Capacity-Constrained Online Convex Optimization with Delayed Feedback

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

Abstract

Online learning with delayed feedback typically assumes that the learner can track all pending rounds until their feedback arrives. In practice, tracking resources are finite, and feedback from rounds that cannot be tracked is permanently lost. In this paper, we study delayed online convex optimization (OCO) under a hard capacity constraint, where at most C pending rounds can be tracked at any time. To model delay information, we introduce a semi-clairvoyant model that refines the clairvoyant assumption from prior work: rather than requiring delays to be known at prediction time, the learner observes delay expirations online, consistent with the classical unconstrained delayed setting. Our approach proceeds via a reduction to a novel "delayed and weighted" OCO problem, using a scheduler that randomizes which rounds are tracked and importance-weights the resulting observations. For the base problem, we propose and analyze Delayed-Weighted FTRL and its bandit analogue, obtaining regret bounds that characterize how weights interact with delayed feedback in both settings. Combining these bounds with our schedulers yields regret guarantees for capacity-constrained OCO under convex and strongly convex losses, for both first-order and bandit feedback. For first-order feedback, capacity $C = \Omega(\log T)$ suffices to recover the standard delayed OCO rates up to logarithmic factors. For bandit feedback, the standard delayed BCO rates are instead modulated by $(1 + \sigma_{\max}/C)$ factors, where $\sigma_{\max}$ is the maximum number of outstanding observations. This allows the regret bound to degrade gracefully when $C < \sigma_{\max}$. We complement our analysis with numerical experiments.

1 Introduction

Online convex optimization is a canonical framework for sequential decision-making under uncertainty [Hazan, 2016, Orabona, 2019]: a player repeatedly selects an action from a convex domain, incurs a convex loss, and receives either full gradient information (*first-order feedback*) or only the incurred scalar loss (*bandit feedback*). We refer to these two settings as Online Convex Optimization (OCO) and Bandit Convex Optimization (BCO), respectively. A long line of work studies delayed feedback in OCO and BCO, as well as in sequential decision-making more broadly, where the observation from round t arrives only at round $t + d_t$ [Joulani et al., 2013, Quanrud and Khashabi, 2015, Cesa-Bianchi et al., 2016]. The vast majority of this literature, however, takes for granted that the player has unlimited capacity to track pending rounds, that is, every outstanding feedback will eventually be observed, regardless of how many are pending simultaneously.

In many practical settings, only a limited number of rounds can be tracked concurrently. A natural example arises in clinical trials: a study can actively monitor only a bounded number of patients at any given time, and outcomes are observed only for the patients selected for monitoring — side

*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).

effects may still manifest in unmonitored patients, but the trial never observes them. Ryabchenko et al. [2025] formalized this limitation as a hard *capacity constraint*, capping the number of concurrently tracked rounds by a *capacity* parameter C, and studied the resulting problem in the multi-armed bandit setting.

This paper extends the capacity-constrained framework from finite-action settings to convex domains, deriving regret bounds for both OCO and BCO with explicit dependence on the capacity C. The bandit setting is notoriously difficult even without delays or capacity constraints, as the player must optimize over a continuous domain from scalar loss observations alone. Beyond the move to continuous domains, we relax the clairvoyant delay model of prior work, which assumes delays are known at prediction time; our framework instead reveals delays only upon expiration, matching the information naturally available in standard delayed online learning and more closely reflecting the classical clairvoyant model in online job scheduling [e.g., Borodin and El-Yaniv, 1998].

Contributions. Following the reduction approach of Ryabchenko et al. [2025], we decompose capacity-constrained delayed OCO into two components: a *scheduler*, which randomly selects the rounds to track subject to the capacity constraint, and a *base learner*, which receives the resulting feedback stream with importance weights correcting for untracked rounds. Extending this reduction to convex domains introduces two new challenges: first, designing OCO and BCO algorithms that can accommodate feedback that is simultaneously delayed and non-uniformly weighted; and second, controlling the stochastic weight process induced by the scheduler in the presence of convex losses and bandit gradient estimation. Our contributions are as follows:

- *Delayed-weighted OCO and BCO* (Section 3). We introduce OCO and BCO with feedback that is simultaneously delayed and weighted, extending the scale-free online learning framework of Orabona and Pál [2018] to delayed feedback. We analyze a delayed-weighted variant of Follow-the-Regularized-Leader and its bandit analogue, obtaining regret bounds that identify pairwise interactions between weights of rounds with overlapping delays as the structural cost of combining weighting with delays. These results are new for convex domains and may be of independent interest. This framework underlies our reduction for the capacity-constrained setting.
- *Unified proxy-delay scheduling* (Section 4). To handle delays that are unknown at prediction time, Ryabchenko et al. [2025] introduced the idea of sampling *proxy delays* to estimate the true delays. We unify these constructions into a general framework parameterized by proxy-delay distributions, and introduce a new scheduler calibrated to the maximum number of outstanding observations, $\sigma_{\max}$, that yields uniform importance weights, a property crucial for our strongly convex and bandit guarantees. Importantly, the algorithm does not require prior knowledge of $\sigma_{\max}$, as it can be removed via the doubling trick (Appendix F).
- *Capacity-dependent regret bounds* (Section 5). Combining the delayed-weighted regret bounds with the proxy-delay schedulers yields regret guarantees for capacity-constrained OCO and BCO under both convex and strongly convex losses; see Table 1.

  For first-order feedback, capacity $C = \Omega(\log T)$ suffices to match the optimal unconstrained rates in terms of the total delay d_{tot} and the maximum number of outstanding observations $\sigma_{\max}$, up to logarithmic factors, since $\sigma_{\max} \leqslant \sqrt{2d_{\text{tot}}}$. For bandit feedback in dimension k, the capacity constraint modulates the dimension-dependent terms by factors of order $1 + \sigma_{\max}/C$, degrading gracefully as capacity decreases. The delay-dependent terms match the optimal $\sigma_{\max}$ scaling of Ryabchenko et al. [2026] in the strongly convex case; for convex losses they scale as $\sqrt{T\sigma_{\max}}$ rather than $\sqrt{d_{\text{tot}}}$, but still improve over the $\sqrt{Td_{\max}}$ dependence of Wan et al. [2024a], since $\sigma_{\max} \leqslant d_{\max}$ in general. The regime $\sigma_{\max} \ll d_{\max}$ is particularly natural in queueing systems: $\sigma_{\max}$ is governed by the peak arrival rate [e.g., Little, 1961], while $d_{\max}$ is driven by the tail of the service-time distribution, which can produce extreme outliers even when the average service time is moderate. See Appendix H for supporting numerical experiments.

Remark 1.1 (Extensions). Our framework allows the player to *preempt* tracked rounds—that is, to stop tracking them before their feedback arrives. There is also a non-preemptive variant of the problem, in which once a round is tracked, it cannot be dropped until its feedback is delivered. Several of our schedulers are non-preemptive by construction and achieve analogous guarantees in this setting; see Appendix G. Additionally, although our results are stated under the semi-clairvoyant and clairvoyant frameworks, they extend to the fully non-clairvoyant setting, where delays d_t are revealed only at times $t + d_t$ for rounds whose feedback is observed, provided that sufficient information about delay-related quantities such as d_{tot} or $\sigma_{\max}$ is available for tuning the algorithm parameters.

Table 1: Notation: T is the horizon, k is the dimension, $d_{\text{tot}} = \sum_{t=1}^{T} d_t$ is the total delay, $\sigma_{\max} \leqslant \min\{d_{\max}, \sqrt{2d_{\text{tot}}}\}$ is the maximum number of outstanding observations at any round, C is the capacity. Our results are the first ones for the capacity-constrained setting (with $C = \Omega(\ln T)$); all prior work implicitly assumes unlimited capacity ($C = \infty$).

Loss type	**Regret bounds for OCO with delays**		
	Unconstrained ($C = \infty$)		Constrained ($C = \Omega(\ln T)$)
	Quanrud and Khashabi [2015]	Qiu et al. [2025]	This work
Convex	$\sqrt{d_{\text{tot}}} + \sqrt{T}$	N/A	$\sqrt{d_{\text{tot}}} + \sqrt{T}$
Strongly convex	N/A	$\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\} + \ln T$	$(\sigma_{\max} + 1) \ln T$

Loss type	**Regret bounds for BCO with delays (up to log-factors)**		
	Unconstrained ($C = \infty$)		Constrained ($C = \Omega(\ln T)$)
	Wan et al. [2024a]	Ryabchenko et al. [2026]	This work
Convex	$\sqrt{T d_{\max}} + T^{3/4}\sqrt{k}$	$\sqrt{d_{\text{tot}}} + T^{3/4}\sqrt{k}$	$\sqrt{T\sigma_{\max}} + T^{3/4}(1 + \frac{\sigma_{\max}}{C})^{1/4}\sqrt{k}$
Strongly convex	$d_{\max} + T^{2/3}k^{2/3}$	$\sigma_{\max} + T^{2/3}k^{2/3}$	$\sigma_{\max} + T^{2/3}(1 + \frac{\sigma_{\max}}{C})^{1/3}k^{2/3}$

Related work. The capacity-constrained framework was introduced by Ryabchenko et al. [2025] for experts and multi-armed bandits; this work extends it to OCO and BCO. 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], Joulani et al. [2016], Flaspohler et al. [2021] for convex losses and by Qiu et al. [2025] for strongly convex losses; earlier, Wan et al. [2022] achieved the rate $O(d_{\max} \ln T)$. Delayed BCO was initiated by Héliou et al. [2020] and refined by Bistritz et al. [2022], Wan et al. [2024a]. A unified view via reduction to non-delayed OCO appears in Ryabchenko et al. [2026], achieving optimal delay-dependent terms with respect to $\sigma_{\max}$ and d_{tot}. Beyond OCO, delayed feedback has been studied in adversarial bandits [Cesa-Bianchi et al., 2016, Thune et al., 2019, Zimmert and Seldin, 2020], MDPs [Lancewicki et al., 2022, Jin et al., 2022], and combinatorial settings [van der Hoeven et al., 2023]. Additional related work is discussed in Appendix A.

2 Problem Setting and Notation

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 of dimension $k \in \mathbb{N}$, equipped with a norm $\|\cdot\|$ and its dual norm $\|\cdot\|_\star$.

We study a T-round game between a player and an adversary with delayed, capacity-constrained feedback (Figure 1). Before the game begins, the adversary fixes a sequence of convex, differentiable losses $f_t : \mathcal{K} \to \mathbb{R}$ and delays $d_t \in \{0, \ldots, T - t\}$ for $t \in [T]$.[4] At each round t, the player selects $x_t \in \mathcal{K}$ and incurs the (unobserved) loss $f_t(x_t)$. The corresponding feedback—$\nabla f_t(x_t)$ in the *first-order* model or $f_t(x_t)$ in the *bandit* model—arrives only at the end of round $t + d_t$, and is observed if and only if t belongs at that time to a *tracking set* $\mathcal{S} \subseteq [T]$ of maximum size $C \in \mathbb{N}$ (the *capacity*), maintained by the player throughout the game. At the start of each round, the player may *preempt* (i.e., remove from $\mathcal{S}$) any subset of round indices, irrevocably discarding their pending feedback; they may then add the current round's index to $\mathcal{S}$, subject to the capacity constraint. If $C = T$, the learner can track every round and observe all delayed feedback, recovering the standard delayed-feedback setting. The goal is to achieve strong performance while maintaining a tracking set of sublinear size.

We consider two natural frameworks for capacity-constrained delayed feedback, which differ in how the delay sequence is revealed to the player. In the *clairvoyant* framework, d_t is revealed to the player at the start of round t. In the *semi-clairvoyant* framework, the player learns d_t only at round $t + d_t$: if $t \in \mathcal{S}$, the player receives indexed feedback for round t, $(t, \nabla f_t(x_t))$ or $(t, f_t(x_t))$; otherwise, the player receives an empty acknowledgment $(t, \perp)$, from which d_t can be inferred as the elapsed time

[4] The restriction $d_t \leqslant T - t$ is without loss of generality: feedback arriving after round T cannot affect any prediction.

since round t. Semi-clairvoyance thus preserves exactly the delay information available at runtime in standard delayed online learning [Cesa-Bianchi et al., 2016]; the capacity constraint restricts only the *content* of the feedback, not the player's knowledge of when feedback would have arrived.

Capacity-Constrained Online Convex Optimization with Delayed Feedback

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

Player initializes empty tracking set $\mathcal{S}$ of maximum size C.
For each round $t = 1, 2, \dots, T$:

0. If the framework is clairvoyant, then the player observes delay d_t.
1. **Predict:** The player predicts $x_t \in \mathcal{K}$, incurring loss $f_t(x_t)$.
2. **Update tracking set (preempt & add):**
   - The player may preempt (remove from $\mathcal{S}$) any subset of indices, possibly none.
   - The player may add round index t to the tracking set $\mathcal{S}$, provided $|\mathcal{S}| < C$.
3. **Receive feedback:**
   For all s such that $s + d_s = t$:
   - If $s \in \mathcal{S}$, the player observes feedback from round s: $(s, \nabla f_s(x_s))$ in the first-order feedback model, or $(s, f_s(x_s))$ in the bandit feedback model, and removes s from $\mathcal{S}$.
   - If $s \notin \mathcal{S}$ and the framework is semi-clairvoyant, then the player observes $(s, \perp)$, signaling that the delay for round s has expired, i.e., $d_s = t - s$.

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

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 $x^* \in \arg\min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x)$ 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. The first-order guarantees in Theorems 3.1 and 5.1 also use the Euclidean norm.

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 $0 < r < R$.*

Notation. For $n \in \mathbb{N}$, we let $[n] = \{1, \dots, n\}$. For any sequence $(\mathrm{a}_t)_{t=1}^{T}$, indexed over $[T]$, 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 $cS = \{cx : x \in S\}$.

2.1 Notation for Delays and the Tracking Set

The interplay between delays and the capacity constraint requires keeping careful track of which past rounds have produced feedback, which are still pending, and which the player is currently tracking. We introduce this notation and present a few standard identities in delayed online learning.

Observed and backlog sets. At the start of each round $t \in [T]$, the past rounds split into those whose feedback has already arrived and those whose feedback is still pending. We denote these by

$$\mathcal{O}_t \;=\; \{s \in [t-1] : s + d_s < t\}, \qquad \mathcal{B}_t \;=\; \{s \in [t-1] : t \leqslant s + d_s\},$$

respectively; together they partition the past, $\mathcal{O}_t \cup \mathcal{B}_t = [t-1]$ and $\mathcal{O}_t \cap \mathcal{B}_t = \varnothing$.

We write $\sigma_t = |\mathcal{B}_t|$ for the backlog set size (or just *backlog*) at round t, and define

$$\sigma_{\max} \;=\; \max_{t\in[T]} \sigma_t, \qquad d_{\max} \;=\; \max_{t\in[T]} d_t, \qquad d_{\mathrm{tot}} \;=\; \sum_{t=1}^{T} d_t,$$

the maximum backlog, maximum delay, and total delay over the game. The next lemma collects two identities relating these quantities, which we will use repeatedly to bound the cost of the backlog.

