + D\, H_\eta \quad \textit{where} \quad H_\eta = \sum_{n=1}^N (1 - \eta_n/\eta_{n-\lambda_n}).$$

Proof. From Lemma D.5, we write

$$\begin{aligned}
\mathfrak{D}_N &= \sum_{n=1}^N \|z_n - z_{n-\lambda_n}\| \\
&\leqslant \sum_{n=1}^N \left(\eta_n \| \sum_{m=n-\lambda_n}^{n-1} c_m \|_\star + (1 - \eta_n/\eta_{n-\lambda_n}) D \right) \\
&= \sum_{n=1}^N \eta_n \| \sum_{m=n-\lambda_n}^{n-1} c_m \|_\star + D H_\eta.
\end{aligned}$$

□

E Guarantees for OCO and BCO with Delays: Proofs

E.1 Proof of Theorem 5.2

Theorem 5.2 (Restated). *For any base algorithm $\mathcal{B}$ and comparator $u \in \mathcal{K}$, $\mathcal{W}_{\text{OCO}}(\mathcal{B})$ guarantees*

$$R_T(u) \leqslant \mathfrak{R}_T^{\text{drift}}(u; G),$$

where $\mathfrak{R}_T^{\text{drift}}(u; W)$ is the drift-penalized regret of $\mathcal{B}$ with loss vectors $c_n = \widetilde{g}_n$ and lags $\lambda_n = \widetilde{d}_n^\star$. Under λ-strong convexity (2.3), it further holds that

$$R_T(u) \leqslant \mathfrak{R}_T^{\text{drift}}(u; 3G) - \tfrac{\lambda}{2} \sum_{n=1}^T \|z_n - u\|^2.$$

Proof. The drift-penalized regret of $\mathcal{B}$ for these loss vectors and lags is defined as

$$\mathfrak{R}_T^{\text{drift}}(u; W) = \sum_{n=1}^T \langle \widetilde{g}_n,\, z_n - u \rangle + W \sum_{n=1}^T \|z_{n-\widetilde{d}_n^\star} - z_n\|.$$

We treat the general convex case as 0-strongly convex. For $\lambda \geqslant 0$, the λ-strong convexity of the loss functions f_t gives

$$\begin{aligned}
&\sum_{t=1}^T (f_t(x_t) - f_t(u)) \\
&\leqslant \sum_{t=1}^T \langle g_t,\, x_t - u \rangle - \tfrac{\lambda}{2} \sum_{t=1}^T \|x_t - u\|^2 \\
&\overset{\text{(a)}}{=} \sum_{n=1}^T \langle \widetilde{g}_n,\, \widetilde{x}_n - u \rangle - \tfrac{\lambda}{2} \sum_{n=1}^T \|\widetilde{x}_n - u\|^2 \\
&\overset{\text{(b)}}{=} \sum_{n=1}^T \langle \widetilde{g}_n,\, z_{n-\widetilde{d}_n^\star} - u \rangle - \tfrac{\lambda}{2} \sum_{n=1}^T \|z_{n-\widetilde{d}_n^\star} - u\|^2 \\
&\leqslant \sum_{n=1}^T \langle \widetilde{g}_n,\, z_n - u \rangle + \sum_{n=1}^T \|\widetilde{g}_n\|_\star \|z_{n-\widetilde{d}_n^\star} - z_n\| - \tfrac{\lambda}{2} \sum_{n=1}^T \|z_{n-\widetilde{d}_n^\star} - u\|^2 \\
&\overset{\text{(c)}}{\leqslant} \sum_{n=1}^T \langle \widetilde{g}_n,\, z_n - u \rangle + G \sum_{n=1}^T \|z_{n-\widetilde{d}_n^\star} - z_n\| - \tfrac{\lambda}{2} \sum_{n=1}^T \|z_{n-\widetilde{d}_n^\star} - u\|^2,
\end{aligned}$$

where (a) replaces sequences $\{x_t\}_{t=1}^T, \{g_t\}_{t=1}^T$ with their observation-orderings (Definition 3.5), (b) substitutes $\widetilde{x}_n = z_{n-\widetilde{d}_n^\star}$ which follows from Lemma 3.6, and (c) applies the gradient norm bound (Assumption 2.2). For the general convex case ($\lambda = 0$), this immediately gives $R_T(u) \leqslant \mathfrak{R}_T^{\text{drift}}(u; G)$. For the strongly convex case ($\lambda > 0$), we convert the distance terms from $z_{n-\widetilde{d}_n^\star}$ to z_n. Since $\|z_{n-\widetilde{d}_n^\star} - u\|, \|z_n - u\| \leqslant \frac{2G}{\lambda}$ by Fact B.1, we have

$$\tfrac{\lambda}{2}(\|z_n - u\|^2 - \|z_{n-\widetilde{d}_n^\star} - u\|^2) = \tfrac{\lambda}{2} \langle z_{n-\widetilde{d}_n^\star} + z_n - 2u,\, z_n - z_{n-\widetilde{d}_n^\star} \rangle \leqslant 2G \|z_{n-\widetilde{d}_n^\star} - z_n\|.$$

Summing over n and rearranging yields $R_T(u) \leqslant \mathfrak{R}_T^{\text{drift}}(u; 3G) - \frac{\lambda}{2} \sum_{n=1}^T \|z_n - u\|^2$. □

E.2 Proofs of Corollaries 5.3 and 5.4

Theorem E.1. *$\mathcal{W}_{\text{OCO}}(\mathcal{B})$, where $\mathcal{B}$ executes P-FTRL (4) updates, guarantees that in the*

$$\begin{aligned}
\textit{general convex case}\colon\ R_T(u) &\leqslant \tfrac{D^2}{2\eta_T} + G^2 \sum_{n=1}^T \eta_n (\widetilde{\sigma}_n^\star + 1), \\
\textit{strongly convex case (2.3)}\colon\ R_T(u) &\leqslant \sum_{n=1}^T \tfrac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2} \|z_n - u\|^2 \\
&\quad + 3G^2 \sum_{n=1}^T \eta_n (\widetilde{d}_n^\star + 1) + 3GDH_\eta,
\end{aligned}$$

where $H_\eta = \sum_{n=1}^T (1 - \eta_n/\eta_{n-\widetilde{d}_n^\star})$ denotes the total learning rate misalignment.

Proof. Both results follow from Theorem 5.2 once we bound the standard regret $\mathfrak{R}_T(u) = \sum_{n=1}^T \langle \widetilde{g}_n,\, z_n - u\rangle$ and prediction drift $\mathfrak{D}_T = \sum_{n=1}^T \|z_{n-\widetilde{d}_n^\star} - z_n\|$ components. To control $\mathfrak{R}_T(u)$ in both cases, apply Theorem 4.1 for $\|\widetilde{g}_n\|_\star \leqslant G$ and $\|z_n - u\| \leqslant D$, as follows

$$\mathfrak{R}_T(u) \leqslant \sum_{n=1}^T \tfrac{1/\eta_n - 1/\eta_{n-1}}{2} \|z_n - u\|^2 + \tfrac{1}{2}\sum_{n=1}^T \eta_n \|g_n\|_\star^2 \leqslant \tfrac{D^2}{2\eta_T} + G^2 \sum_{n=1}^T \eta_n.$$

To control $\mathfrak{D}_T$ in the general convex case, use Theorem 4.1 to write

$$\mathfrak{D}_T \leqslant \sum_{n=1}^T \eta_n \widetilde{\sigma}_n^\star \|\widetilde{g}_n\|_\star \leqslant G \sum_{n=1}^T \eta_n \widetilde{\sigma}_n^\star.$$

Then, the general convex case follows from the corresponding inequality in Theorem 5.2:

$$\begin{aligned} R_T(u) &\leqslant \mathfrak{R}_T(u) + G\,\mathfrak{D}_T \leqslant \tfrac{D^2}{2\eta_T} + G^2 \textstyle\sum_{n=1}^T \eta_n + G^2 \sum_{n=1}^T \eta_n\, \widetilde{\sigma}_n^\star \\ &\leqslant \tfrac{D^2}{2\eta_T} + G^2 \textstyle\sum_{n=1}^T \eta_n\, (\widetilde{\sigma}_n^\star + 1). \end{aligned}$$

To control $\mathfrak{D}_T$ in the strongly convex case, use Lemma 4.2 to write

$$\mathfrak{D}_T \leqslant \textstyle\sum_{n=1}^T \eta_n \| \sum_{m=n-\widetilde{d}_n^\star}^{n-1} \widetilde{g}_m \|_\star + D\, H_\eta \leqslant G \sum_{n=1}^T \eta_n \widetilde{d}_n^\star + D\, H_\eta,$$

where $H_\eta = \sum_{n=1}^T (1 - \eta_n/\eta_{n-\widetilde{d}_n^\star})$ in this case with lags $\lambda_n = \widetilde{d}_n^\star$. Finally, the strongly convex case follows from the corresponding inequality in Theorem 5.2:

$$\begin{aligned} R_T(u) &\leqslant \mathfrak{R}_T(u) - \tfrac{\lambda}{2}\textstyle\sum_{n=1}^T \|z_n - u\|^2 + 3G\,\mathfrak{D}_T \\ &\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2} \|z_n - u\|^2 + G^2 \sum_{n=1}^T \eta_n + 3G^2 \sum_{n=1}^T \eta_n \widetilde{d}_n^\star + 3GDH_\eta \\ &\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2} \|z_n - u\|^2 + 3G^2 \sum_{n=1}^T \eta_n (\widetilde{d}_n^\star + 1) + 3GDH_\eta. \end{aligned}$$

This concludes the proof for both cases. □

Theorem E.2. *$\mathcal{W}_{\text{OCO}}(\mathcal{B})$, where $\mathcal{B}$ executes OMD (5) updates, guarantees that in the*

$$\begin{aligned} &\textit{general convex case}\colon && R_T(u) \leqslant \tfrac{D^2}{2\eta_T} + G^2 \textstyle\sum_{n=1}^T \eta_n\, (\widetilde{\sigma}_n^\star + 1), \\ &\textit{strongly convex case (2.3)}\colon && R_T(u) \leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2} \|z_n - u\|^2 + 3G^2 \sum_{n=1}^T \eta_n (\widetilde{\sigma}_n^\star + 1). \end{aligned}$$

Proof. As in the proof of Theorem E.1, we use Theorem 4.1 to show that

$$\mathfrak{R}_T(u) \leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1}}{2} \|z_n - u\|^2 + G^2 \sum_{n=1}^T \eta_n \quad \text{and} \quad \mathfrak{D}_T \leqslant G \textstyle\sum_{n=1}^T \eta_n\, \widetilde{\sigma}_n^\star.$$