Lemma 2.6 (Backlog properties)**.** *For any delays $(d_t)_{t=1}^T$ with $d_t \in \{0,\dots,T-t\}$, the following hold:*

$$\sum_{t=1}^{T} \sigma_t = d_{\mathrm{tot}} \qquad \textit{and} \qquad \sigma_{\max} \leqslant \min\{\sqrt{2d_{\mathrm{tot}}},\, d_{\max}\}.$$

The first identity is the standard average-case quantity governing regret rates in delayed online learning [Cesa-Bianchi et al., 2016]; the second controls the worst-case backlog.

Tracking set evolution. Among the pending rounds, the player tracks only a subset. To reason about which rounds are tracked at each point in time, we consider the post-preemption snapshot

$$\mathcal{S}_t \;=\; \{s \in [T] : s \text{ lies in } \mathcal{S} \text{ immediately after preemption in round } t\}\,.$$

It satisfies $\mathcal{S}_t \subseteq \mathcal{B}_t$, $|\mathcal{S}_t| \leqslant C$, and $\mathcal{S}_{t+1} \subseteq \mathcal{S}_t \cup \{t\}$. The first inclusion holds because every round in $\mathcal{S}$ has been played but has not yet delivered feedback and round t itself has not yet been added at this point. The second is the capacity constraint. The third reflects that between round t and round $t+1$, the set may only gain index t (if added) and may only lose indices (through feedback or preemption).

2.2 Notation for Bandit Feedback and Smoothing

The bandit setting requires constructing gradient estimates from function values alone. We follow the standard single-point approach of Flaxman et al. [2005], which replaces each loss function with a smoothed surrogate whose gradient admits an unbiased estimator from a single bandit query.

Smoothing. Under Assumption 2.5, the inclusion $r\mathbb{B}^k \subseteq \mathcal{K}$ implies $(1-\delta/r)\,\mathcal{K} + \delta\,\mathbb{B}^k \subseteq \mathcal{K}$ for all $\delta \in (0, r]$, and so any point of the shrunken domain $(1-\delta/r)\mathcal{K}$ admits a δ-ball of perturbations that stays in $\mathcal{K}$. For such δ and any integrable $f : \mathcal{K} \to \mathbb{R}$, we define the *$\delta$-smoothing* of f as

$$f^\delta : (1-\delta/r)\mathcal{K} \to \mathbb{R}, \qquad f^\delta(x) \;=\; \mathbb{E}_{v\sim\mathrm{Unif}(\mathbb{B}^k)}\big[f(x+\delta v)\big].$$

The key properties of f^δ are summarized in the following theorem.

Theorem 2.7 (Flaxman et al. [2005])**.** *For any $\delta \in (0, r]$ and continuous $f : \mathcal{K} \to \mathbb{R}$:*

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 first item underpins single-point gradient estimation; the remaining two ensure that f^δ preserves the regularity of f and stays uniformly close to it.

3 Online Convex Optimization with Delayed and Weighted Feedback

This section introduces a weighted variant of OCO with delayed feedback (Figure 2) that serves as the foundation for the reductions in subsequent sections. The game itself involves no tracking set management; the connection to capacity-constrained OCO emerges only through the reductions. Beyond its role in our reductions, this game gives, to our knowledge, the first treatment of *scale-free* online convex optimization [Orabona and Pál, 2018] under delayed feedback.

In this game, the adversary fixes losses $f_t : \mathcal{K} \to \mathbb{R}$, delays $d_t \in \{0,\dots,T-t\}$, and weights $w_t \geqslant 0$ in advance. At each round t, the player predicts $x_t \in \mathcal{K}$ and incurs the weighted loss $w_t f_t(x_t)$. The feedback for round t is revealed at the end of round $t + d_t$, together with the weight w_t.

Online Convex Optimization with Delayed and Weighted Feedback

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

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

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

Figure 2: OCO with weights and delays under first-order or bandit feedback.

We measure performance through the *weighted regret* against a comparator $u \in \mathcal{K}$,

$$R_T^w(u) \;=\; \textstyle\sum_{t=1}^T w_t\big(f_t(x_t) - f_t(u)\big),$$

and its expectation $\bar{R}_T^w(u) = \mathbb{E}[R_T^w(u)]$. We highlight that this problem is *not* equivalent to delayed OCO on the rescaled losses $f_t'(x) = w_t f_t(x)$: since the weights w_t may vary widely, the vacuous rescaling reduction yields bounds scaling with $\max_t w_t$, whereas we seek finer weight dependence — a property central to the reductions developed later.

We now produce algorithms for this game in both first-order and bandit feedback models, following the delayed-update paradigm of Flaspohler et al. [2021]: at each round, the learner applies an update based on all feedback items that arrive in that round, with the modification that each item is now scaled by its weight w_s. Section 3.1 develops this for first-order feedback, yielding a weighted variant of Follow-The-Regularized-Leader (FTRL). Section 3.2 extends the construction to bandit feedback using the single-point estimator of Flaxman et al. [2005] and prediction drift techniques from Ryabchenko et al. [2026].

3.1 First-Order Feedback: DW-FTRL

Given a non-increasing learning-rate sequence $(\eta_t)_{t=1}^T \subset (0, \infty)$, define the *regularization increments* $\alpha_1 = 1/\eta_1$ and $\alpha_t = 1/\eta_t - 1/\eta_{t-1}$. The *delayed-weighted FTRL* (DW-FTRL) updates are given by

$$\text{DW-FTRL:} \quad x_{t+1} = \operatorname*{arg\,min}_{x \in \mathcal{K}} \big\langle x, \textstyle\sum_{s \in \mathcal{O}_{t+1}} w_s \nabla f_s(x_s) \big\rangle + \sum_{s=1}^t \frac{\alpha_s}{2} \|x_s - x\|^2. \tag{1}$$

Each update aggregates the weighted gradients revealed so far and regularizes against the running sequence of iterates. The following theorem bounds the weighted regret of DW-FTRL.

Theorem 3.1. *For all $u \in \mathcal{K}$, predictions $(x_t)_{t=1}^T$ generated by DW-FTRL (1) satisfy*

$$R_T^w(u) \leqslant \tfrac{D^2}{2\eta_T} + G^2 \textstyle\sum_{t=1}^T \eta_{t-1}(w_t^2 + w_t \sum_{s \in \mathcal{B}_t} w_s),$$

where $\eta_0 = \eta_1$. Moreover, under λ-strong convexity (2.3),

$$R_T^w(u) \leqslant \textstyle\sum_{t=1}^T \frac{\alpha_t - w_t\lambda}{2} \|x_t - u\|^2 + G^2 \sum_{t=1}^T \eta_{t-1}(w_t^2 + w_t \sum_{s \in \mathcal{B}_t} w_s).$$

In the delayed FTRL bounds of Theorem 3.1, weighting manifests through pairwise interactions $w_t w_s$ for every pair of rounds t, s such that $s \in \mathcal{B}_t$, i.e., the feedback of s is still pending at t. Without delays the backlog is empty and only the diagonal terms w_t^2 remain, recovering the standard weighted FTRL bounds from scale-free online learning [Orabona and Pál, 2018]. Conversely, setting all weights to $w_t = 1$ recovers standard delayed FTRL bounds [see, e.g., Qiu et al., 2025].

3.2 Bandit Feedback: DW-FTBL

We extend the delayed-weighted updates to bandit feedback through the classical *center-perturbation construction* of Flaxman et al. [2005], applied to a non-increasing sequence of smoothing parameters $(\delta_t)_{t=1}^T \subset (0, r]$. Each round we sample an independent exploration direction $u_t \sim \mathrm{Unif}(\mathbb{S}^{k-1})$ and pick some $\sigma(u_1, \dots, u_{t-1})$-measurable center $y_t \in \mathcal{K}$ to play the perturbed point

$$x_t \;=\; (1 - \delta_t/r)\, y_t + \delta_t\, u_t,$$

which lies in $\mathcal{K}$ for $\delta_t \leqslant r$ due to Assumption 2.5. Letting $\mathcal{F}_t = \sigma(u_1, \dots, u_{t-1})$, the sequence of centers y_t is $\mathcal{F}_t$-measurable while the predictions x_t are $\mathcal{F}_{t+1}$-measurable. By Theorem 2.7, the bandit feedback $f_t(x_t)$ yields the single-point estimator

$$\widehat{g}_t \;=\; \tfrac{k}{\delta_t}\, f_t(x_t)\, u_t, \qquad \mathbb{E}[\widehat{g}_t \mid \mathcal{F}_t] \;=\; \nabla f_t^{\delta_t}\big((1-\delta_t/r)\, y_t\big),$$

unbiased for the gradient of the smoothed loss $f_t^{\delta_t}$ at the shrunken center $(1-\delta_t/r)\, y_t$.

It remains to specify the $\mathcal{F}_t$-measurable centers $(y_t)_{t=1}^T$. Given a non-increasing learning-rate sequence $(\eta_t)_{t=1}^T$ with regularization increments $(\alpha_t)_{t=1}^T$ as in Section 3.1, we define the *delayed-weighted Follow-The-Bandit-Leader* (DW-FTBL) updates by

$$\text{DW-FTBL:} \quad y_{t+1} = \operatorname*{arg\,min}_{y \in \mathcal{K}} \langle y, \textstyle\sum_{s \in \mathcal{O}_{t+1}} w_s\, \widehat{g}_s \rangle + \sum_{s=1}^t \frac{\alpha_s}{2} \|y_s - y\|^2. \tag{2}$$

This is the bandit analog of DW-FTRL (1), with the weighted estimators $w_s\, \widehat{g}_s$ in place of the weighted gradients $w_s\, \nabla f_s(x_s)$. Since each $\widehat{g}_s$ is $\mathcal{F}_{s+1}$-measurable, the resulting centers $(y_t)_{t=1}^T$ are indeed $\mathcal{F}_t$-adapted. The next theorem bounds the expected regret of both updates.

Theorem 3.2. *For all $u \in \mathcal{K}$, predictions $(x_t)_{t=1}^T$ generated by DW-FTBL (2) satisfy*

$$\bar{R}_T^w(u) \leqslant \tfrac{D^2}{2\eta_T} + 2G^2 \textstyle\sum_{t=1}^T \eta_{t-1} \Big(\frac{k^2 M^2 w_t^2}{G^2 \delta_t^2} + \frac{k M w_t}{G \delta_t} \sqrt{\sum_{s \in \mathcal{B}_t} w_s^2} + w_t \sum_{s \in \mathcal{B}_t} w_s \Big) + \tfrac{3GD}{r} \textstyle\sum_{t=1}^T w_t \delta_t,$$

where $\eta_0 = \eta_1$. Moreover, under λ-strong convexity (2.3),

$$\begin{aligned} \bar{R}_T^w(u) \leqslant\; & \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\alpha_t - w_t \lambda}{2} \|y_t - u\|^2 \Big] \\ & + 2G^2 \textstyle\sum_{t=1}^T \eta_{t-1} \Big(\frac{k^2 M^2 w_t^2}{G^2 \delta_t^2} + \frac{k M w_t}{G \delta_t} \sqrt{\sum_{s \in \mathcal{B}_t} w_s^2} + w_t \sum_{s \in \mathcal{B}_t} w_s \Big) + \tfrac{10 G^2}{\lambda r} \textstyle\sum_{t=1}^T w_t \delta_t. \end{aligned}$$

Compared to Theorem 3.1, weighting has a more severe effect under bandit feedback. Beyond the OCO-style interaction $w_t \sum_{s \in \mathcal{B}_t} w_s$, the bound carries an additional cross term $\frac{k M w_t}{G \delta_t} (\sum_{s \in \mathcal{B}_t} w_s^2)^{1/2}$, through which the variance of pending gradient estimators accumulates in ℓ^2 rather than ℓ^1. For uniform weights $w_t \equiv w$, AM-GM absorbs this term into its neighbors, recovering the delayed bandit rate of Ryabchenko et al. [2026] for $w = 1$. With non-uniform weights, the $(\sum_{s \in \mathcal{B}_t} w_s^2)^{1/2}$ factor introduces an ℓ^2 dependence on the pending weights that has no counterpart in the first-order setting.

4 Reduction via Proxy Delay Scheduling

We now reduce capacity-constrained OCO with delayed feedback (Figure 1) to delayed-weighted OCO from the previous section (Figure 2). For concreteness, we present this reduction under the semi-clairvoyant framework (i.e., the delay is not known at prediction time), but it also applies to the clairvoyant and non-clairvoyant frameworks; see Remark 4.2. The reduction is implemented by a wrapper $\mathcal{W}(\mathcal{A}, \mathfrak{D})$ (Algorithm 1) that takes as input an algorithm $\mathcal{A}$ for delayed and weighted OCO and is parameterized by a sequence of *proxy delay distributions* $\mathfrak{D} = (\mathfrak{D}_t)_{t=1}^T$, which govern how the wrapper maintains the tracking set.

Proxy Delay Scheduling. The distributions $\mathfrak{D}$ are used to sample proxy delays $d'_t \sim \mathfrak{D}_t$ on $\mathbb{Z}_{\geqslant -1} \cup \{\infty\}$ at each round t, determining round t's lifetime in the tracking set: round t is added to $\mathcal{S}$ only if $d'_t \geqslant 0$ and $\mathcal{S}$ is not full, and is preempted from $\mathcal{S}$ at round $t + d'_t + 1$ if its feedback has not arrived by then. Feedback from round t is therefore observed if and only if the proxy delay covers the true delay (i.e., $d'_t \geqslant d_t$) and the tracking set was not full when round t was considered for admission (i.e., $|\mathcal{S}_t| < C$). The choice of $\mathfrak{D}$ navigates a fundamental tradeoff: heavier-tailed distributions are more likely to cover true delays but keep rounds in $\mathcal{S}$ longer, risking saturation.

Algorithm Overview. The wrapper treats a base algorithm $\mathcal{A}$ for OCO with delayed and weighted feedback as a black box and feeds it a simulated feedback stream constructed from the observations available in the original capacity-constrained problem. Using $\mathfrak{D}$, the weights are set as importance-weighted indicators of whether the corresponding feedback is observed.

Algorithm 1: Semi-Clairvoyant Proxy-Delay Importance-Weighted Wrapper $\mathcal{W}(\mathcal{A}, \mathfrak{D})$

Access : Algorithm $\mathcal{A}$ for OCO with delayed and weighted feedback (Figure 2).
Inputs : Proxy delay distributions $\mathfrak{D} = (\mathfrak{D}_t)_{t=1}^T \subset \Delta(\mathbb{Z}_{\geqslant -1} \cup \{\infty\})$.