The general convex case similarly follows from its corresponding inequality in Theorem 5.2:

$$\begin{aligned} R_T(u) &\leqslant \mathfrak{R}_T(u) + G\,\mathfrak{D}_T \leqslant \tfrac{D^2}{2\eta_T} + G^2 \textstyle\sum_{n=1}^T \eta_n + G^2 \sum_{n=1}^T \eta_n\, \widetilde{\sigma}_n^\star \\ &\leqslant \tfrac{D^2}{2\eta_T} + G^2 \textstyle\sum_{n=1}^T \eta_n\, (\widetilde{\sigma}_n^\star + 1). \end{aligned}$$

The strongly convex case also follows from its corresponding inequality in Theorem 5.2:

$$\begin{aligned} R_T(u) &\leqslant \mathfrak{R}_T(u) - \tfrac{\lambda}{2}\textstyle\sum_{n=1}^T \|z_n - u\|^2 + 3G\,\mathfrak{D}_T \\ &\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2} \|z_n - u\|^2 + G^2 \sum_{n=1}^T \eta_n + 3G^2 \sum_{n=1}^T \eta_n \widetilde{\sigma}_n^\star \\ &\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2} \|z_n - u\|^2 + 3G^2 \sum_{n=1}^T \eta_n (\widetilde{\sigma}_n^\star + 1). \end{aligned}$$

This concludes the proof for both cases. □

Corollary 5.3 (Restated)**.** *$\mathcal{W}_{\text{OCO}}(\mathcal{B})$, where $\mathcal{B}$ runs P-FTRL (4) or OMD (5) with $\eta_n = \frac{D/G}{\sqrt{n + \sum_{m=1}^n \widetilde{\sigma}_m^\star}}$, guarantees*

$$R_T(x^*) = O\left(GD\left[\sqrt{d_{\text{tot}}} + \sqrt{T}\right]\right).$$

Proof. Using the general convex case of Theorem E.1 for P-FTRL and Theorem E.2 for OMD, we can write

$$\begin{aligned} R_T(x^*) &\leqslant \tfrac{D^2}{2\eta_T} + G^2 \textstyle\sum_{n=1}^T \eta_n(\widetilde{\sigma}_n^\star + 1) \\ &\overset{\text{(a)}}{\leqslant} GD\sqrt{T + \textstyle\sum_{n=1}^T \widetilde{\sigma}_n^\star} + 2GD\sqrt{\textstyle\sum_{n=1}^T(\widetilde{\sigma}_n^\star + 1)} \\ &\overset{\text{(b)}}{=} 3GD\sqrt{T + d_{\text{tot}}} \\ &\leqslant 6GD\left[\sqrt{d_{\text{tot}}} + \sqrt{T}\right], \end{aligned}$$

where (a) substitutes learning rate values and applies Fact B.4 for $X_n = \widetilde{\sigma}_n^\star + 1$ and (b) uses the identity $\sum_{t=1}^T \sigma_t^\star = d_{\text{tot}}$ from Theorem 3.4. □

Corollary 5.4 (Restated)**.** *Under λ-strong convexity (2.3), $\mathcal{W}_{\text{OCO}}(\mathcal{B})$, where $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{1}{n\lambda}$, guarantees*

$$R_T(x^*) = O\left(\tfrac{G^2}{\lambda}\left[\min\left\{\sigma_{\max}\ln T,\, \sqrt{d_{\text{tot}}}\right\} + \ln T\right]\right).$$

Running OMD (5) with $\eta_n = \frac{1}{n\lambda}$ instead guarantees

$$R_T(x^*) = O\left(\tfrac{G^2}{\lambda}(\sigma_{\max}+1)\ln T\right).$$

Proof. We prove the result for P-FTRL first. For learning rates $\eta_n = \frac{1}{n\lambda}$, by letting $D \leqslant \frac{2G}{\lambda}$ (see Fact B.1), term $H_\eta = \sum_{n=1}^T \eta_n(1/\eta_n - 1/\eta_{n-\widetilde{d}_n^\star})$ can be bounded from above as

$$H_\eta = \textstyle\sum_{n=1}^T \eta_n(\lambda n - \lambda(n - \widetilde{d}_n^\star)) = \lambda \sum_{n=1}^T \eta_n \widetilde{d}_n^\star \leqslant (2G/D)\sum_{n=1}^T \eta_n \widetilde{d}_n^\star.$$

Then, using the strongly convex case of Theorem E.1, we write

$$\begin{aligned} R_T(x^*) &\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2}\|z_n - x^*\|^2 + 3G^2\sum_{n=1}^T \eta_n(\widetilde{d}_n^\star + 1) + 3GDH_\eta \\ &\overset{\text{(a)}}{\leqslant} 0 + 3G^2\textstyle\sum_{n=1}^T \eta_n(\widetilde{d}_n^\star + 1) + 6G^2\sum_{n=1}^T \eta_n \widetilde{d}_n^\star \\ &\overset{\text{(b)}}{\leqslant} 9G^2\textstyle\sum_{n=1}^T \frac{\widetilde{d}_n^\star + 1}{\lambda n} \\ &\overset{\text{(c)}}{\leqslant} \tfrac{9G^2}{\lambda}\left(\min\{\sigma_{\max}\ln(eT),\, 2\sqrt{d_{\text{tot}}}\} + \ln(eT)\right), \end{aligned}$$

where (a) substitutes learning rate values in the first term and applies the above bound on H_η, (b) substitutes $\eta_n = 1/(\lambda n)$, and (c) uses Fact C.4 for $\widetilde{d}_n^\star$ and Fact B.3 to get the desired bound. For OMD, using the strongly convex case of Theorem E.2, we write

$$\begin{aligned} R_T(x^*) &\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2}\|z_n - x^*\|^2 + 3G^2\sum_{n=1}^T \eta_n(\widetilde{\sigma}_n^\star + 1) \\ &\overset{\text{(a)}}{\leqslant} 0 + 3G^2\textstyle\sum_{n=1}^T \frac{\widetilde{\sigma}_n^\star + 1}{\lambda n} \\ &\overset{\text{(b)}}{\leqslant} \tfrac{3G^2}{\lambda}\left(\sigma_{\max}\ln(eT) + \ln(eT)\right), \end{aligned}$$

where (a) substitutes learning rate values and (b) uses Fact C.4 for $\widetilde{\sigma}_n^\star$ to get the final bound. □

E.3 Proof of Theorem 5.6

In this section, we prove Theorem 5.6. We begin by recalling notation from the main body.

Algorithm $\mathcal{B}$ generates base-predictions $\{z_n\}_{n=1}^T$, and in round t the wrapper $\mathcal{W}_{BCO}(\mathcal{B})$ produces prediction $x_t = (1-\delta_t/r)\, y_t + \delta_t u_t$, where $y_t = \bar{z}$ is the current base-prediction and $u_t \sim \text{Unif}(\mathbb{S}^{k-1})$ is sampled independently. Define the shrunk suggestions $y_t^\delta = (1 - \delta_t/r)\, y_t \in (1 - \delta_t/r)\mathcal{K}$, so that $x_t = y_t^\delta + \delta_t u_t$.

Under observation-ordering (Definition 3.5), these quantities become

$$\widetilde{y}_n = z_{n-\widetilde{d}_n^\star}, \quad \widetilde{y}_n^\delta = (1 - \tfrac{\widetilde{\delta}_n}{r})\, z_{n-\widetilde{d}_n^\star}, \quad \widetilde{x}_n = \widetilde{y}_n^\delta + \widetilde{\delta}_n \widetilde{u}_n.$$

For the smoothed losses $\widetilde{f}_n^\delta(x) = \mathbb{E}_{v\sim\text{Unif}(\mathbb{B}^k)}[\widetilde{f}_n(x + \widetilde{\delta}_n v)]$ defined on $(1 - \frac{\widetilde{\delta}_n}{r})\mathcal{K}$, let $\widetilde{g}_n^\delta = \nabla \widetilde{f}_n^\delta(\widetilde{y}_n^\delta)$. By Theorem 5.5, the single-point estimator $\widetilde{\widetilde{g}}_n^\delta = \frac{k}{\widetilde{\delta}_n}\widetilde{f}_n(\widetilde{x}_n)\widetilde{u}_n$ satisfies $\mathbb{E}[\widetilde{\widetilde{g}}_n^\delta \mid z_{n-\widetilde{d}_n^\star}] = \widetilde{g}_n^\delta$.

Theorem 5.6 (Restated). *For any base-algorithm $\mathcal{B}$, non-increasing sequence $(\delta_t)_{t\geqslant 1} \subset (0, r]$, and comparator $u \in \mathcal{K}$, $\mathcal{W}_{\text{BCO}}(\mathcal{B})$ guarantees*

$$\bar{R}_T(u) \leqslant \mathbb{E}\left[\mathfrak{R}_T^{\text{drift}}(u; G)\right] + \tfrac{6GR\delta_{\text{tot}}}{r},$$

where $\mathfrak{R}_T^{\text{drift}}(u; W)$ is the drift-penalized regret of $\mathcal{B}$ on loss vectors $c_n = \widetilde{\widetilde{g}}_n^{\delta}$ with lags $\lambda_n = \widetilde{d}_n^{\star}$, and $\delta_{\text{tot}} = \sum_{t=1}^{T} \delta_t$.

Under λ-strong convexity (2.3), it further holds that

$$\bar{R}_T(u) \leqslant \mathbb{E}\Big[\mathfrak{R}_T^{\text{drift}}(u; 3G) - \tfrac{\lambda}{2}\textstyle\sum_{n=1}^{T} \|z_n - u\|^2\Big] + \tfrac{10GR\delta_{\text{tot}}}{r}.$$

Proof. Fix comparator $v \in \mathcal{K}$. For $n \in [T]$, let $\widetilde{v}_n^{\delta} = (1 - \widetilde{\delta}_n/r)v$ which satisfies $\|\widetilde{v}_n^{\delta} - v\| = \frac{\widetilde{\delta}_n \|v\|}{r} \leqslant \frac{R\widetilde{\delta}_n}{r}$. This gives

$$\begin{aligned}
\bar{R}_T(v) &= \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\widetilde{f}_n(\widetilde{x}_n) - \widetilde{f}_n(v))\right] \\
&= \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\widetilde{f}_n^{\delta}(\widetilde{y}_n^{\delta}) - \widetilde{f}_n^{\delta}(\widetilde{v}_n^{\delta}))\right] \\
&+ \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\widetilde{f}_n(\widetilde{x}_n) - \widetilde{f}_n(\widetilde{y}_n^{\delta}))\right] + \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\widetilde{f}_n(\widetilde{y}_n^{\delta}) - \widetilde{f}_n^{\delta}(\widetilde{y}_n^{\delta}))\right] \\
&+ \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\widetilde{f}_n^{\delta}(\widetilde{v}_n^{\delta}) - \widetilde{f}_n(\widetilde{v}_n^{\delta}))\right] + \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\widetilde{f}_n(\widetilde{v}_n^{\delta}) - \widetilde{f}_n(v))\right] \\
&\overset{(a)}{\leqslant} \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\widetilde{f}_n^{\delta}(\widetilde{y}_n^{\delta}) - \widetilde{f}_n^{\delta}(\widetilde{v}_n^{\delta}))\right] + \textstyle\sum_{n=1}^{T}\left(G\widetilde{\delta}_n + G\widetilde{\delta}_n + G\widetilde{\delta}_n + \tfrac{GR}{r}\widetilde{\delta}_n\right) \\
&\overset{(b)}{\leqslant} \mathbb{E}\left[\textstyle\sum_{n=1}^{T}\left(\langle \widetilde{g}_n^{\delta}, \widetilde{y}_n^{\delta} - \widetilde{v}_n^{\delta}\rangle - \tfrac{\lambda}{2}\|\widetilde{y}_n^{\delta} - \widetilde{v}_n^{\delta}\|^2\right)\right] + \tfrac{4GR\delta_{\text{tot}}}{r} \\
&\overset{(c)}{=} \mathbb{E}\left[\textstyle\sum_{n=1}^{T}\left((1 - \tfrac{\widetilde{\delta}_n}{r})\langle \widetilde{g}_n^{\delta}, z_{n-\widetilde{d}_n^{\star}} - v\rangle - \tfrac{\lambda}{2}(1 - \tfrac{\widetilde{\delta}_n}{r})^2 \|z_{n-\widetilde{d}_n^{\star}} - v\|^2\right)\right] + \tfrac{4GR\delta_{\text{tot}}}{r} \\
&= \mathbb{E}\left[\textstyle\sum_{n=1}^{T}\left(\langle \widetilde{g}_n^{\delta}, z_n - v\rangle - \tfrac{\lambda}{2}\|z_{n-\widetilde{d}_n^{\star}} - v\|^2 + \langle \widetilde{g}_n^{\delta}, z_{n-\widetilde{d}_n^{\star}} - z_n\rangle\right)\right] + \tfrac{4GR\delta_{\text{tot}}}{r} \\
&+ \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\tfrac{\widetilde{\delta}_n}{r}\langle \widetilde{g}_n^{\delta}, v - z_{n-\widetilde{d}_n^{\star}}\rangle + (\tfrac{\lambda\widetilde{\delta}_n}{r} - \tfrac{\lambda\widetilde{\delta}_n^2}{2r^2})\|z_{n-\widetilde{d}_n^{\star}} - v\|^2)\right].
\end{aligned} \tag{6}$$

where (a) uses G-Lipschitzness of $\widetilde{f}_n$ and Theorem 5.5 to bound the last four terms in the previous line, (b) uses λ-strong convexity of $\widetilde{f}_n^{\delta}$, and (c) substitutes $\widetilde{y}_n^{\delta} = (1 - \frac{\widetilde{\delta}_n}{r})z_{n-\widetilde{d}_n^{\star}}$ and $\widetilde{v}_n^{\delta} = (1 - \frac{\widetilde{\delta}_n}{r})v$. The final term in (6) can be bounded using Cauchy-Schwarz inequality as follows:

$$\begin{aligned}
&\mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\tfrac{\widetilde{\delta}_n}{r}\langle \widetilde{g}_n^{\delta}, v - z_{n-\widetilde{d}_n^{\star}}\rangle + (\tfrac{\lambda\widetilde{\delta}_n}{r} - \tfrac{\lambda\widetilde{\delta}_n^2}{2r^2})\|z_{n-\widetilde{d}_n^{\star}} - v\|^2)\right] \\
&\qquad \leqslant \mathbb{E}\left[\textstyle\sum_{n=1}^{T}(\tfrac{\widetilde{\delta}_n}{r}\|\widetilde{g}_n^{\delta}\|_{\star}\|z_{n-\widetilde{d}_n^{\star}} - v\| + \tfrac{\lambda\widetilde{\delta}_n}{r}\|z_{n-\widetilde{d}_n^{\star}} - v\|^2)\right] \\
&\qquad \overset{(d)}{\leqslant} \textstyle\sum_{n=1}^{T}\left(\tfrac{GD\widetilde{\delta}_n}{r} + \tfrac{\lambda D^2\widetilde{\delta}_n}{r}\right) = \tfrac{(G+\lambda D)D\delta_{\text{tot}}}{r},
\end{aligned} \tag{7}$$

where (d) follows from the facts that $\|z_{n-\widetilde{d}_n^{\star}} - v\| \leqslant D$ and $\|\widetilde{g}_n^{\delta}\|_{\star} \leqslant G$ due to Theorem 5.5. For every $n \in [T]$, the direction $\widetilde{u}_n$ and base-prediction z_n are independent: $\widetilde{u}_n$ is sampled when making the n-th prediction and only affects the n-th observation $\widetilde{f}_n(\widetilde{x}_n) = \widetilde{f}_n((1 - \widetilde{\delta}_n/r)\, z_{n-\widetilde{d}_n^{\star}} + \widetilde{\delta}_n\widetilde{u}_n)$, whereas z_n is the output of $\mathcal{B}$ after processing the $(n-1)$-th observation. By Theorem 5.5,

$$\mathbb{E}[\widetilde{\widetilde{g}}_n^{\delta} \mid z_n] = \mathbb{E}\left[\tfrac{k}{\widetilde{\delta}_n}\widetilde{f}_n(\widetilde{x}_n)\widetilde{u}_n \mid z_n\right] = \nabla\widetilde{f}_n^{\delta}\left((1 - \tfrac{\widetilde{\delta}_n}{r})\, z_{n-\widetilde{d}_n^{\star}}\right) = \widetilde{g}_n^{\delta}.$$

Hence, the expected regret term $\mathbb{E}[\mathfrak{R}(u)]$ satisfies

$$\mathbb{E}\left[\mathfrak{R}(u)\right] = \mathbb{E}\left[\textstyle\sum_{n=1}^{T}\langle \widetilde{\widetilde{g}}_n^{\delta}, z_n - u\rangle\right] \overset{(e)}{=} \mathbb{E}\left[\textstyle\sum_{n=1}^{T}\langle \widetilde{g}_n^{\delta}, z_n - u\rangle\right]. \tag{8}$$

Combining (6), (7), (8) and the facts that $D \leqslant 2R$ and $\langle \widetilde{g}_n^{\delta}, z_{n-\widetilde{d}_n^{\star}} - z_n\rangle \leqslant G\|z_{n-\widetilde{d}_n^{\star}} - z_n\|$, we have

$$\bar{R}_T(v) \leqslant \mathbb{E}\left[\mathfrak{R}_T(v) - \tfrac{\lambda}{2}\textstyle\sum_{n=1}^{T}\|z_{n-\widetilde{d}_n^{\star}} - v\|^2 + G\mathfrak{D}_T\right] + \tfrac{(6G+2\lambda D)R\delta_{\text{tot}}}{r}. \tag{9}$$

This concludes the proof for the convex case ($\lambda = 0$). It remains to consider the strongly convex case ($\lambda > 0$). For all $n \in [T]$,

$$\tfrac{\lambda}{2}\left(\|z_n - u\|^2 - \|z_{n-\tilde{d}_n^\star} - u\|^2\right) = \tfrac{\lambda}{2}\langle z_{n-\tilde{d}_n^\star} + z_n - 2u,\, z_n - z_{n-\tilde{d}_n^\star}\rangle \leqslant \lambda D\,\|z_{n-\tilde{d}_n^\star} - z_n\|.$$

Combining with (9) and letting $D \leqslant \frac{2G}{\lambda}$ (see Fact B.1) yields

$$\bar{R}_T(u) \leqslant \mathbb{E}\left[\mathfrak{R}_T(u) - \tfrac{\lambda}{2}\textstyle\sum_{n=1}^T \|z_n - u\|^2 + 3G\,\mathfrak{D}_T\right] + \frac{10GR\delta_{\text{tot}}}{r}.$$

This completes the proof. □

E.4 Proofs of Corollaries 5.7 and 5.8

Theorem E.3. *$\mathcal{W}_{\text{BCO}}(\mathcal{B})$, where $\mathcal{B}$ executes P-FTRL (4) updates, guarantees that in the*

$$\begin{aligned}
\textit{general convex case}\colon\ \bar{R}_T(u) &\leqslant \tfrac{D^2}{2\eta_T} + \textstyle\sum_{n=1}^T \eta_n\left(G\,\mathbb{E}[\|\Gamma_n\|] + (kM/\widetilde{\delta}_n)^2\right)\\
&\quad + GDH_\eta + \tfrac{6GR\delta_{\text{tot}}}{r},\\
\textit{strongly convex case (2.3)}\colon\ \bar{R}_T(u) &\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2}\,\mathbb{E}[\|z_n - u\|^2]\\
&\quad + \textstyle\sum_{n=1}^T \eta_n\left(3G\,\mathbb{E}[\|\Gamma_n\|] + (kM/\widetilde{\delta}_n)^2\right)\\
&\quad + 3GDH_\eta + \tfrac{10GR\delta_{\text{tot}}}{r},
\end{aligned}$$

where $\Gamma_n = \sum_{m=n-\tilde{d}_n^\star}^{n-1} \widetilde{\widetilde{g}}_m^\delta$ and $H_\eta = \sum_{n=1}^T (1 - \eta_n/\eta_{n-\tilde{d}_n^\star})$.

Proof. Both results follow from Theorem 5.6 once we bound $\mathfrak{R}_T(u)$ and $\mathfrak{D}_T$ components of $\mathfrak{R}_T^{\text{drift}}(u; W)$. To control $\mathfrak{R}_T(u)$ in both cases, apply Theorem 4.1 for $\|\widetilde{\widetilde{g}}_n^\delta\| \leqslant kM/\widetilde{\delta}_n$ and $\|z_n - u\| \leqslant D$, as follows

$$\mathfrak{R}_T(u) \leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1}}{2}\,\|z_n - u\|^2 + \frac{1}{2}\sum_{n=1}^T \eta_n \|\widetilde{\widetilde{g}}_n^\delta\|^2 \leqslant \frac{D^2}{2\eta_T} + \textstyle\sum_{n=1}^T \eta_n (kM/\widetilde{\delta}_n)^2.$$

To control $\mathfrak{D}_T = \sum_{n=1}^T \|z_n - z_{n-\tilde{d}_n^\star}\|$, we use Lemma 4.2 to write

$$\mathfrak{D}_T \leqslant \textstyle\sum_{n=1}^T \eta_n\left(\|\sum_{m=n-\tilde{d}_n^\star}^{n-1} \widetilde{\widetilde{g}}_m^\delta\|_\star + (1/\eta_n - 1/\eta_{n-\tilde{d}_n^\star})D\right) \leqslant \textstyle\sum_{n=1}^T \eta_n \|\Gamma_n\| + DH_\eta.$$