Initialize tracking set $\mathcal{S} = \varnothing$ of maximum size C.
For round $t = 1, 2, \ldots, T$:

1. **Predict:** Start round t for $\mathcal{A}$ and output its prediction as x_t.
2. **Update tracking set (preempt & add):**
   - Preempt expired tracking: For all $\tilde{s} \in \mathcal{S}$ such that $\tilde{s} + d'_{\tilde{s}} < t$, remove $\tilde{s}$ from $\mathcal{S}$.
   - Add current round: Sample proxy delay $d'_t \sim \mathfrak{D}_t$. Add t to $\mathcal{S}$ if $|\mathcal{S}| < C$ and $d'_t \geqslant 0$.
3. **Receive feedback:**
   For all s such that $s + d_s = t$:
   - **If** $s \in \mathcal{S}$ (i.e., timeout sufficient: $d'_s \geqslant d_s$):
     Receive (s, v_s); set $w_s = 1/\mathbb{P}(d'_s \geqslant d_s)$; forward (s, w_s, v_s) to $\mathcal{A}$; remove s from $\mathcal{S}$.
   - **If** $s \notin \mathcal{S}$ (i.e., timeout expired early: $d'_s < d_s$, or skipped due to capacity C):
     Receive $(s, \perp)$; set $w_s = 0$; forward $(s, 0, 0)$ to $\mathcal{A}$.

Every round t, the wrapper outputs $\mathcal{A}$'s prediction x_t as its own. At round $s + d_s$, if s is still in $\mathcal{S}$ the wrapper forwards (s, w_s, v_s) to $\mathcal{A}$ with importance weight $w_s = 1/\mathbb{P}(d'_s \geqslant d_s)$ and observation $v_s \in \{\nabla f_s(x_s),\ f_s(x_s)\}$; otherwise it forwards the zero placeholder $(s, 0, 0)$. The following theorem formalizes the properties of the random weights and bounds the wrapper's regret by the weighted regret of $\mathcal{A}$ plus a penalty for rounds in which the tracking set $\mathcal{S}$ is full.

Theorem 4.1 (Regret of Algorithm 1)**.** *Suppose the proxy delays are sampled independently and* $\mathbb{P}(d'_t \geqslant d_t) > 0$ *for every* $t \in [T]$*. The weights forwarded to* $\mathcal{A}$ *satisfy*

$$w_t = \frac{\mathbb{I}(|\mathcal{S}_t| < C)\,\mathbb{I}(d'_t \geqslant d_t)}{\mathbb{P}(d'_t \geqslant d_t)}, \quad \mathbb{E}[w_t] \leqslant 1, \quad \mathbb{E}[w_t^2] \leqslant \frac{1}{\mathbb{P}(d'_t \geqslant d_t)}, \quad \mathbb{E}[w_t w_s] \leqslant 1,$$

for all distinct $t, s \in [T]$*. Moreover, the expected regret satisfies*

$$\bar{R}_T(u) \ \leqslant\ \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t\big(f_t(x_t) - f_t(u)\big)\Big] + GD\ \textstyle\sum_{t=1}^T \mathbb{P}\big(|\mathcal{S}_t| = C\big).$$

Theorem 4.1 reduces the analysis of capacity-constrained OCO with delayed feedback to bounding two quantities: the weighted regret of $\mathcal{A}$ (Theorem 3.1 or 3.2) and the saturation probabilities $\mathbb{P}(|\mathcal{S}_t| = C)$. Under the oblivious-adversary assumption, the tracking sets and weights are determined entirely by the fixed delay sequence and the scheduler's random draws. Conditioning on these draws fixes the weights while preserving the base learner's randomization. Theorems 3.1 and 3.2 therefore apply conditionally, and averaging over scheduler randomness yields bounds on the expected weighted regret. The choice of proxy-delay distributions $\mathfrak{D}_t$ controls the tradeoff between the weighted-regret bound and the saturation probabilities. The next section presents concrete choices of proxy-delay distributions.

Remark 4.2 (Applicability across frameworks). Beyond the semi-clairvoyant framework, this reduction immediately extends to the clairvoyant framework, since the player can simulate the semi-clairvoyant empty acknowledgments from known delays. It also extends to the non-clairvoyant framework provided that $\mathcal{A}$ ignores zero-weight observations, i.e., its predictions are unchanged when a zero-weight item is forwarded; the wrapper can then simply skip the forwarding step for untracked rounds. The algorithms of Section 3 satisfy this property, since a zero weight zeroes out the corresponding gradient term in the update.

4.1 Concrete Proxy distributions

We now give concrete choices of proxy distributions $\mathfrak{D}$. Lemma 4.3 states a generic sufficient condition under which the tracking set saturates with probability at most $\exp(-C)$, and the two corollaries instantiate this condition for specific families of $\mathfrak{D}_t$. The first scheduler, an adaptation of the Pareto-based construction of Ryabchenko et al. [2025], sets the per-round observation probability p_t inversely proportional to the delay d_t. The second is new to this work: it sets a uniform per-round observation probability p inversely proportional to the maximum backlog $\sigma_{\max} = \max_t |\mathcal{B}_t|$ (Section 2.1), trading adaptivity for a constant weight scale $1/p$ across rounds, which is essential for the bounds we obtain in the strongly convex and bandit settings. This scheduler requires

knowledge of $\sigma_{\max}$; however, the resulting regret bounds in Section 5 can nevertheless be extended to unknown $\sigma_{\max}$ via a novel doubling trick in Appendix F.

Lemma 4.3. *Suppose the proxy delay distributions $\mathfrak{D} = (\mathfrak{D}_t)_{t=1}^T$ from Algorithm 1 satisfy*

$$\textstyle\sum_{s\in\mathcal{B}_t\cup\{t\}} \mathbb{P}(d'_s \geqslant t-s) \;\leqslant\; \frac{C}{8} \qquad \text{for every } t\in[T].$$

Then $\mathbb{P}\big(|\mathcal{S}_t| = C\big) \leqslant \exp(-C)$ for every $t \in [T]$.

The bound follows immediately from a Chernoff inequality on the size of $\mathcal{S}_t$.

Corollary 4.4 (Pareto scheduler)**.** *For each $t \in [T]$, let $H_t = \sum_{j=1}^t \frac{1}{j}$. Take $\mathfrak{D}_t$ to be the shifted floored Pareto distribution* [5]

$$\mathfrak{D}_t \;=\; \Big\lfloor \mathrm{Pareto}\big(\tfrac{C}{16H_t},\, 1\big) \Big\rfloor - 1.$$

Then $\mathbb{P}(d'_t \geqslant d_t) = \min\big\{1,\, \frac{C}{16H_t(d_t+1)}\big\}$ and $\mathbb{P}(|\mathcal{S}_t| = C) \leqslant \exp(-C)$ for every $t \in [T]$.

Corollary 4.5 (Bernoulli scheduler)**.** *Suppose $\sigma_{\max}$ is known. Let $p = \min\big\{1,\, \frac{C}{8(\sigma_{\max}+1)}\big\}$, and for every $t \in [T]$ take $\mathfrak{D}_t$ to be the two-point distribution on $\{-1, \infty\}$ with*

$$\mathbb{P}_{d'\sim\mathfrak{D}_t}(d' = \infty) \;=\; p, \qquad \mathbb{P}_{d'\sim\mathfrak{D}_t}(d' = -1) \;=\; 1-p.$$

Then $\mathbb{P}(d'_t \geqslant d_t) = p$ and $\mathbb{P}(|\mathcal{S}_t| = C) \leqslant \exp(-C)$ for every $t \in [T]$.

The two corollaries sit at opposite ends of a tradeoff. Corollary 4.4 provides $p_t = \tilde{O}(\frac{C}{d_t+1})$, adapting to each round's individual delay, while Corollary 4.5 achieves the uniform $p = O(\frac{C}{\sigma_{\max}+1})$. Each is useful in different regimes, as the next section illustrates.

5 Regret Bounds for Capacity-Constrained OCO with delayed feedback

Combining the weighted regret bounds of Section 3 with the proxy-delay wrapper of Section 4 yields regret guarantees for capacity-constrained OCO and BCO under delayed feedback. The two theorems below cover the convex and strongly convex cases under first-order and bandit feedback, respectively, and constitute the main results of this work. The mild restriction $C \geqslant \ln T + 1$ ensures that the saturation probabilities of Lemma 4.3 sum to a constant over the horizon. The learning rates η_t and smoothing parameters δ_t depend on problem quantities that may not be known in advance: either d_{tot} and T, or $\sigma_{\max}$. The former dependence can be removed by the two-dimensional doubling trick of Bistritz et al. [2022], and the latter by a new doubling trick developed in Appendix F.

Theorem 5.1 (Capacity-Constrained OCO with Delays)**.** *Suppose $C \geqslant \ln T + 1$, and consider $\mathcal{W}(\mathcal{A}, \mathfrak{D})$ (Algorithm 1), where $\mathcal{A}$ is DW-FTRL (1). With learning rates $\eta_t = \frac{D/G}{\sqrt{d_{\text{tot}}+T}}$ and $\mathfrak{D}$ from Corollary 4.4, $\mathcal{W}(\mathcal{A}, \mathfrak{D})$ guarantees*

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

Under λ-strong convexity (2.3), for rates $\eta_t = \frac{1}{t\lambda}$ and $\mathfrak{D}$ from Corollary 4.5, $\mathcal{W}(\mathcal{A}, \mathfrak{D})$ guarantees

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

Small capacities. The logarithmic-capacity requirement arises from controlling the saturation penalty $GDTe^{-C}$ and is not claimed to be necessary. For a known fixed delay d, the block-sampling construction of Ryabchenko et al. [2025, Section 3] can be adapted to OCO as follows. Partition the horizon into blocks of $d+1$ rounds, play one action throughout each block, and track one uniformly sampled round per block. The sampled gradient estimates the average block-loss gradient without bias and arrives by the end of the following block. Two OCO learners on alternating blocks therefore achieve expected regret $O(GD\sqrt{T(d+1)})$ with capacity $C = 2$. Whether constant capacity suffices to recover the total-delay guarantee for arbitrary delay sequences remains open.

To state the BCO bounds, we introduce the dimensionless ratio $\nu = \frac{M}{Gr}$, where M is the absolute value bound (Assumption 2.4).

[5] A Pareto(x_m, α) random variable has tail CDF $\mathbb{P}(X \geqslant x) = \min\{1,\, (x_m/x)^\alpha\}$ for $x > 0$.

Theorem 5.2 (Capacity-Constrained BCO with Delays)**.** *Suppose* $C \geqslant \ln T + 1$*, and consider* $\mathcal{W}(\mathcal{A}, \mathfrak{D})$*, where* $\mathcal{A}$ *is DW-FTBL* (2) *and* $\mathfrak{D}$ *is from Corollary 4.5. With smoothing parameters* $\delta_t = r \min\left\{1, \frac{\sqrt{\nu k}}{t^{1/4}}\left(1 + \frac{\sigma_{\max}}{C}\right)^{1/4}\right\}$ *and learning rates* $\eta_t = \frac{D/G}{t^{1/2}\sqrt{\sigma_{\max} + (\frac{\nu k}{\delta_t/r})^2(1+\frac{\sigma_{\max}}{C})}}$*,* $\mathcal{W}(\mathcal{A}, \mathfrak{D})$ *guarantees*

$$\bar{R}_T(x^*) = O\Big(GD\left(\sqrt{T\sigma_{\max}} + T^{3/4}\left(1 + \tfrac{\sigma_{\max}}{C}\right)^{1/4}\sqrt{\nu k}\right)\Big).$$

Under λ*-strong convexity (2.3), with* $\eta_t = \frac{1}{t\lambda}$ *and* $\delta_t = r \min\left\{1, \left(\frac{\nu^2 k^2 \ln(t+1)}{t}\right)^{1/3}\left(1 + \frac{\sigma_{\max}}{C}\right)^{1/3}\right\}$*,*

$$\bar{R}_T(x^*) = O\Big(\tfrac{G^2}{\lambda}\left(\sigma_{\max}\ln T + (T^2 \ln T)^{1/3}\left(1 + \tfrac{\sigma_{\max}}{C}\right)^{1/3}(\nu k)^{2/3}\right)\Big).$$

Under the uniform-weight scheduler, importance weighting contributes a factor $\mathbb{E}[w_t^2] = O\big(1 + \frac{\sigma_{\max}}{C}\big)$ to the gradient estimator's second-moment bound. Balancing this contribution against smoothing error yields the capacity exponents $\frac{1}{4}$ and $\frac{1}{3}$ for convex and strongly convex losses, respectively.

Both theorems follow from Theorem 4.1 by combining the weighted regret bounds of Theorems 3.1 and 3.2 with the saturation bounds of Corollaries 4.4 and 4.5; full proofs are deferred to Appendix E.

6 Discussion and Future Work

This work gives the first regret bounds for capacity-constrained OCO, in both first-order and bandit settings under convex and strongly convex losses. For first-order OCO, logarithmic capacity suffices to match unconstrained delayed rates; for BCO, capacity enters regret through $(1 + \sigma_{\max}/C)$ factors.

Several directions remain open. First, our BCO bounds for convex losses scale as $\sqrt{T\sigma_{\max}}$ in the delay-dependent term rather than the optimal $\sqrt{d_{\text{tot}}}$ of the unconstrained setting; closing this gap or establishing a separation is an open problem. Second, the capacity-constrained framework bears a structural resemblance to finite-buffer queueing systems, where jobs that find the buffer full are dropped. Our schedulers exploit this connection, but the precise relationship between queueing theory and online learning under capacity constraints warrants deeper exploration. Third, extending from oblivious to *adaptive* adversaries, who select losses and delays based on the player's history, appears to require fundamentally new tools and is a compelling direction for future work.

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*, 2008.

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *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.

Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, pages 20–29, 1996.

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.

Allan Borodin and Ran El-Yaniv. *Online Computation and Competitive Analysis*. Cambridge University Press, Cambridge, 1998.

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

Graham Cormode and Shan Muthukrishnan. An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithms*, 55(1):58–75, 2005.

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*, volume 202 of *Proceedings of Machine Learning Research*, pages 9374–9395. PMLR, 2023.

Philippe Flajolet, Éric Fusy, Olivier Gandouet, and Frédéric Meunier. HyperLogLog: the analysis of a near-optimal cardinality estimation algorithm. In *2007 Conference on Analysis of Algorithms (AofA 07)*, volume AH of *DMTCS Proceedings*, pages 137–156, 2007.

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 Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2005.

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):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, Andras Gyorgy, and Csaba Szepesvári. Online learning under delayed feedback. In *International Conference on Machine Learning*, pages 1453–1461. PMLR, 2013.

Pooria Joulani, Andras Gyorgy, 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.

Orin Levy, Liad Erez, Alon Peled-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, pages 35742–35773, 2025.

Richard J. Lipton and Andrew Tomkins. Online interval scheduling. In *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '94, pages 302–311, USA, 1994. Society for Industrial and Applied Mathematics.

John D. C. Little. A proof for the queuing formula: $L = \lambda W$. *Operations Research*, 9(3):383–387, 1961.