Then, the general convex case follows from the corresponding inequality in Theorem 5.6, as follows:

$$\begin{aligned}
\bar{R}_T(u) &\leqslant \mathbb{E}\left[\mathfrak{R}_T(u) + G\,\mathfrak{D}_T\right] + \tfrac{6GR\delta_{\text{tot}}}{r}\\
&\leqslant \tfrac{D^2}{2\eta_T} + \textstyle\sum_{n=1}^T \eta_n (kM/\widetilde{\delta}_n)^2 + G\sum_{n=1}^T \eta_n \mathbb{E}[\|\Gamma_n\|] + GDH_\eta + \tfrac{6GR\delta_{\text{tot}}}{r}\\
&= \tfrac{D^2}{2\eta_T} + \textstyle\sum_{n=1}^T \eta_n\left(G\,\mathbb{E}[\|\Gamma_n\|] + (kM/\widetilde{\delta}_n)^2\right) + GDH_\eta + \tfrac{6GR\delta_{\text{tot}}}{r}.
\end{aligned}$$

The strongly convex case follows from the corresponding inequality in Theorem 5.6:

$$\begin{aligned}
\bar{R}_T(u) &\leqslant \mathbb{E}\left[\mathfrak{R}_T(u) - \tfrac{\lambda}{2}\textstyle\sum_{n=1}^T \|z_n - u\|^2 + 3G\mathfrak{D}_T\right] + \tfrac{10GR\delta_{\text{tot}}}{r}\\
&\leqslant \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2}\mathbb{E}[\|z_n - u\|^2] + \sum_{n=1}^T \eta_n \left(\frac{kM}{\widetilde{\delta}_n}\right)^2\\
&\quad + 3G\left(\textstyle\sum_{n=1}^T \eta_n \mathbb{E}[\|\Gamma_n\|] + DH_\eta\right) + \tfrac{10GR\delta_{\text{tot}}}{r}\\
&= \textstyle\sum_{n=1}^T \frac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2}\,\mathbb{E}[\|z_n - u\|^2]\\
&\quad + \textstyle\sum_{n=1}^T \eta_n\left(3G\,\mathbb{E}[\|\Gamma_n\|] + (kM/\widetilde{\delta}_n)^2\right) + 3GDH_\eta + \tfrac{10GR\delta_{\text{tot}}}{r}.
\end{aligned}$$

This concludes the proof for both cases. □

Lemma E.4. *For $\Gamma_n = \sum_{m=n-\tilde{d}_n^\star}^{n-1} \widetilde{\widetilde{g}}_m^\delta$ from Theorem E.3 and $\delta_n' = \delta_{\rho(n)+d_{\rho(n)}}$, it holds that*

$$G\,\mathbb{E}\left[\|\Gamma_n\|\right] \leqslant 4(G^2\,\tilde{d}_n^\star + (kM/\delta_n')^2).$$

Proof. Note that $\|\widetilde{g}_n^\delta\| \leqslant G$ and $\|\widetilde{\widehat{g}}_n^\delta\| \leqslant \frac{kM}{\widetilde{\delta}_n}$. With probability one, it holds that

$$\|\Gamma_n\| \leqslant \sum_{m=n-\widetilde{d}_n^\star}^{n-1} \|\widetilde{g}_m^\delta\| + \| \underbrace{\sum_{m=n-\widetilde{d}_n^\star}^{n-1} (\widetilde{\widehat{g}}_m^\delta - \widetilde{g}_m^\delta)}_{=:M_n} \| \leqslant G\widetilde{d}_n^\star + \|M_n\|.$$

Since $\delta'_n = \delta_{\rho(n)+d_{\rho(n)}}$, it holds that $\delta'_n \leqslant \delta_t$ for all $t \in \{s : r_s \in I_{\rho(n)}\}$, because for all such t we have $l_t < r_t < r_{\rho(n)} < l_{\rho(n)+d_{\rho(n)}+1}$ (i.e., $t \leqslant \rho(n) + d_{\rho(n)}$) and sequence $(\delta_t)_{t=1}^T$ is non-increasing. Let $G' = G + kM/\delta'_n$ so that $\|\widehat{g}_t^\delta - g_t^\delta\| \leqslant G'$ for all $t \in \{s : r_s \in I_{\rho(n)}\}$. Let $v_t = (\widehat{g}_t^\delta - g_t^\delta)\,\mathbb{I}(r_t \in I_{\rho(n)})$ for $t \in [T]$, so that $M_n = \sum_{t=1}^T v_t$, $\mathbb{E}[v_t] = 0$, and $\|v_t\| \leqslant G'\,\mathbb{I}(r_t \in I_{\rho(n)})$. Consider sigma-algebras $\mathcal{F}_t^u = \sigma(u_1, ..., u_{t-1})$. Then, for all $t \in [T]$, $v_t \in \mathcal{F}_{t+1}^u$ and $\mathbb{E}[v_t|\mathcal{F}_t^u] = 0$. Therefore, it holds that

$$\begin{aligned}
\mathbb{E}\left[\|M_n\|^2\right] &= \sum_{t=1}^T \mathbb{E}\left[\|v_t\|^2\right] + 2\sum_{t=1}^T \sum_{s=t+1}^T \mathbb{E}\left[\langle v_t, v_s\rangle\right] \\
&\leqslant \sum_{t=1}^T (G')^2\,\mathbb{I}(r_t \in I_{\rho(n)}) + 2\sum_{t=1}^T \sum_{s=t+1}^T \mathbb{E}\left[\langle v_t, \mathbb{E}[v_s|\mathcal{F}_s^u]\rangle\right] \\
&= (G')^2\,\widetilde{d}_n^\star + 0.
\end{aligned}$$

Combining these results and applying Jensen's inequality, we prove the inequality

$$\begin{aligned}
G\,\mathbb{E}\left[\|\Gamma_n\|\right] &\leqslant G^2\,\widetilde{d}_n^\star + G\sqrt{\mathbb{E}[\|M_n\|^2]} \leqslant G^2\,\widetilde{d}_n^\star + G(G + kM/\delta'_n)\sqrt{\widetilde{d}_n^\star} \\
&\leqslant 2G^2\,\widetilde{d}_n^\star + (G^2\widetilde{d}_n^\star)^{1/2}(kM/\delta'_n) \leqslant 4(G^2\,\widetilde{d}_n^\star + (kM/\delta'_n)^2).
\end{aligned}$$

This concludes the proof of the lemma. □

Corollary 5.7 (Restated)**.** *$\mathcal{W}_{\text{BCO}}(\mathcal{B})$, where $(\delta_t)_{t\geqslant 1}$ is set to either a round-dependent schedule $\delta_t = r\,\min\left\{1, \frac{\sqrt{\nu k}}{t^{1/4}}\right\}$ or a fixed schedule $\delta_t = r\,\min\left\{1, \frac{\sqrt{\nu k}}{T^{1/4}}\right\}$ and $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{D/G}{\sqrt{n + \sum_{m=1}^n (\widetilde{\sigma}_m^\star + (\nu k r/\delta'_m)^2)}}$, guarantees*

$$\bar{R}_T(x^*) = O\left(GD\left[\sqrt{d_{\text{tot}}} + T^{3/4}\sqrt{\nu k}\right]\right).$$

Proof. The proof for a fixed schedule is the same as for a varying schedule, so we only present the latter. From Theorem E.3 and Lemma E.4, we have the following

$$\begin{aligned}
\bar{R}_T(x^*) &\leqslant \frac{D^2}{2\eta_T} + \sum_{n=1}^T \eta_n \left(G\,\mathbb{E}[\|\Gamma_n\|] + (kM/\widetilde{\delta}_n)^2\right) + GDH_\eta + \frac{6GR\delta_{\text{tot}}}{r} \\
&\leqslant \frac{D^2}{\eta_T} + 5G^2 \sum_{n=1}^T \eta_n \left(\widetilde{d}_n^\star + (\nu k r/\delta'_n)^2\right) + GDH_\eta + \frac{6GD\delta_{\text{tot}}}{r}.
\end{aligned}$$

We bound the terms $1/\eta_T$, $\sum_{n=1}^T \eta_n \left(\widetilde{d}_n^\star + (\nu k r/\delta'_n)^2\right)$, and H_η separately.

1. As $\delta_T = \min_{t\in[T]} \delta_t$ and $\sum_{n=1}^T \widetilde{\sigma}_n^\star = d_{\text{tot}}$, it holds that

$$\frac{1}{\eta_T} = \frac{G}{D}\sqrt{T + \sum_{m=1}^T (\widetilde{\sigma}_m^\star + (\nu k r/\delta'_m)^2)} \leqslant \frac{G}{D}\left(\sqrt{T + d_{\text{tot}}} + \frac{\nu k r}{\delta_T}\sqrt{T}\right).$$

2. As $\sum_{m=1}^n \widetilde{d}_m^\star \leqslant \sum_{m=1}^n \widetilde{\sigma}_n^\star$ for all $n \in [T]$ according to Fact C.2, we write

$$\begin{aligned}
\sum_{n=1}^T \eta_n \left(\widetilde{d}_n^\star + (\nu k r/\delta'_n)^2\right) &\leqslant \frac{D}{G}\sum_{n=1}^T \frac{\widetilde{d}_n^\star + (\nu k r/\delta'_n)^2}{\sqrt{\sum_{m=1}^n (\widetilde{d}_m^\star + (\nu k r/\delta'_m)^2)}} \\
&\leqslant \frac{2D}{G}\sqrt{\sum_{n=1}^T (\widetilde{d}_n^\star + (\nu k r/\delta'_n)^2)} \\
&\leqslant \frac{2D}{G}\left(\sqrt{T + d_{\text{tot}}} + (\nu k r/\delta_T)\sqrt{T}\right),
\end{aligned}$$

where the second inequality follows from Fact B.4 with $X_n = \widetilde{d}_n^\star + (\nu k r/\delta'_n)^2$.

3. For all $n \in [T]$, it holds that

$$\sum_{m=n}^{n+\widetilde{\sigma}_n^\star} \eta_m \leqslant \sum_{m=n}^{n+\widetilde{\sigma}_n^\star} \frac{D/G}{\sqrt{m+\sum_{k=n}^{m}\widetilde{\sigma}_k^\star}} \overset{\text{(a)}}{\leqslant} \sum_{m=0}^{\widetilde{\sigma}_n^\star} \frac{D/G}{\sqrt{(m+1)+\sum_{k=0}^{m}(\widetilde{\sigma}_n^\star-k)}}$$
$$\overset{\text{(b)}}{\leqslant} \sum_{m=0}^{\widetilde{\sigma}_n^\star} \frac{2D/G}{\sqrt{(m+1)(\widetilde{\sigma}_n^\star+1)}} \overset{\text{(c)}}{\leqslant} \frac{4D}{G},$$

where (a) follows from Fact C.3, (b) applies inequality $\sum_{k=0}^{m}(\widetilde{\sigma}_n^\star - k) = \frac{(m+1)(2\widetilde{\sigma}_n^\star - m)}{2} \geqslant \frac{(m+1)\widetilde{\sigma}_n^\star}{2}$, and (c) uses Fact B.4 to conclude $\sum_{m=0}^{\widetilde{\sigma}_n^\star} \frac{1}{\sqrt{m+1}} \leqslant 2\sqrt{\widetilde{\sigma}_n^\star + 1}$. Therefore, by Lemma D.6 and the bound on $1/\eta_T$ above, we have

$$H_\eta \leqslant \frac{\max_{n\in[T]} \sum_{m=n}^{n+\widetilde{\sigma}_n^\star} \eta_m}{\eta_T} \leqslant 8\left(\sqrt{T + d_{\text{tot}}} + (\nu k r/\delta_T)\sqrt{T}\right).$$

Combining all these results, we have

$$\bar{R}_T(x^*) = O\left(GD\left(\sqrt{T + d_{\text{tot}}} + (\nu k r/\delta_T)\sqrt{T} + \delta_{\text{tot}}/r\right)\right).$$

Next, we provide the bounds on r/δ_T and δ_{tot}/r in order to derive the final bound. Trivially, it holds that $r/\delta_T \leqslant \max\left\{1, \frac{T^{1/4}}{\sqrt{\nu k}}\right\}$. Also, note that $\sum_{t=1}^{T} \frac{1}{t^{1/4}} \leqslant 1 + \int_1^T x^{-1/4}\,dx = 1 + \frac{4}{3}(T^{3/4} - 1) \leqslant \frac{4}{3}T^{3/4}$, so that

$$\delta_{\text{tot}}/r = \sum_{t=1}^{T} \min\{1, \tfrac{\sqrt{\nu k}}{t^{1/4}}\} \leqslant \sum_{t=1}^{T} \tfrac{\sqrt{\nu k}}{t^{1/4}} \leqslant \tfrac{4}{3}T^{3/4}\sqrt{\nu k}.$$

Thus, we have $r/\delta_T \leqslant \max\left\{1, \frac{T^{1/4}}{\sqrt{\nu k}}\right\}$ and $\delta_{\text{tot}}/r \leqslant \frac{4}{3}T^{3/4}\sqrt{\nu k}$. Finally, substituting the bounds for r/δ_T and δ_{tot}/r, we have

$$\bar{R}_T(x^*) = O\left(GD\left(\sqrt{T + d_{\text{tot}}} + T^{3/4}\sqrt{\nu k} + T^{1/2}(\nu k)\right)\right) = O\left(GD\left(\sqrt{d_{\text{tot}}} + T^{3/4}\sqrt{\nu k}\right)\right).$$

□

Corollary 5.8 (Restated). *Under λ-strong convexity (2.3), $\mathcal{W}_{\text{BCO}}(\mathcal{B})$, where $(\delta_t)_{t\geqslant 1}$ is set to either a round-dependent schedule $\delta_t = r \min\left\{1, (\frac{\nu^2 k^2 \ln(et)}{t})^{1/3}\right\}$ or a fixed schedule $\delta_t = r \min\left\{1, (\frac{\nu^2 k^2 \ln(eT)}{T})^{1/3}\right\}$ and $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{1}{n\lambda}$, guarantees*

$$\bar{R}_T(x^*) = O\Big(\tfrac{G^2}{\lambda}\left[\min\{\sigma_{\max}\ln T,\ \sqrt{d_{\text{tot}}}\} + T^{2/3}(\ln T)^{1/3}(\nu k)^{2/3}\right]\Big).$$

Proof. The proof for a fixed schedule is the same as for a varying schedule, so we only present the latter. From Theorem E.3 and Lemma E.4, noting that $1/\eta_n - 1/\eta_{n-1} - \lambda = 0$, we have the following

$$\bar{R}_T(x^*) \leqslant 0 + \sum_{n=1}^{T} \eta_n \left(3G\,\mathbb{E}[\|\Gamma_n\|] + (kM/\widetilde{\delta}_n)^2\right) + 3GDH_\eta + \tfrac{10GR\delta_{\text{tot}}}{r},$$
$$\leqslant 13G^2 \sum_{n=1}^{T} \eta_n \left(\widetilde{d}_n^\star + (\nu k r/\delta_n')^2\right) + 3GDH_\eta + \tfrac{10GD\delta_{\text{tot}}}{r}.$$

We bound the terms $\sum_{n=1}^{T} \eta_n \left(\widetilde{d}_n^\star + (\nu k r/\delta_n')^2\right)$ and H_η separately.

1. Using Fact C.4 for dual delays $\widetilde{d}_n^\star$ and Fact B.3, we write

$$\sum_{n=1}^{T} \eta_n \left(\widetilde{d}_n^\star + (\nu k r/\delta_n')^2\right) \leqslant \tfrac{1}{\lambda} \min\left\{\sigma_{\max}\ln(eT),\ 2\sqrt{d_{\text{tot}}}\right\} + \tfrac{1}{\lambda}(\nu k r/\delta_T)^2 \ln(eT).$$

2. Substituting learning rate values into H_η and applying Fact C.4 for dual delays $\widetilde{d}_n^\star$ yields

$$H_\eta = \sum_{n=1}^{T}(1 - \tfrac{\eta_n}{\eta_{n-\widetilde{d}_n^\star}}) = \sum_{n=1}^{T} \tfrac{\widetilde{d}_n^\star}{n} \leqslant \min\left\{\sigma_{\max}\ln(eT),\ 2\sqrt{d_{\text{tot}}}\right\}.$$

Combining all these results and letting $D \leqslant \frac{2G}{\lambda}$ (see Fact B.1), we have

$$\begin{aligned}\bar{R}_T(x^*) &\leqslant \left(\tfrac{13G^2}{\lambda} + 3GD\right) \min\left\{\sigma_{\max}\ln(eT),\, 2\sqrt{d_{\text{tot}}}\right\} + \tfrac{13G^2}{\lambda}(\nu k r/\delta_T)^2\ln(eT) + \tfrac{10GD\delta_{\text{tot}}}{r} \\ &= O\left(\tfrac{G^2}{\lambda}\left(\min\left\{\sigma_{\max}\ln(eT),\, \sqrt{d_{\text{tot}}}\right\} + (\nu k r/\delta_T)^2\ln(eT) + \delta_{\text{tot}}/r\right)\right).\end{aligned}$$

Next, we provide the bounds on r/δ_T and δ_{tot}/r in order to derive the final bound. Trivially, it holds that $r/\delta_T \leqslant \max\left\{1, (\frac{T}{\nu^2k^2\ln(eT)})^{1/3}\right\}$. Also, note that $\sum_{t=1}^T \frac{1}{t^{1/3}} \leqslant 1 + \int_1^T x^{-1/3}\,dx = 1 + \frac{3}{2}(T^{2/3} - 1) \leqslant \frac{3}{2}T^{2/3}$, so that

$$\delta_{\text{tot}}/r = \sum_{t=1}^T \min\left\{1, (\tfrac{\nu^2k^2\ln(et)}{t})^{1/3}\right\} \leqslant \sum_{t=1}^T \tfrac{(\nu^2k^2\ln(eT))^{1/3}}{t^{1/3}} \leqslant \tfrac{3}{2}T^{2/3}\ln(eT)^{1/3}(\nu k)^{2/3}.$$

Thus, we have $r/\delta_T \leqslant \max\left\{1, (\frac{T}{\nu^2k^2\ln(eT)})^{1/3}\right\}$ and $\delta_{\text{tot}}/r \leqslant \frac{3}{2}T^{2/3}\ln(eT)^{1/3}(\nu k)^{2/3}$. Finally, substituting the bounds for r/δ_T and δ_{tot}/r, we have

$$\begin{aligned}\bar{R}_T(x^*) &= O\left(\tfrac{G^2}{\lambda}\left(\min\{\sigma_{\max}\ln(eT),\, \sqrt{d_{\text{tot}}}\} + T^{2/3}\ln(eT)^{1/3}(\nu k)^{2/3} + \nu^2k^2\ln(eT)\right)\right) \\ &= O\left(\tfrac{G^2}{\lambda}\left(\min\{\sigma_{\max}\ln(eT),\, \sqrt{d_{\text{tot}}}\} + T^{2/3}\ln(eT)^{1/3}(\nu k)^{2/3}\right)\right).\end{aligned}$$

□

F Bandit Convex Optimization with Two-Point Delayed Feedback

In this section, we extend the analysis of BCO with delayed feedback (see Section 2) to the two-point feedback setting Agarwal et al. [2010]. In each round t, the player selects a pair of points $x_t^1, x_t^2 \in \mathcal{K}$ (the same domain as Section 2), incurs loss $\frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big)$, and later receives the delayed feedback $(t, f_t(x_t^1), f_t(x_t^2))$ in round $t + d_t$.

BCO with Two-Point Delayed Feedback

- *Latent Parameters:* number of rounds T.
- *Pre-game:* adversary selects loss functions $f_t : \mathcal{K} \to \mathbb{R}$ and delays $d_t \in \{0, \dots, T - t\}$ for all $t \in [T]$.

For each round $t = 1, 2, \dots, T$:

1. The player predicts points (x_t^1, x_t^2) and incurs loss $\frac{1}{2}(f_t(x_t^1) + f_t(x_t^2))$.
2. For all $s \leqslant t$ such that $s + d_s = t$, the environment reveals $(s, f_s(x_s^1), f_s(x_s^2))$.

Figure 5: Bandit convex optimization with two-point delayed feedback.

The two-point regret against a comparator $u \in \mathcal{K}$ is defined as $R_T^{\text{2p}}(u) = \sum_{t=1}^T(\frac{1}{2}(f_t(x_t^1) + f_t(x_t^2)) - f_t(u))$ and its expectation as $\bar{R}_T^{\text{2p}}(u) = \mathbb{E}[R_T^{\text{2p}}(u)]$. Letting $x^* \in \operatorname{argmin}_{x\in\mathcal{K}} \sum_{t=1}^T f_t(x)$ denote the best action in hindsight, the goal is to minimize $\bar{R}_T^{\text{2p}}(x^*)$.

Preliminaries on two-point gradient estimation. We briefly review the standard two-point gradient estimator. By Theorem 5.5, the gradient of the δ-smoothed loss f^δ admits a spherical expectation representation; by symmetry of $u \sim \text{Unif}(\mathbb{S}^{k-1})$, this can be written in the two-point form

$$\nabla f^\delta(x) = \mathbb{E}_{u\sim\text{Unif}(\mathbb{S}^{k-1})}\left[\tfrac{k}{2\delta}\big(f(x + \delta u) - f(x - \delta u)\big)u\right].$$

Accordingly, the estimator $\frac{k}{2\delta}\big(f(x + \delta u) - f(x - \delta u)\big)u$ is unbiased for $\nabla f^\delta(x)$ [Agarwal et al., 2010, Shamir, 2017]. Moreover, unlike the one-point estimator, its second moment does not scale with $1/\delta$, as captured by the following lemma.