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.

S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends® in Theoretical Computer Science*, 1(2):117–236, 2005.

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

Francesco Orabona and Dávid Pál. Scale-free online learning. *Theoretical Computer Science*, 716: 50–69, 2018. Preliminary version in ALT 2015.

Michael L. Pinedo. *Scheduling: Theory, Algorithms, and Systems*. Springer, Cham, 6 edition, 2022.

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*, volume 267 of *Proceedings of Machine Learning Research*, pages 50448–50479. PMLR, 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. A reduction from delayed to immediate feedback for online convex optimization with improved guarantees. In *Advances in Neural Information Processing Systems*, 2026. To appear.

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.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2 edition, 2026.

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.

Gerhard J. Woeginger. On-line scheduling of jobs with fixed start and end times. *Theoretical Computer Science*, 130(1):5–16, 1994.

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*, pages 928–936, 2003.

A Additional Related Work

Comparison to Ryabchenko et al. [2025]. The capacity-constrained framework was recently introduced by Ryabchenko et al. [2025] for finite-action settings, specifically prediction with expert advice and multi-armed bandits. To manage the tracking set under the capacity limit, they developed a modular approach that decouples the observation scheduling policy from the base learning algorithm, proposing techniques such as batch partitioning and Pareto-distributed proxy delays. While we adopt their high-level separation of schedulers and base learners, extending this framework to continuous convex domains (OCO and BCO) introduces fundamentally new theoretical challenges. In particular, the proxy-delay scheduler of Ryabchenko et al. [2025] inherently induces non-uniform observation probabilities; while discrete-action algorithms can absorb these via time-varying loss scales, such variable importance weights are highly detrimental to continuous bandit gradient estimation and strongly convex objectives. To address this, we introduce the delayed-weighted OCO and BCO framework, and design a novel unified proxy-delay scheduler calibrated to the maximum number of outstanding observations $\sigma_{\max}$ that yields the uniform importance weights necessary for our continuous-domain bounds. Furthermore, their preemptive scheduler requires prior knowledge of the maximum delay $d_{\max}$ in the non-clairvoyant setting. In contrast, our approach accommodates a semi-clairvoyant model and removes the need for prior knowledge of delay bounds via the doubling trick.

Connections to Online Scheduling. Our model is closely related to online job and interval scheduling [Lipton and Tomkins, 1994, Woeginger, 1994, Borodin and El-Yaniv, 1998, Pinedo, 2022]. In online interval scheduling, jobs arrive sequentially with release times and processing intervals, and the algorithm must assign jobs to a bounded number of machines while respecting overlap constraints. Under the correspondence with our setting, rounds become jobs, tracking capacity corresponds to the number of available machines, and delays determine the durations for which machines remain occupied. Tracking a round occupies one unit of capacity until feedback arrives or the round is preempted. The quantity $\sigma_{\max}$ corresponds to the peak concurrency level of overlapping jobs, whereas $d_{\max}$ reflects the longest processing duration and may therefore be much larger. This perspective helps explain regimes in which $\sigma_{\max} \ll d_{\max}$, which naturally arise in queueing systems with moderate average load but heavy-tailed service times. The distinction between preemptive and non-preemptive schedulers in our framework also parallels classical scheduling models.

Connections to Streaming Algorithms. Our framework also bears conceptual similarities to streaming algorithms [Alon et al., 1996, Muthukrishnan, 2005]. Streaming algorithms process massive data streams under strict memory constraints, typically maintaining compact randomized summaries or sketches rather than storing the entire stream. Analogously, in our setting the learner cannot retain all pending observations and must selectively track only a bounded subset of rounds. Both settings therefore rely on randomized sampling and compressed representations to preserve global performance guarantees despite limited resources. Classic examples include the AMS sketch [Alon et al., 1996], Count-Min Sketch [Cormode and Muthukrishnan, 2005], and HyperLogLog [Flajolet et al., 2007].

Online convex optimization. Online convex optimization (OCO) and its feedback models have been studied extensively. Early work by Zinkevich [2003] introduced online convex programming and established $O(\sqrt{T})$ regret bounds for online gradient methods. For first-order feedback, subsequent research 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] provided $O(\ln T)$ guarantees, which match the $\Omega(\ln T)$ lower bound of Abernethy et al. [2008]. Comprehensive background and further developments can be found in standard texts [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 a 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 the theoretical guarantees. For two-point feedback, they achieved $\widetilde{O}(T^{1/2}k^2)$ regret for convex losses and $O(k^2 \ln T)$ regret under strong convexity. For single-point feedback, they also demonstrated improved rates under additional curvature, including a $T^{2/3}$-type

guarantee for strongly convex losses (specifically, $O((T^2 \ln T)^{1/3} k^{2/3})$ up to problem-dependent constants). Complementary results for stochastic convex optimization under bandit feedback were established by Agarwal et al. [2011]. For two-point feedback, Shamir [2017] later introduced an optimal algorithm with an improved dimension dependence of $O(T^{1/2}k^{1/2})$. A parallel line of work designs (nearly) minimax-optimal algorithms for general BCO that approach the $\Omega(k\sqrt{T})$ lower bound of Shamir [2013] [see, e.g., Bubeck and Eldan, 2016, Hazan and Li, 2016, Lattimore, 2019]. Because these methods typically incur prohibitive computational costs, exhibiting exponential or high-degree polynomial dependence on the dimension k and/or the horizon T, we refer the reader to Lattimore [2026] for a detailed treatment and further references.

Online convex optimization with delayed feedback. Early investigations into online learning with delayed feedback date back to Weinberger and Ordentlich [2002]. A systematic treatment of delayed-feedback online learning and reduction-style approaches was developed by Joulani et al. [2013], while Joulani et al. [2016] introduced a black-box reduction that processes feedback strictly in the order of its arrival. Beyond generic black-box techniques, a substantial body of literature focuses on algorithms explicitly 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 similarly received dedicated attention, beginning with delayed extensions of bandit gradient methods [Héliou et al., 2020] and advancing toward delay-adaptive bounds with improved delay dependence [Bistritz et al., 2022, Wan et al., 2024a]. Recently, Qiu et al. [2026] reduced delayed feedback to online learning with time-varying movement costs in the dynamic regret setting.

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]. Such delays emerge naturally under resource limitations, including the capacity constraints we study [Ryabchenko et al., 2025]. Delayed feedback has additionally been explored in adversarial contextual bandits with general function approximation [Levy et al., 2025], as well as in reinforcement learning and Markov decision processes (MDPs) operating 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 Technical Preliminaries

Here, we collect auxiliary facts and standard results used throughout the proofs.

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 (Ryabchenko et al. [2025], Fact F.2)**.** *Let $t \in \mathbb{N}$. For the sequence of harmonic numbers $(H_s)_{s=1}^t = (\sum_{s'=1}^s 1/s')_{s=1}^t$, it holds that $\sum_{s=1}^t \frac{1}{2H_s(t-s+1)} \leqslant 1$.*

Proof. For every $t \in \mathbb{N}$,

$$\begin{aligned}
\sum_{s=1}^t \frac{1}{2H_s(t-s+1)} &= \sum_{s=1}^{\lceil t/2\rceil} \frac{1}{2H_s(t-s+1)} + \sum_{s=\lceil t/2\rceil+1}^t \frac{1}{2H_s(t-s+1)} \\
&\leqslant \sum_{s=1}^{\lceil t/2\rceil} \frac{1}{2\lceil t/2\rceil} + \sum_{s=\lceil t/2\rceil+1}^t \frac{1}{2H_{\lceil t/2\rceil}(t-s+1)} \\
&\leqslant \frac{\lceil t/2\rceil}{2\lceil t/2\rceil} + \frac{H_{\lceil t/2\rceil}}{2H_{\lceil t/2\rceil}} \\
&= 1.
\end{aligned}$$

□

The following results are standard in the analysis of FTRL.

Theorem B.3 (Orabona [2019], adapted from Lemmas 7.1 and 7.5)**.** *Let $\mathcal{K} \subset \mathbb{R}^k$ be a convex compact set, let $(\eta_n)_{n=0}^N \subset (0,\infty)$ be a non-increasing sequence, and let $\psi_0, \dots, \psi_N : \mathbb{R}^k \to \mathbb{R}$ be differentiable functions such that each ψ_n is $(1/\eta_n)$-strongly convex with respect to $\|\cdot\|$ on $\mathcal{K}$. Assume $z_1 \in \arg\min_{z\in\mathcal{K}} \psi_0(z)$ and $\psi_0(z_1) = 0$. Given loss vectors $(c_n)_{n=1}^N \subset \mathbb{R}^k$, define $z_{n+1} = \arg\min_{z\in\mathcal{K}} \{\psi_n(z) + \sum_{m=1}^n \langle c_m, z\rangle\}$ for $n = 1, \dots, N$. Then the minimizers z_{n+1} are unique. Moreover, for all $u \in \mathcal{K}$,*

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

Lemma B.4 (Qiu et al. [2025], Lemma A.2 (modified))**.** *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 $z_i \in \arg\min_{x\in\mathcal{K}} \langle w_i, x\rangle + \phi_i(x)$. 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

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

□

C OCO with Delayed and Weighted Feedback: Proofs

We prove the regret bounds for DW-FTRL (Theorem 3.1) and DW-FTBL (Theorem 3.2) stated in Section 3. Both proofs follow the delayed-update template: we compare the actual iterates against a fictitious sequence that has access to all gradients without delay, and bound the drift between them using the stability lemma (Lemma B.4). In the bandit case, we control the variance of pending gradient estimators using a technique of Ryabchenko et al. [2026] (Lemma C.1).

Theorem 3.1 (Restated)**.** *For all $u \in \mathcal{K}$, predictions $(x_t)_{t=1}^T$ generated by DW-FTRL (1) satisfy*

$$R_T^w(u) \leqslant \tfrac{D^2}{2\eta_T} + G^2 \sum_{t=1}^T \eta_{t-1}(w_t^2 + w_t \textstyle\sum_{s\in\mathcal{B}_t} w_s),$$

where $\eta_0 = \eta_1$. Moreover, under λ-strong convexity (2.3),

$$R_T^w(u) \leqslant \sum_{t=1}^T \tfrac{\alpha_t - w_t\lambda}{2}\|x_t - u\|^2 + G^2 \sum_{t=1}^T \eta_{t-1}(w_t^2 + w_t \textstyle\sum_{s\in\mathcal{B}_t} w_s).$$

Proof. Consider a sequence $(\widetilde{x}_t)_{t=1}^T \subset \mathcal{K}$, defined by $\widetilde{x}_1 = x_1$ and

$$\widetilde{x}_{t+1} = \arg\min_{x\in\mathcal{K}} \langle x, \textstyle\sum_{s\in[t]} w_s \nabla f_s(x_s)\rangle + \sum_{s=1}^t \tfrac{\alpha_s}{2}\|x_s - x\|^2.$$

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

$$\begin{aligned}\sum_{t=1}^T w_t(f_t(x_t) - f_t(u)) &\leqslant \sum_{t=1}^T \langle w_t \nabla f_t(x_t), x_t - u\rangle - \tfrac{\lambda}{2}\sum_{t=1}^T w_t\|x_t - u\|^2 \\ &\leqslant \sum_{t=1}^T \langle w_t \nabla f_t(x_t), \widetilde{x}_t - u\rangle - \tfrac{\lambda}{2}\sum_{t=1}^T w_t\|x_t - u\|^2 \\ &\quad + G\textstyle\sum_{t=1}^T w_t\|x_t - \widetilde{x}_t\|,\end{aligned}$$

where the final inequality follows from Assumption 2.2 and the triangle inequality.

To bound the first term, apply Theorem B.3 to $(\widetilde{x}_t)_{t=1}^T$ with $\psi_t(x) = \sum_{s=1}^t \frac{\alpha_s}{2}\|x_s - x\|^2$ and $\psi_0(x) = \frac{1/\eta_1}{2}\|x_1 - x\|^2$ so that $x_1 = \arg\min_{x\in\mathcal{K}} \psi_0(x)$. Note that ψ_t are $(1/\eta_t)$-strongly convex while ψ_0 is $(1/\eta_1)$-strongly convex. Hence, it holds that

$$\sum_{t=1}^T \langle w_t \nabla f_t(x_t), \widetilde{x}_t - u\rangle \leqslant \sum_{t=1}^T \tfrac{\alpha_t}{2}\|x_t - u\|^2 + \sum_{t=1}^T \tfrac{\eta_{t-1} w_t^2 \|\nabla f_t(x_t)\|^2}{2}.$$

To control $\|x_t - \widetilde{x}_t\|$, apply Lemma B.4 (with appropriate $\phi_0 = \phi_1 = \psi_{t-1}$),

$$\|x_t - \widetilde{x}_t\| \leqslant \eta_{t-1} \left\| \sum_{s\in\mathcal{B}_t} w_s \nabla f_s(x_s) \right\|.$$

Combining these results and Assumption 2.2, we can write

$$R_T^w(u) \leqslant \sum_{t=1}^T \tfrac{\alpha_t - w_t\lambda}{2} \|x_t - u\|^2 + G^2 \sum_{t=1}^T \eta_{t-1}(w_t^2 + w_t \textstyle\sum_{s\in\mathcal{B}_t} w_s).$$

By substituting $\lambda = 0$ or $\lambda > 0$, we obtain the claimed results in both cases. □

Theorem 3.2 (Restated). *For all $u \in \mathcal{K}$, predictions $(x_t)_{t=1}^T$ generated by DW-FTBL (2) satisfy*

$$\bar{R}_T^w(u) \leqslant \tfrac{D^2}{2\eta_T} + 2G^2 \textstyle\sum_{t=1}^T \eta_{t-1} \Big(\frac{k^2M^2w_t^2}{G^2\delta_t^2} + \frac{kMw_t}{G\delta_t} \sqrt{\sum_{s\in\mathcal{B}_t} w_s^2} + w_t \sum_{s\in\mathcal{B}_t} w_s \Big) + \tfrac{3GD}{r} \sum_{t=1}^T w_t\delta_t,$$

where $\eta_0 = \eta_1$. Moreover, under λ-strong convexity (2.3),

$$\begin{aligned}\bar{R}_T^w(u) &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\alpha_t - w_t\lambda}{2} \|y_t - u\|^2 \Big] \\ &\quad + 2G^2 \textstyle\sum_{t=1}^T \eta_{t-1} \Big(\frac{k^2M^2w_t^2}{G^2\delta_t^2} + \frac{kMw_t}{G\delta_t} \sqrt{\sum_{s\in\mathcal{B}_t} w_s^2} + w_t \sum_{s\in\mathcal{B}_t} w_s \Big) + \tfrac{10G^2}{\lambda r} \sum_{t=1}^T w_t\delta_t.\end{aligned}$$

Proof. Consider a sequence $(\widetilde{y}_t)_{t=1}^T$, defined by $\widetilde{y}_1 = y_1$ and

$$\widetilde{y}_{t+1} = \operatorname*{arg\,min}_{y\in\mathcal{K}} \langle y, \textstyle\sum_{s\in[t]} w_s \widehat{g}_s \rangle + \sum_{s=1}^t \frac{\alpha_s}{2} \|y_s - y\|^2.$$

By induction, it follows simultaneously that, for every $t \in [T]$, (1) y_t is $\mathcal{F}_t$-measurable, (2) $\widetilde{y}_t$ is $\mathcal{F}_t$-measurable, and (3) x_t and $\widehat{g}_t$ are $\mathcal{F}_{t+1}$-measurable.

Fix comparator $v \in \mathcal{K}$. For $t \in [T]$, let $v_t^\delta = (1 - \delta_t/r)v$ which satisfies $\|v_t^\delta - v\| = \frac{\delta_t\|v\|}{r}$. We treat the general convex case as 0-strongly convex. For $\lambda \geqslant 0$, the λ-strong convexity of f_t gives

$$\begin{aligned}\bar{R}_T^w(v) &= \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t(x_t) - f_t(v)) \Big] \\ &= \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t^\delta(y_t^\delta) - f_t^\delta(v_t^\delta)) \Big] \\ &\quad + \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t(x_t) - f_t(y_t^\delta)) \Big] + \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t(y_t^\delta) - f_t^\delta(y_t^\delta)) \Big] \\ &\quad + \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t^\delta(v_t^\delta) - f_t(v_t^\delta)) \Big] + \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t(v_t^\delta) - f_t(v)) \Big] \\ &\overset{(a)}{\leqslant} \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t^\delta(y_t^\delta) - f_t^\delta(v_t^\delta)) \Big] + \textstyle\sum_{t=1}^T w_t \Big(G\delta_t + G\delta_t + G\delta_t + \frac{G\|v\|}{r}\delta_t \Big) \\ &\overset{(b)}{\leqslant} \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t \big(\langle \nabla f_t^\delta(y_t^\delta),\, y_t^\delta - v_t^\delta \rangle - \frac{\lambda}{2}\|y_t^\delta - v_t^\delta\|^2 \big) \Big] + \frac{2GD}{r} \sum_{t=1}^T w_t\delta_t \\ &\overset{(c)}{=} \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t \big((1 - \frac{\delta_t}{r}) \langle \nabla f_t^\delta(y_t^\delta),\, y_t - v \rangle - \frac{\lambda}{2}(1 - \frac{\delta_t}{r})^2 \|y_t - v\|^2 \big) \Big] + \frac{2GD}{r} \sum_{t=1}^T w_t\delta_t \\ &= \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t \big(\langle \nabla f_t^\delta(y_t^\delta),\, \widetilde{y}_t - v \rangle - \frac{\lambda}{2}\|y_t - v\|^2 \big) \Big] \\ &\quad + \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t \big(\langle \nabla f_t^\delta(y_t^\delta),\, y_t - \widetilde{y}_t \rangle \big) \Big] \\ &\quad + \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t \big(\frac{\delta_t}{r} \langle \nabla f_t^\delta(y_t^\delta),\, v - y_t \rangle + (\frac{\lambda\delta_t}{r} - \frac{\lambda\delta_t^2}{2r^2}) \|y_t - v\|^2 \big) \Big] + \frac{2GD}{r} \sum_{t=1}^T w_t\delta_t. \qquad (3)\end{aligned}$$

where (a) uses G-Lipschitzness of f_t and Theorem 2.7 to bound the last four terms in the previous line, (b) uses λ-strong convexity of f_t^δ and $3G\delta_t + \frac{G\|v\|}{r}\delta_t \leqslant \frac{2GD}{r}\delta_t$, since $r\mathbb{B}^k \subseteq \mathcal{K}$ and the diameter bound give $\|v\| + r \leqslant D$ and $2r \leqslant D$, and (c) substitutes $y_t^\delta = (1 - \frac{\delta_t}{r})y_t$ and $v_t^\delta = (1 - \frac{\delta_t}{r})v$.

We analyze the terms in (3) separately. First, using the fact that $\mathbb{E}[\widehat{g}_t|\mathcal{F}_t] = \nabla f_t^\delta(y_t^\delta)$ while $\widetilde{y}_t$ is $\mathcal{F}_t$-measurable, we write

$$\begin{aligned}\mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t \Big(\langle \nabla f_t^\delta(y_t^\delta),\, \widetilde{y}_t - v \rangle - \frac{\lambda}{2}\|y_t - v\|^2 \Big) \Big] &= \mathbb{E}\Big[\textstyle\sum_{t=1}^T \big(\langle w_t\widehat{g}_t,\, \widetilde{y}_t - v \rangle - \frac{w_t\lambda}{2}\|y_t - v\|^2 \big) \Big] \\ &\overset{(d)}{\leqslant} \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\alpha_t - w_t\lambda}{2}\|y_t - v\|^2 + \sum_{t=1}^T \frac{\eta_{t-1}w_t^2\|\widehat{g}_t\|^2}{2} \Big] \\ &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\alpha_t - w_t\lambda}{2}\|y_t - v\|^2 \Big] + G^2 \sum_{t=1}^T \eta_{t-1} \frac{k^2M^2w_t^2}{G^2\delta_t^2},\end{aligned}$$

where (d) applies Theorem B.3 to $(\widetilde{y}_t)_{t=1}^T$.

Secondly, applying Assumption 2.2 and Theorem 2.7, we write

$$\begin{aligned}
\mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(\langle \nabla f_t^\delta(y_t^\delta),\, y_t - \widetilde{y}_t\rangle)\Big] &\leqslant G\;\mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t\|y_t - \widetilde{y}_t\|\Big] \\
&\overset{\text{(e)}}{\leqslant} G\;\mathbb{E}\Big[\textstyle\sum_{t=1}^T \eta_{t-1} w_t \,\|\sum_{s\in\mathcal{B}_t} w_s \widehat{g}_s\|\Big] \\
&\overset{\text{(f)}}{\leqslant} 2G\textstyle\sum_{t=1}^T \eta_{t-1} w_t \left(G\sum_{s\in\mathcal{B}_t} w_s + \frac{kM}{\delta_t}\sqrt{\sum_{s\in\mathcal{B}_t} w_s^2}\right) \\
&= 2G^2\textstyle\sum_{t=1}^T \eta_{t-1}\left(\frac{kMw_t}{G\delta_t}\sqrt{\sum_{s\in\mathcal{B}_t} w_s^2} + w_t\sum_{s\in\mathcal{B}_t} w_s\right),
\end{aligned}$$

where (e) applies Lemma B.4 (with appropriate $\phi_0 = \phi_1$) and (f) uses Lemma C.1 on the set $\mathcal{B}_t$.

For the last term, it holds that

$$\mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t\left(\frac{\delta_t}{r}\langle\nabla f_t^\delta(y_t^\delta),\, v - y_t\rangle + (\frac{\lambda\delta_t}{r} - \frac{\lambda\delta_t^2}{2r^2})\,\|y_t - v\|^2\right)\Big] \leqslant \frac{(G+\lambda D)D}{r}\textstyle\sum_{t=1}^T w_t\delta_t.$$

Combining these bounds, we have

$$\begin{aligned}
\bar{R}_T(v) &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\alpha_t - w_t\lambda}{2}\|y_t - v\|^2\Big] + G^2\textstyle\sum_{t=1}^T \eta_{t-1}\frac{k^2M^2w_t^2}{G^2\delta_t^2} \\
&\quad + 2G^2\textstyle\sum_{t=1}^T \eta_{t-1}\left(\frac{kMw_t}{G\delta_t}\sqrt{\sum_{s\in\mathcal{B}_t} w_s^2} + w_t\sum_{s\in\mathcal{B}_t} w_s\right) \\
&\quad + \frac{(G+\lambda D)D}{r}\textstyle\sum_{t=1}^T w_t\delta_t + \frac{2GD}{r}\sum_{t=1}^T w_t\delta_t \\
&\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\alpha_t - w_t\lambda}{2}\|y_t - v\|^2\Big] \\
&\quad + 2G^2\textstyle\sum_{t=1}^T \eta_{t-1}\left(\frac{k^2M^2w_t^2}{G^2\delta_t^2} + \frac{kMw_t}{G\delta_t}\sqrt{\sum_{s\in\mathcal{B}_t} w_s^2} + w_t\sum_{s\in\mathcal{B}_t} w_s\right) \\
&\quad + \frac{3GD+\lambda D^2}{r}\textstyle\sum_{t=1}^T w_t\delta_t.
\end{aligned}$$

By substituting $\lambda = 0$ or $\lambda > 0$, while letting $D \leqslant \frac{2G}{\lambda}$ in the strongly convex case (see Fact B.1), we obtain the claimed results in both cases. □

Lemma C.1. *Under the conditions of Theorem 3.2, for every set $\mathcal{B} \subseteq [T]$ with $\delta = \min_{t\in\mathcal{B}}\delta_t$,*

$$\mathbb{E}[\|\textstyle\sum_{t\in\mathcal{B}} w_t\widehat{g}_t\|] \leqslant 2\left(G\sum_{t\in\mathcal{B}} w_t + \frac{kM}{\delta}\sqrt{\sum_{t\in\mathcal{B}} w_t^2}\right).$$

Proof. For every $t \in \mathcal{B}$, let $g_t = \nabla f_t^\delta(y_t^\delta) = \mathbb{E}[\widehat{g}_t|\mathcal{F}_t]$ and note

$$\|g_t\| \leqslant G \qquad \text{and} \qquad \|\widehat{g}_t\| \leqslant \tfrac{kM}{\delta}.$$

Consequently, from the triangle inequality and Jensen's inequality, it follows that

$$\begin{aligned}
\mathbb{E}[\|\textstyle\sum_{t\in\mathcal{B}} w_t\widehat{g}_t\|] &\leqslant \mathbb{E}[\|\textstyle\sum_{t\in\mathcal{B}} w_t g_t\|] + \mathbb{E}[\|\textstyle\sum_{t\in\mathcal{B}} w_t(\widehat{g}_t - g_t)\|] \\
&\leqslant G\textstyle\sum_{t\in\mathcal{B}} w_t + \sqrt{\mathbb{E}\big[\|\textstyle\sum_{t\in\mathcal{B}} w_t(\widehat{g}_t - g_t)\|^2\big]}.
\end{aligned}$$

The second term under the square root can be bounded as

$$\begin{aligned}
\mathbb{E}\Big[\|\textstyle\sum_{t\in\mathcal{B}} w_t(\widehat{g}_t - g_t)\|^2\Big] &= \mathbb{E}\big[\textstyle\sum_{t\in\mathcal{B}} w_t^2\|\widehat{g}_t - g_t\|^2\big] \\
&\quad + 2\,\mathbb{E}\Big[\textstyle\sum_{t=1}^T\sum_{s=t+1}^T \mathbb{I}(t,s\in\mathcal{B})\, w_t w_s\,\langle\widehat{g}_t - g_t, \widehat{g}_s - g_s\rangle\Big] \\
&\leqslant \textstyle\sum_{t\in\mathcal{B}} w_t^2(G + \frac{kM}{\delta})^2 + 0.
\end{aligned}$$

Combining these results, we write

$$\begin{aligned}
\mathbb{E}[\|\textstyle\sum_{t\in\mathcal{B}} w_t\widehat{g}_t\|] &\leqslant G\textstyle\sum_{t\in\mathcal{B}} w_t + (G + \frac{kM}{\delta})\sqrt{\sum_{t\in\mathcal{B}} w_t^2} \\
&\leqslant 2\left(G\textstyle\sum_{t\in\mathcal{B}} w_t + \frac{kM}{\delta}\sqrt{\sum_{t\in\mathcal{B}} w_t^2}\right).
\end{aligned}$$

This concludes the proof of the lemma. □

D Reduction via Proxy Delay Scheduling: Proofs

We prove the results of Section 4. The proof of Theorem 4.1 extends the importance-weighting argument of Ryabchenko et al. [2025] from multi-armed bandits to the convex setting. The saturation bound (Lemma 4.3) and the Pareto scheduler (Corollary 4.4) simplify the corresponding analysis by reducing it to a single Chernoff condition on the tracking set size. The Bernoulli scheduler (Corollary 4.5), which calibrates to the maximum backlog $\sigma_{\max}$, is new to this work.

Theorem 4.1 (Restated). *Suppose the proxy delays are sampled independently and $\mathbb{P}(d'_t \geqslant d_t) > 0$ for every $t \in [T]$. The weights forwarded to $\mathcal{A}$ satisfy*

$$w_t = \tfrac{\mathbb{I}(|\mathcal{S}_t|<C)\,\mathbb{I}(d'_t \geqslant d_t)}{\mathbb{P}(d'_t \geqslant d_t)}, \quad \mathbb{E}[w_t] \leqslant 1, \quad \mathbb{E}[w_t^2] \leqslant \tfrac{1}{\mathbb{P}(d'_t \geqslant d_t)}, \quad \mathbb{E}[w_t w_s] \leqslant 1,$$

for all distinct $t, s \in [T]$. Moreover, the expected regret satisfies

$$\bar{R}_T(u) \;\leqslant\; \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t\big(f_t(x_t) - f_t(u)\big)\Big] + GD\,\textstyle\sum_{t=1}^T \mathbb{P}\big(|\mathcal{S}_t| = C\big).$$

Proof. From the structure of Algorithm 1, feedback from round t is observed in round $t + d_t$ if and only if $|S_t| < C$ and $d'_t \geqslant d_t$. Consequently, we have the identity $w_t = \frac{\mathbb{I}(|\mathcal{S}_t|<C)\,\mathbb{I}(d'_t \geqslant d_t)}{\mathbb{P}(d'_t \geqslant d_t)}$.

Therefore, for each $t \in [T]$ and $s \in [T]\backslash\{t\}$, it holds that

$$\mathbb{E}[w_t] \leqslant \mathbb{E}\left[\tfrac{\mathbb{I}(d'_t \geqslant d_t)}{\mathbb{P}(d'_t \geqslant d_t)}\right] = 1, \quad \mathbb{E}[w_t^2] \leqslant \mathbb{E}\left[\tfrac{\mathbb{I}(d'_t \geqslant d_t)}{(\mathbb{P}(d'_t \geqslant d_t))^2}\right] = \tfrac{1}{\mathbb{P}(d'_t \geqslant d_t)},$$
$$\mathbb{E}[w_t w_s] \leqslant \mathbb{E}\left[\tfrac{\mathbb{I}(d'_t \geqslant d_t)\,\mathbb{I}(d'_s \geqslant d_s)}{\mathbb{P}(d'_t \geqslant d_t)\,\mathbb{P}(d'_s \geqslant d_s)}\right] = \mathbb{E}\left[\tfrac{\mathbb{I}(d'_t \geqslant d_t)}{\mathbb{P}(d'_t \geqslant d_t)}\right]\mathbb{E}\left[\tfrac{\mathbb{I}(d'_s \geqslant d_s)}{\mathbb{P}(d'_s \geqslant d_s)}\right] = 1,$$

where the last inequality applies the fact that proxy delays d'_t and d'_s are sampled independently.