Lemma F.1 (Lemma 5, Shamir [2017]). *There exists a numerical constant $c > 0$ such that for every G-Lipschitz function $f : \mathcal{K} \to \mathbb{R}$, $\delta \in (0, r]$, and $x \in (1 - \delta/r)\mathcal{K}$, it holds that*

$$\mathbb{E}_{u\sim\text{Unif}(\mathbb{S}^{k-1})}\left[\left\|\tfrac{k}{2\delta}\big(f(x + \delta u) - f(x - \delta u)\big)u\right\|^2\right] \leqslant c\,kG^2.$$

Two-point Delayed BCO via Drift-Penalized OLO. The wrapper $\mathcal{W}_{\text{2p-BCO}}$ (Algorithm 3) follows the structure of $\mathcal{W}_{\text{BCO}}$, with two-point gradient estimates instead of single-point estimates. As before, the algorithm takes as input a non-increasing sequence of smoothing parameters $(\delta_t)_{t\geqslant 1} \subset (0, r]$ fixed in advance. In round t, given the current base-prediction $y_t = \bar{z}$ from $\mathcal{B}$, the algorithm samples an exploration direction $u_t \sim \text{Unif}(\mathbb{S}^{k-1})$ and predicts $x_t^1 = x_t + \delta_t u_t$, $x_t^2 = x_t - \delta_t u_t$ for $x_t = (1 - \delta_t/r)y_t$. Upon receiving observations $(s, f_s(x_s^1), f_s(x_s^2))$, it constructs the two-point gradient estimate $\widehat{g}_s^\delta = \frac{k}{2\delta_s}(f_s(x_s^1) - f_s(x_s^2))u_s$. This estimate satisfies $\mathbb{E}[\widehat{g}_t^\delta \mid x_t] = \nabla f_t^\delta(x_t)$ with reduced variance compared to single-point estimation.

Algorithm 3: $\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$

Access : Algorithm $\mathcal{B}$ for OLO with drift penalization.
Parameters : Non-increasing sequence $(\delta_t)_{t\geqslant 1} \subset (0, r]$.

Initialize $\bar{z}$ with initial output of $\mathcal{B}$.

For round $t = 1, 2, \ldots, T$:

1. Select $x_t = (1 - \delta_t/r)\bar{z}$, sample $u_t \sim \text{Unif}(\mathbb{S}^{k-1})$, and predict $x_t^1 = x_t + \delta_t u_t$, $x_t^2 = x_t - \delta_t u_t$.
2. For each round s such that $s + d_s = t$, in the order of observation:
   - Receive $(s, f_s(x_s^1), f_s(x_s^2))$, forward $(\frac{k}{2\delta_s}(f_s(x_s^1) - f_s(x_s^2))u_s, d_s^\star, \sigma_s^\star)$ to $\mathcal{B}$, and set $\bar{z}$ to the new output of $\mathcal{B}$.

The analysis follows the same structure as for $\mathcal{W}_{\text{BCO}}$. The algorithm updates $\mathcal{B}$ sequentially on tuples $(\widetilde{\widehat{g}}_n^\delta, \widetilde{d}_n^\star, \widetilde{\sigma}_n^\star)$ for $n \in [T]$, where $\widetilde{\widehat{g}}_n^\delta = \frac{k}{2\widetilde{\delta}_n}(\widetilde{f}_n(\widetilde{x}_n^1) - \widetilde{f}_n(\widetilde{x}_n^2))\widetilde{u}_n$, producing base-predictions z_n that yield predictions $\widetilde{x}_n^1 = \widetilde{x}_n + \widetilde{\delta}_n\widetilde{u}_n$ and $\widetilde{x}_n^2 = \widetilde{x}_n - \widetilde{\delta}_n\widetilde{u}_n$ for $\widetilde{x}_n = (1 - \widetilde{\delta}_n/r)\widetilde{y}_n$ and $\widetilde{y}_n = z_{n-\widetilde{d}_n^\star}$. The following theorem bounds the expected two-point regret of $\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$ in terms of the drift-penalized regret of $\mathcal{B}$, with an additional bias of order $\delta_{\text{tot}} = \sum_{t=1}^T \delta_t$ due to smoothing.

Theorem F.2 (Regret of Algorithm 3)**.** *For any base-algorithm $\mathcal{B}$, non-increasing sequence $(\delta_t)_{t\geqslant 1} \subset (0, r]$, and comparator $u \in \mathcal{K}$, $\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$ guarantees*

$$\bar{R}_T^{\text{2p}}(u) \leqslant \mathbb{E}\left[\mathfrak{R}_T^{\text{drift}}(u; G)\right] + \tfrac{7GR\delta_{\text{tot}}}{r},$$

where $\mathfrak{R}_T^{\text{drift}}(u; W)$ is the drift-penalized regret (3) of $\mathcal{B}$ for loss vectors $c_n = \widetilde{\widehat{g}}_n^\delta$ and lags $\lambda_n = \widetilde{d}_n^\star$.

Under λ-strong convexity (2.3), it further holds that

$$\bar{R}^{\text{2p}}(u) \leqslant \mathbb{E}\Big[\mathfrak{R}_T^{\text{drift}}(u; 3G) - \tfrac{\lambda}{2}\textstyle\sum_{n=1}^T \|z_n - u\|^2\Big] + \tfrac{11GR\delta_{\text{tot}}}{r}.$$

Proof. The argument follows the proof of Theorem 5.6. Here the played center x_t plays the role of y_t in that theorem, so the same steps bound $\sum_{t=1}^T \big(f_t(x_t) - f_t(u)\big)$. It remains to relate the two-point loss to the center loss: by G-Lipschitzness,

$$\left|\tfrac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - f_t(x_t)\right| \leqslant \tfrac{G}{2}\big(\|x_t^1 - x_t\| + \|x_t^2 - x_t\|\big) = G\delta_t.$$

Summing over t yields $\left|\sum_{t=1}^T \left[\tfrac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - f_t(x_t)\right]\right| \leqslant G\delta_{\text{tot}} \leqslant \tfrac{GR\delta_{\text{tot}}}{r}$, which gives the claim. □

Selecting an appropriate smoothing sequence $(\delta_t)_{t=1}^T$ and instantiating $\mathcal{B}$ with P-FTRL (4) yields concrete guarantees.

Corollary F.3. *$\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$ with $\delta_t = \frac{r}{\sqrt{t}}$, where $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{D/G}{\sqrt{nk + \sum_{m=1}^n \widetilde{\sigma}_m^\star}}$, guarantees*

$$\bar{R}_T^{\text{2p}}(x^*) = O\left(GD\left[\sqrt{d_{\text{tot}}} + \sqrt{Tk}\right]\right).$$

Composing with the skipping scheme (Algorithm 4), $\mathcal{S}_{\text{skip}}(\mathcal{W}_{\text{2p-BCO}}(\mathcal{B}))$ gets $\bar{R}_T^{\text{2p}}(x^*) = O(GD[\min_{Q\subseteq[T]}\{|Q| + (\sum_{t\notin Q} d_t)^{\frac{1}{2}}\} + \sqrt{Tk}])$.

Corollary F.4. *Under λ-strong convexity (2.3), $\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$ with $\delta_t = \frac{r}{t}$, where $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{1}{n\lambda}$, guarantees*

$$\bar{R}_T^{\text{2p}}(x^*) = O\left(\tfrac{G^2}{\lambda}\left[\min\{\sigma_{\max}\ln(T), \sqrt{d_{\text{tot}}}\} + k\ln T\right]\right).$$

F.1 Proofs of Corollaries F.3 and F.4

Theorem F.5. *$\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$, where $\mathcal{B}$ executes P-FTRL (4) updates, guarantees that in the*

$$\textit{general convex case}\colon\ \bar{R}_T^{\text{2p}}(u) \leqslant \tfrac{D^2}{2\eta_T} + \textstyle\sum_{n=1}^T \eta_n \left(G\,\mathbb{E}[\|\Gamma_n\|] + G^2\,ck\right) + GDH_\eta + \tfrac{7GR\delta_{\text{tot}}}{r},$$

$$\textit{strongly convex case (2.3)}\colon\ \bar{R}_T^{\text{2p}}(u) \leqslant \textstyle\sum_{n=1}^T \tfrac{1/\eta_n - 1/\eta_{n-1} - \lambda}{2}\,\mathbb{E}[\|z_n - u\|^2] + \sum_{n=1}^T \eta_n \left(3G\,\mathbb{E}[\|\Gamma_n\|] + G^2\,ck\right) + 3GDH_\eta + \tfrac{11GR\delta_{\text{tot}}}{r},$$

where $\Gamma_n = \sum_{m=n-\widetilde{d}_n^\star}^{n-1} \widetilde{\widehat{g}}_m^\delta$, $H_\eta = \sum_{n=1}^T (1 - \eta_n/\eta_{n-\widetilde{d}_n^\star})$, and c is the constant from Lemma F.1.

Proof. Follows identically to Theorem E.3, using the bound $\mathbb{E}[\|\widetilde{\widehat{g}}_n^\delta\|^2] \leqslant c\,kG^2$ from Lemma F.1. □

Lemma F.6. *For the constant c from Lemma F.1, $\Gamma_n = \sum_{m=n-\widetilde{d}_n^\star}^{n-1} \widetilde{\widehat{g}}_m^\delta$ satisfies*

$$\mathbb{E}\left[\|\Gamma_n\|_2\right] \leqslant 2G\,(\widetilde{d}_n^\star + ck).$$

Proof. Note that $\|\widetilde{g}_n^\delta\| \leqslant G$, $\mathbb{E}[\widetilde{\widehat{g}}_n^\delta] = \widetilde{g}_n^\delta$ by Theorem 5.5 and $\mathbb{E}[\|\widetilde{\widehat{g}}_n^\delta - \widetilde{g}_n^\delta\|^2] \leqslant ckG^2$ by Lemma F.1. With probability one, we write

$$\|\Gamma_n\| \leqslant \textstyle\sum_{m=n-\widetilde{d}_n^\star}^{n-1} \|\widetilde{g}_m^\delta\| + \|\underbrace{\textstyle\sum_{m=n-\widetilde{d}_n^\star}^{n-1} (\widetilde{\widehat{g}}_m^\delta - \widetilde{g}_m^\delta)}_{=:M_n}\| \leqslant G\widetilde{d}_n^\star + \|M_n\|.$$

Let $v_t = (\widehat{g}_t^\delta - g_t^\delta)\,\mathbb{I}(r_t \in I_{\rho(n)})$, so that $M_n = \sum_{t=1}^T v_t$, $\mathbb{E}[v_t] = 0$, and $\mathbb{E}[\|v_t\|^2] \leqslant ckG^2\,\mathbb{I}(r_t \in I_{\rho(n)})$. Let $\mathcal{F}_t^u = \sigma(u_1, ..., u_{t-1})$. Then, for all $t \in [T]$, $v_t \in \mathcal{F}_{t+1}^u$ and $\mathbb{E}[v_t|\mathcal{F}_t^u] = 0$. Therefore, it holds that

$$\begin{aligned}
\mathbb{E}\left[\|M_n\|^2\right] &= \textstyle\sum_{t=1}^T \mathbb{E}\left[\|v_t\|^2\right] + 2\sum_{t=1}^T \sum_{s=t+1}^T \mathbb{E}\left[\langle v_t, v_s\rangle\right] \\
&\leqslant \textstyle\sum_{t=1}^T ckG^2\,\mathbb{I}(r_t \in I_{\rho(n)}) + 2\sum_{t=1}^T \sum_{s=t+1}^T \mathbb{E}\left[\langle v_t, \mathbb{E}[v_s|\mathcal{F}_s^u]\rangle\right] \\
&= c\,kG^2\,\widetilde{d}_n^\star + 0.
\end{aligned}$$