For the regret bound, we write

$$\begin{aligned}
\bar{R}_T(u) &= \mathbb{E}\Big[\textstyle\sum_{t=1}^T (f_t(x_t) - f_t(u))\Big] \\
&\overset{(a)}{\leqslant} \mathbb{E}\Big[\textstyle\sum_{t=1}^T \mathbb{I}(|\mathcal{S}_t| < C)\,(f_t(x_t) - f_t(u))\Big] + GD\,\mathbb{E}\Big[\textstyle\sum_{t=1}^T \mathbb{I}(|\mathcal{S}_t| = C)\Big] \\
&\overset{(b)}{=} \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\mathbb{I}(|\mathcal{S}_t|<C)\,\mathbb{I}(d'_t \geqslant d_t)}{\mathbb{P}(d'_t \geqslant d_t)}\,(f_t(x_t) - f_t(u))\Big] + GD\,\textstyle\sum_{t=1}^T \mathbb{P}(|\mathcal{S}_t| = C) \\
&\overset{(c)}{=} \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t\,(f_t(x_t) - f_t(u))\Big] + GD\,\textstyle\sum_{t=1}^T \mathbb{P}(|\mathcal{S}_t| = C),
\end{aligned}$$

where (a) combines Assumptions 2.1 and 2.2, (b) uses the fact that d'_t is sampled from $\mathfrak{D}_t$ at the start of round t independently of S_t and x_t, and (c) applies the identity $w_t = \frac{\mathbb{I}(|\mathcal{S}_t|<C)\,\mathbb{I}(d'_t \geqslant d_t)}{\mathbb{P}(d'_t \geqslant d_t)}$. □

Lemma 4.3 (Restated). *Consider proxy delay distributions $\mathfrak{D} = (\mathfrak{D}_t)_{t=1}^T$ such that for every $t \in [T]$*

$$\textstyle\sum_{s \in \mathcal{B}_t \cup \{t\}} \mathbb{P}(d'_s \geqslant t - s) \leqslant \frac{C}{8}.$$

Then, $\mathbb{P}\big(|\mathcal{S}_t| = C\big) \leqslant \exp(-C)$ for every $t \in [T]$.

Proof. Fix $t \in [T]$. For each $s \in \mathcal{B}_t \cup \{t\}$, define $X_s = \mathbb{I}(d'_s \geqslant t - s)$.

By independence of the proxy-delay draws across rounds, the random variables $(X_s)_{s \in \mathcal{B}_t \cup \{t\}}$ are independent Bernoulli random variables. Moreover, by construction of the proxy scheduler, $|\mathcal{S}_t| \leqslant \sum_{s \in \mathcal{B}_t \cup \{t\}} X_s$. Therefore, writing $X = \sum_{s \in \mathcal{B}_t \cup \{t\}} X_s$, we have

$$\mathbb{P}(|\mathcal{S}_t| = C) \leqslant \mathbb{P}(|\mathcal{S}_t| \geqslant C) \leqslant \mathbb{P}(X \geqslant C).$$

Let $\mu = \mathbb{E}[X] = \sum_{s \in \mathcal{B}_t \cup \{t\}} \mathbb{P}(d'_s \geqslant t - s)$. By assumption, $\mu \leqslant C/8$.

If $\mu = 0$, then $\mathbb{P}(X \geqslant C) = 0$. Assume henceforth that $\mu > 0$, and define

$$\alpha = C/\mu - 1 \in [7, \infty).$$

By the multiplicative Chernoff bound (e.g., see Theorem 2.3.1 from Vershynin [2026]),

$$\begin{aligned}\mathbb{P}(X \geqslant C) &= \mathbb{P}\big(X \geqslant (1+\alpha)\mu\big) \\ &\leqslant \left(\frac{e^{\alpha}}{(1+\alpha)^{1+\alpha}}\right)^{\mu} \\ &= \exp\left(C\left(\tfrac{\alpha}{1+\alpha} - \ln(1+\alpha)\right)\right).\end{aligned}$$

As the function $f(x) = \ln(1+x) - \frac{x}{1+x}$ is increasing on $[0,\infty)$, α satisfies

$$\tfrac{\alpha}{1+\alpha} - \ln(1+\alpha) \leqslant 7/8 - \ln 8 < -1.$$

Therefore,

$$\mathbb{P}\big(|\mathcal{S}_t| = C\big) \leqslant \mathbb{P}(X \geqslant C) \leqslant e^{-C}.$$

□

Corollary 4.4 (Restated). *For each $t \in [T]$, let $H_t = \sum_{j=1}^{t} \frac{1}{j}$. Take $\mathfrak{D}_t$ to be the shifted floored Pareto distribution*

$$\mathfrak{D}_t \;=\; \Big\lfloor \mathrm{Pareto}\big(\tfrac{C}{16H_t},\, 1\big) \Big\rfloor - 1.$$

Then $\mathbb{P}(d'_t \geqslant d_t) = \min\big\{1,\, \frac{C}{16H_t(d_t+1)}\big\}$ and $\mathbb{P}(|\mathcal{S}_t| = C) \leqslant \exp(-C)$ for every $t \in [T]$.

Proof. Fix $t \in [T]$. For each $s \leqslant t$, by the Pareto tail formula,

$$\mathbb{P}(d'_s \geqslant t-s) = \mathbb{P}\Big(\mathrm{Pareto}\Big(\tfrac{C}{16H_s}, 1\Big) \geqslant t-s+1\Big) = \min\Big\{1, \tfrac{C}{16H_s(t-s+1)}\Big\}.$$

It remains to verify the condition of Lemma 4.3. It follows from Fact B.2 for harmonic numbers.

Thus, Lemma 4.3 now yields $\mathbb{P}(|\mathcal{S}_t| = C) \leqslant e^{-C}$. □

Corollary 4.5 (Restated). *Let $p = \min\{1, \frac{C}{8(\sigma_{\max}+1)}\}$. For every $t \in [T]$, let the proxy distribution $\mathfrak{D}_t \in \Delta(\{-1,\infty\})$ be given by*

$$\mathbb{P}_{d' \sim \mathfrak{D}_t}(d' = \infty) = p, \qquad \mathbb{P}_{d' \sim \mathfrak{D}_t}(d' = -1) = 1-p.$$

Then, $\mathbb{P}(d'_t \geqslant d_t) = p$ and $\mathbb{P}(|\mathcal{S}_t| = C) \leqslant \exp(-C)$ for every $t \in [T]$.

Proof. Fix $t \in [T]$. Since $d_t \geqslant 0$ and $d'_t \in \{-1,\infty\}$, $\mathbb{P}(d'_t \geqslant d_t) = \mathbb{P}(d'_t = \infty) = p$.

Thus, the condition of Lemma 4.3 is satisfied

$$\sum_{s \in \mathcal{B}_t \cup \{t\}} \mathbb{P}(d'_s \geqslant t-s) = p\,|\mathcal{B}_t \cup \{t\}| = p(\sigma_t+1) \leqslant \tfrac{C(\sigma_t+1)}{8(\sigma_{\max}+1)} \leqslant C/8.$$

Applying Lemma 4.3, we obtain

$$\mathbb{P}(|\mathcal{S}_t| = C) \leqslant \exp(-C).$$

This completes the proof. □

E Guarantees for Capacity-Constrained OCO and BCO: Proofs

In this section, we prove the main regret bounds for capacity-constrained OCO and BCO stated in Section 5, covering both convex and strongly convex losses.

Theorem 5.1 (Restated). *Suppose $C \geqslant \ln T + 1$, and consider $\mathcal{W}(\mathcal{A}, \mathfrak{D})$, where $\mathcal{A}$ is DW-FTRL (1). With learning rates $\eta_t = \frac{D/G}{\sqrt{d_{\mathrm{tot}}+T}}$ and $\mathfrak{D}$ from Corollary 4.4, $\mathcal{W}(\mathcal{A}, \mathfrak{D})$ guarantees*

$$\bar{R}_T(x^*) = O\left(GD\sqrt{d_{\mathrm{tot}} + T}\right).$$

Under λ-strong convexity (2.3), for rates $\eta_t = \frac{1}{t\lambda}$ and $\mathfrak{D}$ from Corollary 4.5, $\mathcal{W}(\mathcal{A}, \mathfrak{D})$ guarantees

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

Proof. Note that $C \geqslant \ln(eT) \geqslant \sum_{t=1}^{T} \frac{1}{t} = H_T$ and $Te^{-C} \leqslant 1$ throughout the proof. Fix comparator $u \in \mathcal{K}$. We prove the bounds on $\bar{R}_T(u)$ for general convex and strongly convex losses separately.

General convex losses: We use fixed learning rate $\eta_t = \eta$ for the value $\eta = \frac{D/G}{\sqrt{d_{\text{tot}}+T}}$. Combining Theorems 4.1, 3.1 and Corollary 4.4, we write

$$\begin{aligned}
\bar{R}_T(u) &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^{T} w_t(f_t(x_t) - f_t(u))\Big] + GD \textstyle\sum_{t=1}^{T} \mathbb{P}(|\mathcal{S}_t| = C) \\
&\leqslant \tfrac{D^2}{\eta} + G^2\eta\, \mathbb{E}\Big[\textstyle\sum_{t=1}^{T}(w_t^2 + w_t \sum_{s\in\mathcal{B}_t} w_s)\Big] + GD\,T\exp(-C) \\
&\leqslant \tfrac{D^2}{\eta} + G^2\eta \textstyle\sum_{t=1}^{T} \left(\max\left\{1, \frac{16H_t}{C}\,(d_t+1)\right\} + \sigma_t\right) + GD \\
&\overset{\text{(a)}}{\leqslant} \tfrac{D^2}{\eta} + G^2\eta\, \left(1 + \tfrac{16H_T}{C}\right)\,(d_{\text{tot}} + T) + GD \\
&\overset{\text{(b)}}{\leqslant} \tfrac{D^2}{\eta} + 17G^2\eta\,(d_{\text{tot}} + T) + GD \\
&= 18GD\sqrt{d_{\text{tot}} + T} + GD,
\end{aligned}$$

where (a) applies $\sum_{t=1}^{T} \sigma_t = d_{\text{tot}}$ from Lemma 2.6 and (b) follows for $C \geqslant H_T$.

Strongly convex losses: Combining Theorems 4.1, 3.1 and Corollary 4.5, while letting $D \leqslant \frac{2G}{\lambda}$ in the strongly convex case (see Fact B.1), we write

$$\begin{aligned}
\bar{R}_T(u) &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^{T} w_t(f_t(x_t) - f_t(u))\Big] + GD \textstyle\sum_{t=1}^{T} \mathbb{P}(|\mathcal{S}_t| = C) \\
&\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^{T} \frac{\alpha_t - w_t\lambda}{2} \|x_t - u\|^2\Big] + G^2\, \mathbb{E}\Big[\textstyle\sum_{t=1}^{T} \eta_{t-1}(w_t^2 + w_t \sum_{s\in\mathcal{B}_t} w_s)\Big] + \tfrac{2G^2}{\lambda}\, Te^{-C} \\
&\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^{T} \frac{\lambda - w_t\lambda}{2} \|x_t - u\|^2\Big] + 2G^2\, \mathbb{E}\Big[\textstyle\sum_{t=1}^{T} \eta_t(w_t^2 + w_t \sum_{s\in\mathcal{B}_t} w_s)\Big] + \tfrac{2G^2}{\lambda}, \qquad (4)
\end{aligned}$$

where the last inequality substitutes $\alpha_t = \lambda$, $\eta_{t-1} \leqslant 2\eta_t$, and $Te^{-C} \leqslant 1$.

To control the first term in (4), we use the identity for w_t in Theorem 4.1,

$$\begin{aligned}
\mathbb{E}\Big[\textstyle\sum_{t=1}^{T} \frac{\lambda - w_t\lambda}{2} \|x_t - u\|^2\Big] &= \tfrac{\lambda}{2} \textstyle\sum_{t=1}^{T} \mathbb{E}\Big[\Big(1 - \frac{\mathbb{I}(|\mathcal{S}_t|<C)\,\mathbb{I}(d_t'\geqslant d_t)}{\mathbb{P}(d_t'\geqslant d_t)}\Big)\|x_t - u\|^2\Big] \\
&\overset{\text{(a)}}{=} \tfrac{\lambda}{2} \textstyle\sum_{t=1}^{T} \mathbb{E}\big[(1 - \mathbb{I}(|\mathcal{S}_t| < C))\,\|x_t - u\|^2\big] \\
&\overset{\text{(b)}}{\leqslant} \tfrac{2G^2}{\lambda} \textstyle\sum_{t=1}^{T} \mathbb{P}(|\mathcal{S}_t| = C) \\
&\leqslant \tfrac{2G^2}{\lambda}\, Te^{-C} \\
&\leqslant \tfrac{2G^2}{\lambda},
\end{aligned}$$

where (a) uses the fact that d_t' is sampled independently of x_t, and (b) lets $D \leqslant \frac{2G}{\lambda}$ (see Fact B.1).

For the second term in (4), we apply $\mathbb{E}[w_t^2] \leqslant \frac{1}{p}$ and $\mathbb{E}[w_s w_t] \leqslant 1$ from Theorem 4.1 for $p = \min\{1, \frac{C}{8(\sigma_{\max}+1)}\}$ in Corollary 4.5 and substitute learning rate values $\eta_t = \frac{1}{t\lambda}$,

$$\begin{aligned}
\mathbb{E}\Big[\textstyle\sum_{t=1}^{T} \eta_t(w_t^2 + w_t \sum_{s\in\mathcal{B}_t} w_s)\Big] &\leqslant \textstyle\sum_{t=1}^{T} \frac{1}{t\lambda} \left(\max\left\{1, \frac{8(\sigma_{\max}+1)}{C}\right\} + \sigma_{\max}\right) \\
&\leqslant \tfrac{9(\sigma_{\max}+1)H_T}{\lambda}.
\end{aligned}$$

Combining these bounds, we finally have

$$\bar{R}_T(u) \leqslant \tfrac{18G^2}{\lambda}(\sigma_{\max}+1)H_T + \tfrac{4G^2}{\lambda} = O\left(\tfrac{G^2}{\lambda}(\sigma_{\max}+1)\ln T\right).$$

This completes the proof. □

We now proceed to the BCO case.