Combining these results and applying Jensen's inequality, we prove the inequality

$$\mathbb{E}\left[\|\Gamma_n\|\right] \leqslant G\,\widetilde{d}_n^\star + \sqrt{\mathbb{E}[\|M_n\|^2]} \leqslant G\,\widetilde{d}_n^\star + G(ck)^{1/2}\sqrt{\widetilde{d}_n^\star} \leqslant 2G\,(\widetilde{d}_n^\star + ck).$$

This concludes the proof of the lemma. □

Corollary F.3 (Restated)**.** *$\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$ with $\delta_t = \frac{r}{\sqrt{t}}$, where $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{D/G}{\sqrt{nk + \sum_{m=1}^n \widetilde{\sigma}_m^\star}}$, guarantees*

$$\bar{R}_T^{\text{2p}}(x^*) = O\left(GD\left[\sqrt{d_{\text{tot}}} + \sqrt{Tk}\right]\right).$$

Composing with the skipping scheme, $\mathcal{S}_{\text{skip}}(\mathcal{W}_{\text{2p-BCO}}(\mathcal{B}))$ gets $\bar{R}_T^{\text{2p}}(x^*) = O(GD[\min_{Q\subseteq[T]}\{|Q| + (\sum_{t\notin Q} d_t)^{\frac{1}{2}}\} + \sqrt{Tk}])$.

Proof. From Theorem F.5 and Lemma F.6, we have the following

$$\begin{aligned}
\bar{R}_T^{\text{2p}}(x^*) &\leqslant \tfrac{D^2}{2\eta_T} + \textstyle\sum_{n=1}^T \eta_n \left(G\,\mathbb{E}[\|\Gamma_n\|] + G^2\,ck\right) + GDH_\eta + \tfrac{7GR\delta_{\text{tot}}}{r} \\
&\leqslant \tfrac{D^2}{\eta_T} + 5(c+1)G^2 \textstyle\sum_{n=1}^T \eta_n(\widetilde{d}_n^\star + k) + GDH_\eta + \tfrac{7GD\delta_{\text{tot}}}{r}.
\end{aligned}$$

We bound the terms $1/\eta_T$, $\sum_{n=1}^T \eta_n(\widetilde{d}_n^\star + k)$, and H_η separately.

1. As $\sum_{n=1}^T \widetilde{\sigma}_n^\star = d_{\text{tot}}$ according to Theorem 3.4, it holds that

$$\tfrac{1}{\eta_T} = \tfrac{G}{D}\sqrt{Tk + \textstyle\sum_{m=1}^T \widetilde{\sigma}_m^\star} \leqslant \tfrac{G}{D}\left(\sqrt{d_{\text{tot}}} + \sqrt{Tk}\right).$$

2. As $\sum_{m=1}^{n} \widetilde{d}_m^\star \leqslant \sum_{m=1}^{n} \widetilde{\sigma}_n^\star$ for all $n \in [T]$ according to Fact C.2, we write

$$\begin{aligned}\sum_{n=1}^{T} \eta_n(\widetilde{d}_n^\star + k) &\leqslant \tfrac{D}{G} \sum_{n=1}^{T} \frac{\widetilde{d}_n^\star + k}{\sqrt{\sum_{m=1}^{n}(\widetilde{d}_m^\star + k)}} \\ &\leqslant \tfrac{2D}{G}\sqrt{\sum_{n=1}^{T}(\widetilde{d}_n^\star + k)} \\ &\leqslant \tfrac{2D}{G}\left(\sqrt{d_{\text{tot}}} + \sqrt{Tk}\right),\end{aligned}$$

where the second inequality follows from Fact B.4 with $X_n = \widetilde{d}_n^\star + k$.

3. Following the similar argument from the proof of Corollary 5.7, it holds that $\max_{n\in[T]} \sum_{m=n}^{n+\widetilde{\sigma}_n^\star} \eta_m \leqslant \frac{4D}{G}$. Then, by Lemma D.6 and the bound on $1/\eta_T$ above, we have

$$H_\eta \leqslant \frac{\max_{n\in[T]} \sum_{m=n}^{n+\widetilde{\sigma}_n^\star} \eta_m}{\eta_T} \leqslant 4\left(\sqrt{d_{\text{tot}}} + \sqrt{Tk}\right).$$

For our choice of δ_t, we have $\delta_{\text{tot}}/r = \sum_{t=1}^{T} \frac{1}{t^{1/2}} \leqslant 2\sqrt{T}$, which follows immediately from Fact B.4. Combining all these results, we have

$$\bar{R}_T^{\text{2p}}(x^*) = O\left(GD\left[\sqrt{d_{\text{tot}}} + \sqrt{Tk}\right]\right).$$

The result for the skipping scheme follows from Theorem G.1. □

Corollary F.4 (Restated). *Under λ-strong convexity (2.3), $\mathcal{W}_{\text{2p-BCO}}(\mathcal{B})$ with $\delta_t = \frac{r}{t}$, where $\mathcal{B}$ runs P-FTRL (4) with $\eta_n = \frac{1}{n\lambda}$, guarantees*

$$\bar{R}_T^{\text{2p}}(x^*) = O\left(\tfrac{G^2}{\lambda}\left[\min\{\sigma_{\max}\ln T,\ \sqrt{d_{\text{tot}}}\} + k\ln T\right]\right).$$

Proof. From Theorem F.5 and Lemma F.6, noting that $1/\eta_n - 1/\eta_{n-1} - \lambda = 0$, we have the following

$$\begin{aligned}\bar{R}_T^{\text{2p}}(x^*) &\leqslant 0 + \sum_{n=1}^{T} \eta_n\left(3G\,\mathbb{E}[\|\Gamma_n\|] + G^2\,ck\right) + 3GDH_\eta + \tfrac{11GR\delta_{\text{tot}}}{r}, \\ &\leqslant 13(c+1)G^2 \sum_{n=1}^{T} \eta_n(\widetilde{d}_n^\star + k) + 3GDH_\eta + \tfrac{11GD\delta_{\text{tot}}}{r}.\end{aligned}$$

We bound the terms $\sum_{n=1}^{T} \eta_n(\widetilde{d}_n^\star + k)$ and H_η separately.

1. Using Fact C.4 for dual delays $\widetilde{d}_n^\star$ and Fact B.3, we write

$$\sum_{n=1}^{T} \eta_n(\widetilde{d}_n^\star + k) \leqslant \tfrac{1}{\lambda}\min\left\{\sigma_{\max}\ln(eT),\ 2\sqrt{d_{\text{tot}}}\right\} + \tfrac{1}{\lambda}k\ln(eT).$$

2. Substituting learning rate values into H_η and applying Fact C.4 for dual delays $\widetilde{d}_n^\star$ yields

$$H_\eta = \sum_{n=1}^{T}(1 - \tfrac{\eta_n}{\eta_{n-\widetilde{d}_n^\star}}) = \sum_{n=1}^{T} \tfrac{\widetilde{d}_n^\star}{n} \leqslant \min\left\{\sigma_{\max}\ln(eT),\ 2\sqrt{d_{\text{tot}}}\right\}.$$

For our choice of δ_t, we have $\delta_{\text{tot}}/r = \sum_{t=1}^{T} \frac{1}{t} \leqslant \ln(eT)$, which follows immediately from Fact B.3. Combining all these results and letting $D \leqslant \frac{2G}{\lambda}$ (see Fact B.1), we have

$$\bar{R}_T^{\text{2p}}(x^*) = O\left(\tfrac{G^2}{\lambda}\left[\min\{\sigma_{\max}\ln T,\ \sqrt{d_{\text{tot}}}\} + k\ln T\right]\right).$$

This concludes the proof of the corollary. □

G Skipping Scheme

The $O(\sqrt{d_{\text{tot}}})$ term in Corollaries 5.3 and 5.7 is pessimistic when delays are highly unbalanced: if a few rounds have exceptionally large delays, it is preferable to skip them (i.e., accept full loss and ignore their feedback), improving the bound to $O(\min_{Q\subseteq[T]}\{|Q| + (\sum_{t\notin Q} d_t)^{\frac{1}{2}}\})$. For instance,

following Thune et al. [2019], if the first $\lfloor \sqrt{T} \rfloor$ rounds have delay $\Theta(T)$ and the rest have 1, we get improvement from $O(T^{3/4})$ to $O(\sqrt{T})$. However, as delays are revealed upon feedback arrival, achieving this adaptively is nontrivial.

To address this, we adopt the skipping technique of Zimmert and Seldin [2020] as an external wrapper, $\mathcal{S}_{\text{skip}}$ (Algorithm 4), making it applicable to any delayed OCO algorithm $\mathcal{A}$. The wrapper acts as an interface between $\mathcal{A}$ and the environment, maintaining a tracking set S of rounds awaiting feedback. When feedback for $s \in S$ arrives, it is forwarded to $\mathcal{A}$; if s remains pending for too long, it is "skipped" by forwarding zero-valued feedback and removing s from S. Thus, in effect, $\mathcal{A}$ is presented with a different delayed OCO problem, one with shorter delays and zero losses on skipped rounds.

Algorithm 4: $\mathcal{S}_{\text{skip}}(\mathcal{A})$

Access : Algorithm $\mathcal{A}$ for OCO (BCO) with delays.

Initialize tracking set $S = \varnothing$ and set $\mathcal{D}_0 = 0$.
For round $t = 1, 2, \ldots, T$:
1. Set $\mathcal{D}_t = \mathcal{D}_{t-1} + |S|$.
2. **For** all $s \in S$ such that $t - s > \sqrt{\mathcal{D}_t}$:
   - Forward $(s, 0)$ to $\mathcal{A}$ and remove s from S.
3. Query $\mathcal{A}$ for the next decision, obtain x_t, play x_t, and insert t into the tracking set S.
4. **For** all s such that $s + d_s = t$, in observation order:
   - Receive (s, v_s) from the environment.
   - **If** $s \in S$ **then** forward (s, v_s) to $\mathcal{A}$, remove s from S.

By construction, for every $t \in [T]$, the wrapper passes feedback with index t to $\mathcal{A}$ at round $t + d'_t$ for some $d'_t \in \{0, \ldots, d_t\}$: either the true feedback (t, v_t) upon arrival or dummy feedback $(t, 0)$ forwarded preemptively. Let $Q^* \subseteq [T]$ denote the set of skipped rounds, i.e., those for which dummy feedback was forwarded, and define $f'_t = f_t \, \mathbb{I}(t \notin Q^*)$. Then, in effect, $\mathcal{S}_{\text{skip}}$ presents $\mathcal{A}$ with a delayed OCO problem having losses $\{f'_t\}_{t=1}^T$ and delays $\{d'_t\}_{t=1}^T$. The following theorem bounds the regret of $\mathcal{S}_{\text{skip}}(\mathcal{A})$, with proof in the following subsection.