Theorem 5.2 (Restated). *Suppose $C \geqslant \ln T + 1$, and consider $\mathcal{W}(\mathcal{A}, \mathfrak{D})$, where $\mathcal{A}$ is DW-FTBL (2) and $\mathfrak{D}$ is from Corollary 4.5. With smoothing parameters $\delta_t = r \min\left\{1, \frac{\sqrt{\nu k}}{t^{1/4}}\,(1 + \frac{\sigma_{\max}}{C})^{1/4}\right\}$ and*

learning rates $\eta_t = \frac{D/G}{t^{1/2}\sqrt{\sigma_{\max}+(\frac{\nu k}{\delta_t/r})^2(1+\frac{\sigma_{\max}}{C})}}$, $\mathcal{W}(\mathcal{A},\mathfrak{D})$ *guarantees*

$$\bar{R}_T(x^*) = O\Big(GD\left(\sqrt{T\sigma_{\max}} + T^{3/4}\left(1+\tfrac{\sigma_{\max}}{C}\right)^{1/4}\sqrt{\nu k}\right)\Big).$$

Under λ*-strong convexity (2.3), with* $\eta_t = \frac{1}{t\lambda}$ *and* $\delta_t = r\,\min\Big\{1, (\frac{\nu^2k^2\ln(t+1)}{t})^{1/3}(1+\frac{\sigma_{\max}}{C})^{1/3}\Big\}$,

$$\bar{R}_T(x^*) = O\Big(\tfrac{G^2}{\lambda}\Big(\sigma_{\max}\ln T + (T^2\ln T)^{1/3}(1+\tfrac{\sigma_{\max}}{C})^{1/3}(\nu k)^{2/3}\Big)\Big).$$

Proof. Recall $\nu = \frac{M}{Gr}$. Note that $C \geqslant \ln(eT) \geqslant \sum_{t=1}^T \frac{1}{t}$ and $Te^{-C} \leqslant 1$ throughout the proof. Fix comparator $u \in \mathcal{K}$. We prove the bounds on $\bar{R}_T(u)$ for convex and strongly convex losses separately.

General convex losses: We use non-increasing smoothing parameters and learning rates

$$\delta_t = r\,\min\Big\{1, \tfrac{\sqrt{\nu k}}{t^{1/4}}(1+\tfrac{\sigma_{\max}}{C})^{1/4}\Big\}, \qquad \eta_t = \frac{D/G}{t^{1/2}\sqrt{\sigma_{\max}+(\frac{\nu k}{\delta_t/r})^2(1+\frac{\sigma_{\max}}{C})}}.$$

Observe that $\delta_{t-1} \leqslant 2\delta_t$ for all $t \geqslant 2$ and consequently $\eta_{t-1} \leqslant 2\eta_t$ for all $t \geqslant 1$.

Also, consider $\delta_{\text{tot}} = \sum_{t=1}^T \delta_t$, for which it holds that

$$\delta_{\text{tot}}/r \leqslant \left(\textstyle\sum_{t=1}^T t^{-1/4}\right)(1+\tfrac{\sigma_{\max}}{C})^{1/4}\sqrt{\nu k} \leqslant \tfrac{4}{3}T^{3/4}(1+\tfrac{\sigma_{\max}}{C})^{1/4}\sqrt{\nu k}.$$

From Corollary 4.5, we additionally get

$$1/p = \max\{1, \tfrac{8(\sigma_{\max}+1)}{C}\} \leqslant 9(1+\tfrac{\sigma_{\max}}{C}).$$

Combining Theorems 4.1, 3.2 and Corollary 4.5, we write

$$\begin{aligned}
\bar{R}_T(u) &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t(x_t) - f_t(u))\Big] + GD\,\sum_{t=1}^T \mathbb{P}(|\mathcal{S}_t| = C) \\
&\leqslant \tfrac{D^2}{\eta_T} + 2G^2\,\textstyle\sum_{t=1}^T \eta_{t-1}\mathbb{E}\Big[(\frac{kMw_t}{G\delta_t})^2 + \frac{kMw_t}{G\delta_t}\sqrt{\sum_{s\in\mathcal{B}_t} w_s^2} + w_t\sum_{s\in\mathcal{B}_t} w_s\Big] \\
&+ \tfrac{3GD}{r}\textstyle\sum_{t=1}^T \mathbb{E}[w_t\delta_t] + GD\,Te^{-C} \\
&\overset{(a)}{\leqslant} \tfrac{D^2}{\eta_T} + 4G^2\,\textstyle\sum_{t=1}^T \eta_t\left((\frac{k\nu}{\delta_t/r})^2/p + \frac{k\nu}{\delta_t/r}\sqrt{\sigma_{\max}/p} + \sigma_{\max}\right) + 3GD\,(\tfrac{\delta_{\text{tot}}}{r}+1) \\
&\overset{(b)}{\leqslant} \tfrac{D^2}{\eta_T} + 72G^2\textstyle\sum_{t=1}^T \eta_t\Big(\sigma_{\max} + (\frac{k\nu}{\delta_t/r})^2(1+\frac{\sigma_{\max}}{C})\Big) + 3GD\,(\tfrac{\delta_{\text{tot}}}{r}+1) \\
&\overset{(c)}{\leqslant} 145GD\,\left(\sqrt{T\sigma_{\max}} + T^{1/2}(\tfrac{k\nu}{\delta_T/r})(1+\tfrac{\sigma_{\max}}{C})^{1/2} + \delta_{\text{tot}}/r + 1\right),
\end{aligned} \tag{5}$$

where (a) uses Jensen's inequality for $\mathbb{E}[(\sum_{s\in\mathcal{B}_t} w_s^2)^{1/2}]^2 \leqslant \mathbb{E}[\sum_{s\in\mathcal{B}_t} w_s^2] \leqslant \sigma_{\max}/p$, (b) uses AM-GM and $1/p \leqslant 9(1+\frac{\sigma_{\max}}{C})$, and (c) substitutes η_t and uses subadditivity of the square root.

Let $\delta_T < r$, since otherwise $T^{1/4} \leqslant (1+\frac{\sigma_{\max}}{C})^{1/4}\sqrt{\nu k}$, and the bound is vacuous $O(T)$. Then,

$$T^{1/2}(\tfrac{k\nu}{\delta/r})(1+\tfrac{\sigma_{\max}}{C})^{1/2} = T^{3/4}(1+\tfrac{\sigma_{\max}}{C})^{1/4}\sqrt{\nu k}, \qquad \delta_{\text{tot}}/r \leqslant \tfrac{4}{3}T^{3/4}(1+\tfrac{\sigma_{\max}}{C})^{1/4}\sqrt{\nu k}.$$

Combining the bounds above for $u = x^*$ yields

$$\bar{R}_T(x^*) = O\left(GD\left(\sqrt{T\sigma_{\max}} + T^{3/4}(1+\tfrac{\sigma_{\max}}{C})^{1/4}\sqrt{\nu k}\right)\right).$$

Strongly convex losses: We use non-increasing smoothing parameters and learning rates

$$\delta_t = r\,\min\Big\{1, (\tfrac{\nu^2k^2\ln(t+1)}{t})^{1/3}(1+\tfrac{\sigma_{\max}}{C})^{1/3}\Big\}, \qquad \eta_t = \tfrac{1}{\lambda t}.$$

Consider $\delta_{\text{tot}} = \sum_{t=1}^T \delta_t$, for which it holds that

$$\begin{aligned}
\delta_{\text{tot}}/r &\leqslant \left(\textstyle\sum_{t=1}^T t^{-1/3}\right)(\ln(T+1))^{1/3}(1+\tfrac{\sigma_{\max}}{C})^{1/3}(\nu k)^{2/3} \\
&\leqslant \tfrac{3}{2}(T^2\ln(T+1))^{1/3}(1+\tfrac{\sigma_{\max}}{C})^{1/3}(\nu k)^{2/3}.
\end{aligned}$$

Combining Theorems 4.1, 3.2 and Corollary 4.5, we write

$$\begin{aligned}\bar{R}_T(u) &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^T w_t(f_t(x_t) - f_t(u))\Big] + GD\,\textstyle\sum_{t=1}^T \mathbb{P}(|\mathcal{S}_t| = C)\\ &\leqslant \mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\alpha_t - w_t\lambda}{2}\|y_t - u\|^2\Big] + \frac{10G^2}{\lambda r}\mathbb{E}\left[\textstyle\sum_{t=1}^T w_t\delta_t\right]\\ &\quad + 4G^2\textstyle\sum_{t=1}^T \eta_t\mathbb{E}\left[\frac{k^2M^2w_t^2}{G^2\delta_t^2} + \frac{kMw_t}{G\delta_t}\sqrt{\textstyle\sum_{s\in\mathcal{B}_t} w_s^2} + w_t\textstyle\sum_{s\in\mathcal{B}_t} w_s\right] + GD, \qquad (6)\end{aligned}$$

where the last inequality additionally applies $\eta_{t-1} \leqslant 2\eta_t$.

To control the first term in (6), we use the identity for w_t in Theorem 4.1,

$$\begin{aligned}\mathbb{E}\Big[\textstyle\sum_{t=1}^T \frac{\lambda - w_t\lambda}{2}\|y_t - u\|^2\Big] &= \tfrac{\lambda}{2}\textstyle\sum_{t=1}^T \mathbb{E}\Big[\Big(1 - \frac{\mathbb{I}(|\mathcal{S}_t|<C)\,\mathbb{I}(d_t'\geqslant d_t)}{\mathbb{P}(d_t'\geqslant d_t)}\Big)\|y_t - u\|^2\Big]\\ &\overset{\text{(d)}}{=} \tfrac{\lambda}{2}\textstyle\sum_{t=1}^T \mathbb{E}\big[(1 - \mathbb{I}(|\mathcal{S}_t| < C))\,\|y_t - u\|^2\big]\\ &\overset{\text{(e)}}{\leqslant} \tfrac{2G^2}{\lambda}\textstyle\sum_{t=1}^T \mathbb{P}(|\mathcal{S}_t| = C)\\ &\leqslant \tfrac{2G^2}{\lambda},\end{aligned}$$

where (d) uses the fact that d_t' is sampled independently of y_t, and (e) lets $D \leqslant \frac{2G}{\lambda}$ (see Fact B.1).

For the second term in (6), we use Theorem 4.1 to write

$$\begin{aligned}\tfrac{G^2}{\lambda r}\mathbb{E}\left[\textstyle\sum_{t=1}^T w_t\delta_t\right] &\leqslant \tfrac{G^2}{\lambda}\,(\delta_{\text{tot}}/r)\\ &\leqslant \tfrac{3G^2}{2\lambda}\,T^{2/3}(\ln(T+1))^{1/3}(1 + \tfrac{\sigma_{\max}}{C})^{1/3}(\nu k)^{2/3}.\end{aligned}$$

For the third term in (6), we follow the analysis of the general convex case in (5) to write

$$\begin{aligned}G^2\textstyle\sum_{t=1}^T \eta_t\mathbb{E}\left[\frac{k^2M^2w_t^2}{G^2\delta_t^2} + \frac{kMw_t}{G\delta_t}\sqrt{\textstyle\sum_{s\in\mathcal{B}_t} w_s^2} + w_t\textstyle\sum_{s\in\mathcal{B}_t} w_s\right]\\ \leqslant \tfrac{18G^2}{\lambda}\left(\textstyle\sum_{t=1}^T \frac{1}{t}\right)\left(\sigma_{\max} + (\tfrac{k\nu}{\delta_T/r})^2(1 + \tfrac{\sigma_{\max}}{C})\right)\\ = O\left(\tfrac{G^2}{\lambda}\left(\sigma_{\max}\ln T + T^{2/3}(\ln T)^{1/3}(1 + \tfrac{\sigma_{\max}}{C})^{1/3}(\nu k)^{2/3}\right)\right).\end{aligned}$$

Combining the bounds above for $u = x^*$ yields

$$\bar{R}_T(x^*) = O\left(\tfrac{G^2}{\lambda}\left(\sigma_{\max}\ln T + T^{2/3}(\ln T)^{1/3}(1 + \tfrac{\sigma_{\max}}{C})^{1/3}(\nu k)^{2/3}\right)\right).$$

This completes the proof. □

F Doubling Trick over Maximum Backlog

The strongly convex and bandit results of Theorems 5.1 and 5.2 require tuning with knowledge of $\sigma_{\max}$, which may not be available in advance. Since the learning rates and smoothing parameters in all our results depend on the round index t but not on the horizon T, our base algorithms are anytime in T, and we need only double over $\sigma_{\max}$. This avoids the two-dimensional doubling over (T, d_{tot}) used by Bistritz et al. [2022].

Epoch structure. Define backlog guesses $S_j = 2^j - 1$ for $j \in \mathbb{Z}_{\geqslant 0}$. The wrapper initializes at epoch $j = 0$ and restarts the base algorithm with parameters tuned to maximum backlog S_j. Within each epoch, the algorithm runs until the observed backlog exceeds S_j, at which point the epoch ends, j is incremented, and the algorithm restarts with the updated guess. Upon starting a new epoch, the wrapper clears the tracking set, discarding any pending observations from the previous epoch. The final epoch index is $J = \lceil \log_2(\sigma_{\max} + 1) \rceil$.

Theorem F.1 (Doubling over $\sigma_{\max}$)**.** *Suppose the base algorithm is anytime in T: for any backlog guess $\widetilde{S}$, on any interval of length T' with maximum backlog at most $\widetilde{S}$, it guarantees regret at most*

$$R(T', \widetilde{S}) \;\leqslant\; \varphi_0(T') + \textstyle\sum_{m=1}^M \varphi_m(T')\,(\widetilde{S} + 1)^{b_m},$$

where $b_m > 0$ and each $\varphi_m : \mathbb{N} \to [0, \infty)$ for $m \in \{0\} \cup [M]$ is non-decreasing. Then the doubling wrapper guarantees

$$R(T, \sigma_{\max}) \;=\; O\Big(\varphi_0(T)\big(1 + \log(\sigma_{\max}+1)\big) + \textstyle\sum_{m=1}^M \varphi_m(T)\,(\sigma_{\max} + 1)^{b_m}\Big).$$

Proof. The wrapper produces at most $J+1$ epochs, indexed $j = 0, 1, \ldots, J$ with $J = \lceil \log_2(\sigma_{\max}+1) \rceil$. Let L_j denote the number of rounds in epoch j. Since $L_j \leqslant T$ and each φ_m is non-decreasing, the regret decomposes as

$$R(T, \sigma_{\max}) \;\leqslant\; \sum_{j=0}^{J} R(L_j, S_j) \;\leqslant\; (J+1)\,\varphi_0(T) \;+\; \sum_{m=1}^{M} \varphi_m(T) \sum_{j=0}^{J} 2^{b_m j}.$$

The first term equals $O\big(\varphi_0(T)(1+\log(\sigma_{\max}+1))\big)$. For the second, since $b_m > 0$, the geometric series satisfies $\sum_{j=0}^{J} 2^{b_m j} = O\big((\sigma_{\max}+1)^{b_m}\big)$. Summing over $m \in [M]$ completes the proof. □

Remark F.2. Every term in the regret bounds of Theorems 5.1 and 5.2 that depends on $\sigma_{\max}$ has $b_m > 0$, so the geometric series absorbs the epoch count for these terms. The $\sigma_{\max}$-independent terms (captured by φ_0) incur a $1+\log(\sigma_{\max}+1)$ overhead because, unlike d_{tot} or T, the maximum backlog does not accumulate across epochs: each restart effectively begins a new game, and φ_0 is charged in full for every epoch. The general convex OCO bound of Theorem 5.1 does not depend on $\sigma_{\max}$; for it, we use the doubling trick of Bistritz et al. [2022] over T and d_{tot} instead.

G Non-Preemptive Framework

Ryabchenko et al. [2025] also studied a more restrictive variant of capacity-constrained online learning in which preemption is disallowed — once a round enters the tracking set $\mathcal{S}$, it must remain until its feedback arrives. Our framework recovers this variant by restricting the support of proxy delay distributions $\mathfrak{D}_t$ to $\{-1, \infty\}$ (Section 4), so that each round is either rejected outright ($d'_t = -1$) or committed to until feedback delivery ($d'_t = \infty$). The two-point schedulers in Corollaries 4.5 and G.1 are non-preemptive in this sense.

Corollary G.1. *For each $t \in [T]$, define $H_t = \sum_{j=1}^{t} \frac{1}{j}$ and $p_t = \min\left\{1, \frac{C}{16 H_t (d_t+1)}\right\}$. Let the proxy distribution $\mathfrak{D}_t \in \Delta(\{-1, \infty\})$ be given by*

$$\mathbb{P}_{d' \sim \mathfrak{D}_t}(d' = \infty) = p_t, \qquad \mathbb{P}_{d' \sim \mathfrak{D}_t}(d' = -1) = 1 - p_t.$$

Then, $\mathbb{P}(d'_t \geqslant d_t) = p_t$ and $\mathbb{P}(|\mathcal{S}_t| = C) \leqslant \exp(-C)$ for every $t \in [T]$.

Proof. Fix $t \in [T]$. Since $\mathfrak{D}_t \in \Delta(\{-1, \infty\})$ and $d_t \in \mathbb{Z}_{\geqslant 0}$, we immediately have $\mathbb{P}(d'_t \geqslant d_t) = \mathbb{P}(d'_t = \infty) = p_t$. The rest follows analogously to the proof of Corollary 4.4. □

Corollary G.1 requires clairvoyance because its admission probability depends on d_t at prediction time. The scheduler of Corollary 4.5 requires no advance knowledge of individual delays and can be implemented using a known upper bound on $\sigma_{\max}$.

Since both schedulers satisfy the same saturation and weight moment bounds as their preemptive counterparts, combining them with the analysis of Theorems 5.1 and 5.2 yields identical regret guarantees in the non-preemptive setting.

H Experiments

Code for reproducing the experiments is publicly available at
`https://github.com/alex-rbch/cc-oco`.

We conduct numerical experiments to study how tracking capacity and scheduling affect regret under first-order and bandit feedback. We consider both convex and strongly convex losses, focusing on settings where informative feedback takes longer to arrive. Such dependence can arise when longer or more difficult tasks produce systematically different losses. Retaining this feedback occupies capacity for longer, while favoring short-delay feedback can discard useful information. We model this dependence using heavy-tailed delays, with both the losses and delays fixed before play, consistent with the oblivious-adversary model. See Appendix H.1 for experimental details.

For each loss family, we vary the capacity C and compare empirical variants of our uniform-weight Bernoulli scheduler (Corollary 4.5) and the Pareto-type scheduler (Corollary 4.4), using the same losses and delays across methods and capacities. Full tracking provides an unconstrained baseline.

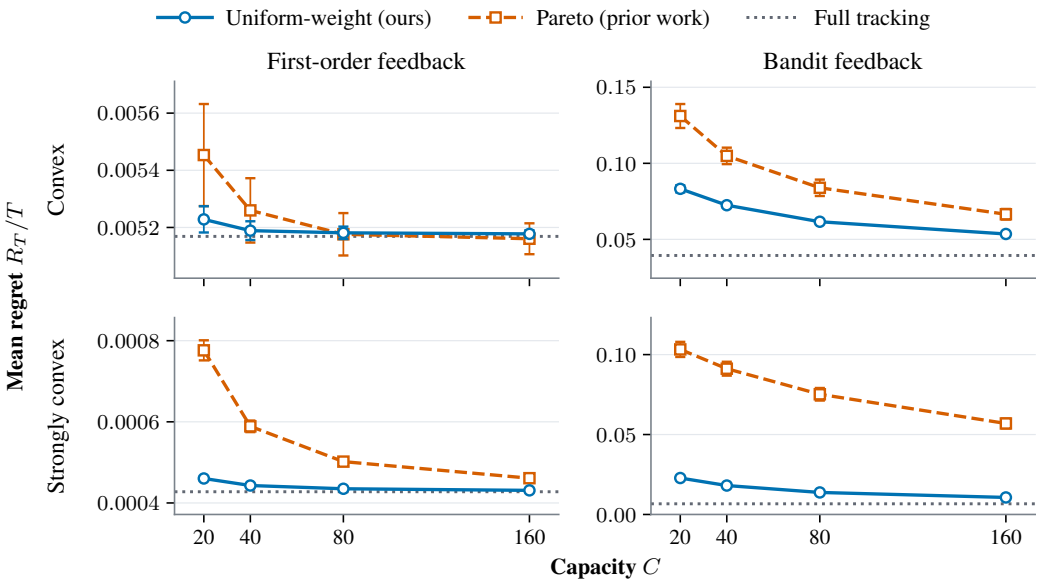


Figure 3: Capacity ablation for convex (top) and strongly convex (bottom) losses under first-order (left) and bandit (right) feedback. Both loss families take values in $[0, 1]$. We use $C \in \{20, 40, 80, 160\}$, horizon $T = 10^6$, and delays sampled with mean 400. The plots show mean regret per round R_T/T. Markers average 100 paired repetitions, with error bars indicating 95% confidence intervals. The mean maximum backlog is $\sigma_{\max} \approx 467$. Horizontal lines indicate full tracking ($C = \infty$).

Figure 3 shows that scheduling has a greater effect under bandit feedback. In OCO, the uniform-weight scheduler is close to full tracking even at small capacities, while the Pareto-type scheduler approaches this baseline as capacity increases. In BCO, uniform weighting gives consistently lower regret for both loss families. Larger capacities narrow the gap, but both methods remain above full tracking, and the difference between schedulers remains substantial for strongly convex losses. These observations are consistent with our analysis: nonuniform importance weights can amplify the noise in bandit gradient estimates.

H.1 Experimental setup

The domain is the unit Euclidean ball in $\mathbb{R}^5$, the horizon is $T = 10^6$, and the tracking capacities are $C \in \{20, 40, 80, 160\}$.

Delays and losses. We draw independent nonnegative random variables Z_t with tail

$$\mathbb{P}(Z_t \geqslant z) = (1 + z/360)^{-1.9}, \qquad z \geqslant 0, \qquad d_t = \min\{1 + \lfloor Z_t \rfloor, T - t\}.$$

The scale is calibrated to target mean delay 400 after horizon clipping. For each environment, we round the empirical 0.4-quantile of the delays upward to the next multiple of 25 to obtain a threshold d^*; this gives $d^* = 125$ for every reported seed. We then generate affine losses

$$f_t(x) = \tfrac{1}{2} + \langle a_t, x \rangle, \qquad a_t = \tfrac{1}{2} g_t \mathbb{I}\{d_t > d^*\}, \qquad g_t = 0.95 g^* + \sqrt{1 - 0.95^2}\, u_t,$$

where $g^* = \mathbf{1}/\sqrt{5}$ and u_t is a uniformly sampled unit vector orthogonal to g^*. Thus $f_t(x) \in [0, 1]$ on the domain, and only longer-delay rounds have nonzero gradients. The entire loss and delay sequence is generated before play. The directional bias 0.95 was selected on separate pilot seeds 0–9. Final results use seeds 100–199.

Strongly convex losses. For the second loss family, we use the same delays and directions and set

$$f_t^{\mathrm{SC}}(x) = \frac{3}{4} f_t(x) + \frac{1}{4} \|x\|^2.$$

Both terms lie in $[0, 1]$ on the unit ball, so their convex combination does too. Each loss is $\frac{1}{2}$-strongly convex, with Hessian $\frac{1}{2}I$. The directional signal remains confined to longer-delay rounds, while every round now also contributes the quadratic term.

Tracking and event order. The simulator processes positive-delay feedback arriving at time t before prediction at time t, with arrivals taking precedence over simultaneous timeout expirations. After prediction, a new positive-delay item is admitted only if a slot is free and its sampled lifetime is positive. Zero-delay feedback is observed immediately after its action without occupying a slot; horizon clipping produces such a delay only at the final round in this experiment. A positive simulator delay d therefore corresponds to delay $d-1$ under the manuscript's end-of-round feedback convention. Across the 100 seeds, $\sigma_{\max}$ has mean 466.67 and ranges from 457 to 485.

Schedulers. Bernoulli independently selects each positive-delay round with probability

$$p = \min\{1,\, 0.75C/\max\{1, \sigma_{\max}\}\},$$

and keeps admitted rounds until feedback arrives. The Pareto baseline samples a nonnegative integer tracking lifetime L with

$$\mathbb{P}(L \geqslant d) = (1 + d/b_C)^{-1}, \qquad d \in \mathbb{Z}_{\geqslant 0},$$

and clips sampled lifetimes at T. We calibrate b_C numerically so that

$$\mathbb{E}[\min\{L, T\}] = \sum_{d=1}^{T} (1 + d/b_C)^{-1} = 0.75C.$$

A lifetime of zero rejects a positive-delay item. Both schedulers enforce the hard cap C. Retained feedback receives inverse-probability weight $1/p$ or $1 + d/b_C$, respectively; dropped feedback contributes zero. These weights correct proxy selection conditional on an available slot. Full tracking retains every round with weight one and no capacity restriction. The proxy distribution and occupancy calibration are empirical adaptations of the schedulers discussed in Section 4; the empirical Pareto baseline differs from the time-varying Pareto distribution in Corollary 4.4.

Learners and tuning. The experimental learners use the weighted-gradient feedback of DW-FTRL (Section 3.1) and the one-point bandit estimator of DW-FTBL (Section 3.2) for first-order and bandit feedback, respectively. Their updates are implemented as per-arrival projected-gradient steps with fixed parameters, as detailed below.

The OCO learner starts at zero and applies a projected gradient update for each arriving weighted gradient, with constant step size

$$\eta_{\mathrm{OCO}} = \frac{2}{\sqrt{T(1 + \bar{d})}}, \qquad \bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t.$$

For OCO, feedback is a_t for the affine losses and $\frac{3}{4}a_t + \frac{1}{2}x_t$ for the strongly convex losses, evaluated at the original played action x_t.

For BCO, the learner maintains a center y_t in the ball of radius $1 - \delta$ and plays $x_t = y_t + \delta v_t$, where v_t is uniform on the unit sphere. Its estimate is $(5/\delta) f_t(x_t) v_t$, with f_t denoting the chosen loss family, multiplied once by the retention weight. Each arrival triggers a projected update of the center using the raw loss and the full importance-weighted estimate. Define $q_C = 1 + \bar{d}/C$ for the constrained methods and $q_C = 1$ for full tracking. The fixed parameters within a BCO run are

$$\delta = \min\{1, \sqrt{5}\, q_C^{1/4} T^{-1/4}\}, \qquad \eta_{\mathrm{BCO}} = \frac{2}{\sqrt{T}\sqrt{\bar{d} + (5/\delta)^2 q_C}}.$$

All multiplicative tuning constants equal one. The same parameter rules are used for both loss families. The constants remain fixed across capacities, while δ and η_{BCO} follow the displayed dependence on C. Tuning uses the realized mean delay and, for Bernoulli retention, the realized maximum backlog.

Pairing and evaluation. The delays, loss directions, and bandit perturbation directions are shared across methods, capacities, and loss families within each seed. The underlying scheduler random streams are also reused across capacities and loss families. Full tracking is run separately for each loss family, and its outcomes are reused unchanged for every C. For $A = \sum_{t=1}^{T} a_t$, the exact static comparator for affine losses is $x^* = -A/\|A\|$ when $A \neq 0$ and $x^* = 0$ otherwise. For the strongly convex losses, it is $x^*_{\mathrm{SC}} = \Pi(-\frac{3A}{2T})$, where Π projects onto the unit ball.

Regret is computed using each learner's actual played actions, including the bandit perturbations. Figure 3 shows mean regret per round R_T/T. Results are averaged across the 100 seeds, with pointwise 95% Student-t confidence intervals using 99 degrees of freedom.

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 state the theoretical contributions and their assumptions. The regret guarantees in Section 5 and the synthetic experiments in Appendix H support the stated claims within this scope.

   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 paper states the assumptions on losses, delays, feedback, and tracking capacity. The Discussion and Future Work section discusses the gap in the delay dependence of the convex BCO bound and the restriction to oblivious adversaries. Appendix H.1 describes the synthetic setting and the empirical scheduler variants used in the experiments.

   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 with its required assumptions. Complete proofs are provided in the appendix, with the main paper giving the relevant definitions, theorem statements, and proof 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: [Yes]

Justification: Appendix H.1 specifies the synthetic loss and delay generation, scheduler and learner implementations, parameter choices, random seeds, pairing, and evaluation procedure. The publicly available code generates the experimental data and reproduces the capacity-ablation results and figure.

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: [Yes]

   Justification: Code is publicly available at `https://github.com/alex-rbch/cc-oco`, as linked in Appendix H. The repository includes dependencies, setup instructions, and commands to run all methods and baselines and regenerate the paper's experimental figure. All data are generated synthetically by the released code; no external datasets are required.

   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: [Yes]

   Justification: Appendix H.1 gives the domain, horizon, capacities, loss and delay distributions, scheduler calibration, update rules, learning rates, smoothing parameters, and static comparators. It distinguishes pilot seeds 0–9 from evaluation seeds 100–199 and specifies which randomness is shared across methods, capacities, and loss families.

   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: [Yes]

Justification: Figure 3 reports means across 100 paired repetitions, with error bars showing pointwise 95% Student-t confidence intervals using 99 degrees of freedom. Appendix H.1 describes the randomization and pairing. The intervals quantify uncertainty in the mean across seeds in the generated environments and algorithm randomness; their computation is included in the released plotting code.

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: [No]

   Justification: Experiments were conducted on a personal computer. Detailed hardware specifications and execution times are not reported.

   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 work 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 studies foundational online optimization methods with synthetic experiments and does not introduce a deployed system or an application with direct societal impact. Any broader impact is 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 with 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 experiments use author-generated synthetic data and author-implemented algorithms, with no external datasets, pretrained models, or third-party research code. Standard software dependencies are listed in the released repository.

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: [Yes]

Justification: The released experimental code includes a README describing installation, dependencies, reproduction commands, configurable parameters, output files, and simulator checks. The experimental design is documented in Appendix H.1. The data are synthetic and do not require participant consent.

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. The proposed schedulers, online learners, and synthetic evaluation do not use LLMs as an important, original, or non-standard component of the research methodology.

    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.