Theorem G.1. *The regret $R_T(u)$ of $\mathcal{S}_{\text{skip}}(\mathcal{A})$ against any comparator $u \in \mathcal{K}$ satisfies $R_T(u) \leqslant R'_T(u) + GD|Q^*|$, where $R'_T(u) = \sum_{t=1}^T (f'_t(x_t) - f'_t(u))$ is the regret of $\mathcal{A}$ on delayed OCO problem with losses and delays $\{f'_t, d'_t\}_{t=1}^T$. Moreover, the number of skipped rounds $|Q^*|$ and the total modified delay $d'_{\text{tot}} = \sum_{t=1}^T d'_t$ satisfy*

$$|Q^*| + \sqrt{d'_{\text{tot}}} = O\left(\min_{Q \subseteq [T]} \left\{ |Q| + \sqrt{\textstyle\sum_{t \notin Q} d_t} \right\}\right).$$

Composing $\mathcal{S}_{\text{skip}}$ with wrappers $\mathcal{W}_{\text{OCO}}$ and $\mathcal{W}_{\text{BCO}}$ immediately improves upon Corollaries 5.3 and 5.7, respectively.

Corollary G.2. *Under the conditions of Corollaries 5.3 and 5.7, applying the skipping wrapper yields:*

$$R_T(x^*) = O\left(GD\left[\min_{Q \subseteq [T]} \left\{ |Q| + \sqrt{\textstyle\sum_{t \notin Q} d_t} \right\} + \sqrt{T}\right]\right),$$
$$\bar{R}_T(x^*) = O\left(GD\left[\min_{Q \subseteq [T]} \left\{ |Q| + \sqrt{\textstyle\sum_{t \notin Q} d_t} \right\} + T^{\frac{3}{4}}\sqrt{\nu k}\right]\right),$$

for $\mathcal{S}_{\text{skip}}(\mathcal{W}_{\text{OCO}}(\mathcal{B}))$ and $\mathcal{S}_{\text{skip}}(\mathcal{W}_{\text{BCO}}(\mathcal{B}))$, respectively.

Proof. Follows from Theorem G.1 combined with Corollaries 5.3 and 5.7 for controlling $R'_T(x^*)$. □

This skipping scheme applies in any setting where a zero loss is admissible, e.g., the general convex case. In the strongly convex case, one could instead skip rounds by substituting $\frac{\lambda}{2}x^2$ loss values. However, $\min_{Q \subseteq [T]}\{|Q| + (\sum_{t \notin Q} d_t)^{1/2}\} = \Omega(\sigma_{\max})$ as shown by Qiu et al. [2025], so our strongly convex bound $O(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\})$ already matches this benchmark up to a $\ln T$ factor.

G.1 Proof of Theorem G.1

Theorem G.1 (Restated). *The regret $R_T(u)$ of $\mathcal{S}_{\text{skip}}(\mathcal{A})$ against any comparator $u \in \mathcal{K}$ satisfies*

$$R_T(u) \leqslant R'_T(u) + GD|Q^*|,$$

where $R'_T(u) = \sum_{t=1}^T (f'_t(x_t) - f'_t(u))$ is the regret of $\mathcal{A}$ on delayed OCO problem with losses and delays $\{f'_t, d'_t\}_{t=1}^T$. Moreover, it holds that

$$|Q^*| + \sqrt{\textstyle\sum_{t=1}^T d'_t} = O\left(\min_{Q\subseteq[T]} \left\{|Q| + \sqrt{\textstyle\sum_{t\notin Q} d_t}\right\}\right).$$

Proof. The regret bound follows directly from the observation that each skipped round contributes at most GD to the regret. Specifically, for all $t \in [T]$, $f_t(x_t) - f_t(u) \leqslant f'_t(x_t) - f'_t(u) + GD \cdot \mathbb{I}(t \in Q^*)$. Summing over t yields the first claim. It remains to establish the second claim regarding the skipped set R. We introduce additional notation: let S_t denote the state of the tracking set S at the start of round t. By construction, $d'_t = \sum_{s=1}^T \mathbb{I}(t \in S_s) \leqslant d_t$. Then, $d'_{\text{tot}} = \sum_{t=1}^T d'_t = \sum_{t=1}^T |S_t|$. As $\mathcal{D}_t = \mathcal{D}_{t-1} + |S_t|$, it also holds that $d'_{\text{tot}} = \mathcal{D}_T$. Furthermore, by the preemption rule, round t is not removed before round $t + d'_t - 1$, which implies $d'_t \leqslant \sqrt{\mathcal{D}_{t-1}} + 1 \leqslant \sqrt{d'_{\text{tot}}} + 1$ for all $t \in [T]$.

Part 1 (Bounding $|Q^*|$). We show that $|Q^*| \leqslant 2\sqrt{d'_{\text{tot}}}$. Let $Q^* = \{t_1, \ldots, t_{|Q^*|}\}$ ordered by time of skipping. We claim that $d'_{t_s} \geqslant s/2$ for all $s \in [|Q^*|]$. Indeed, for each $s \in [|Q^*|]$,

$$d'_{t_s} > \sqrt{\mathcal{D}_{t_s + d'_{t_s}}} = \sqrt{\textstyle\sum_{t=1}^T \min\{d'_t, t_s + d'_{t_s} - t\}} \geqslant \sqrt{\textstyle\sum_{k=1}^s d'_{t_k}},$$

which rearranges to $d'_{t_s} \geqslant \frac{1+\sqrt{1+4\sum_{k=1}^{s-1} d'_{t_k}}}{2}$ by solving the quadratic inequality. We proceed by induction. The base case gives $d'_{t_1} \geqslant 1$. Assuming $d'_{t_k} \geqslant k/2$ for all $k < s$, we obtain

$$d'_{t_s} \geqslant \frac{1 + \sqrt{1 + 4\sum_{k=1}^{s-1} \frac{k}{2}}}{2} = \frac{1 + \sqrt{1 + s(s-1)}}{2} \geqslant \frac{s}{2}.$$

Therefore, we have

$$\sqrt{d'_{\text{tot}}} \geqslant \sqrt{\textstyle\sum_{s=1}^{|Q^*|} d'_{t_s}} \geqslant \sqrt{\textstyle\sum_{s=1}^{|Q^*|} \frac{s}{2}} = \sqrt{\frac{|Q^*|(|Q^*|+1)}{4}} \geqslant \frac{|Q^*|}{2}.$$

Part 2 (Bounding d'_{tot}). Fix an arbitrary $Q \subseteq [T]$. Using $d'_t \leqslant \sqrt{d'_{\text{tot}}} + 1$ for all t, we have

$$\textstyle\sum_{t\notin Q} d_t \geqslant \sum_{t\notin Q} d'_t \geqslant \max\left\{0,\ d'_{\text{tot}} - |Q|(\sqrt{d'_{\text{tot}}} + 1)\right\}.$$

Consequently, we can write

$$\begin{aligned}
|Q| + \sqrt{\textstyle\sum_{t\notin Q} d_t} &\geqslant \min_{q\geqslant 0} \left\{q + \sqrt{\max\{0,\ d'_{\text{tot}} - q(\sqrt{d'_{\text{tot}}} + 1)\}}\right\} \\
&\geqslant \min\left\{\sqrt{d'_{\text{tot}}} - 1,\ \min_{q\in[0,\sqrt{d'_{\text{tot}}}-1]} \left\{q + \sqrt{d'_{\text{tot}} - q(\sqrt{d'_{\text{tot}}} + 1)}\right\}\right\} \\
&= \sqrt{d'_{\text{tot}}} - 1.
\end{aligned}$$

Combining both parts and minimizing over $Q \subseteq [T]$, we conclude

$$|Q^*| + \sqrt{\textstyle\sum_{t=1}^T d'_t} \leqslant 2\sqrt{d'_{\text{tot}}} + \sqrt{d'_{\text{tot}}} = 3\sqrt{d'_{\text{tot}}} = O\left(\min_{Q\subseteq[T]} \left\{|Q| + \sqrt{\textstyle\sum_{t\notin Q} d_t}\right\}\right),$$

completing the proof. □

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: The abstract and introduction accurately summarize the paper's contributions, assumptions, and scope. The stated claims are supported by the formal results and proofs in the main paper and appendix.

   Guidelines:

   - The answer [N/A] 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 [N/A] 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: The introduction discusses the main limitations of the work, including the assumptions required for the theoretical guarantees.

   Guidelines:

   - The answer [N/A] 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: Each theoretical result is stated together with the assumptions under which it holds. Complete proofs are provided in the appendix, while the main paper includes the relevant definitions, theorem statements, and intuition where appropriate.

Guidelines:

- The answer [N/A] 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: [N/A]

   Justification: Not applicable. The paper does not report experimental results.

   Guidelines:

   - The answer [N/A] 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: [N/A]

Justification: Not applicable. The paper is theoretical and does not use experimental code or datasets.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.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://neurips.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, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: Not applicable. The paper does not include experiments, training procedures, test sets, or hyperparameter tuning.

   Guidelines:

   - The answer [N/A] 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: [N/A]

   Justification: Not applicable. The paper does not include experimental results for which error bars or statistical significance estimates would be required.

   Guidelines:

   - The answer [N/A] 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: [N/A]

   Justification: Not applicable. The paper does not include experiments requiring computational resources.

   Guidelines:

   - The answer [N/A] 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: The authors have reviewed the NeurIPS Code of Ethics, and the research described in the paper conforms to it.

   Guidelines:

   - The answer [N/A] 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: [N/A]

    Justification: Not applicable. The paper is foundational theoretical work and does not introduce a deployed system, dataset, model, or application with direct societal impact. Any broader impact would be indirect, through future uses of online optimization methods.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.

- If the authors answer [N/A] 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., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Not applicable. The paper does not release data, models, or systems that pose a high risk of misuse.

    Guidelines:

    - The answer [N/A] 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: [N/A]

    Justification: Not applicable. The paper does not use external datasets, codebases, pretrained models, or other existing assets as part of the contribution.

    Guidelines:

    - The answer [N/A] 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: [N/A]

    Justification: Not applicable. The paper does not introduce or release new datasets, code, models, or other assets.

    Guidelines:

    - The answer [N/A] 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: [N/A]

    Justification: Not applicable. The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] 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: [N/A]

    Justification: Not applicable. The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

- The answer [N/A] 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 rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: Not applicable. LLMs were not used as part of the core methodology, theoretical development, or scientific contribution of the paper.

    Guidelines:

    - The answer [N/A] 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 in the NeurIPS handbook for what should or should not be described.