Decentralized Online Convex Optimization with Unknown Feedback Delays

Hao Qiu[1], **Mengxiao Zhang**[2], **Juliette Achddou**[3]

[1]Università degli Studi di Milano
[2]University of Iowa
[3]UMR 9189 - CRIStAL, Université de Lille, CNRS, Inria, Centrale Lille
hao.qiu@unimi.com, mengxiao-zhang@uiowa.edu, juliette.achddou@inria.fr

Abstract

Decentralized online convex optimization (D-OCO), where multiple agents within a network collaboratively learn optimal decisions in real-time, arises naturally in applications such as federated learning, sensor networks, and multi-agent control. In this paper, we study D-OCO under unknown, time- and agent-varying feedback delays. While recent work has addressed this problem (Nguyen, Kim Thang, and Trystram 2024), existing algorithms assume prior knowledge of the total delay over agents and still suffer from suboptimal dependence on both the delay and network parameters. To overcome these limitations, we propose a novel algorithm that achieves an improved regret bound of $\widetilde{\mathcal{O}}\Big(N\sqrt{d_{\text{tot}}} + N\sqrt{\frac{T}{\sqrt{1-\sigma_2}}}\Big)$, where d_{tot} denotes the average total delay across agents, N is the number of agents, and $1-\sigma_2$ is the spectral gap of the network. We also prove a lower bound showing that our upper bound is tight up to logarithmic factors. Our approach builds upon recent advances in D-OCO (Wan et al. 2024a), but crucially incorporates an adaptive learning rate mechanism via a decentralized communication protocol. This enables each agent to estimate delays locally using a gossip-based strategy without the prior knowledge of the total delay. We further extend our framework to the strongly convex setting and derive a sharper regret bound. Experimental results validate the effectiveness of our approach, showing improvements over existing benchmark algorithms.

Introduction

Decentralized online convex optimization (D-OCO) provides a powerful framework for distributed learning systems where multiple agents collaboratively optimize a global objective while processing local data streams. Specifically, in D-OCO, agents make sequential decisions based on local information and coordinate through peer-to-peer communication networks without relying on a central coordinator. This paradigm has become increasingly important in modern applications including federated learning (Kairouz et al. 2021), wireless sensor networks (Hosseini, Chapman, and Mesbahi 2013; Akbari, Gharesifard, and Linder 2015), real-time control systems (Lesage-Landry and Callaway 2020), and multi-agent robotic systems (Liu and Wu 2018), where centralized processing is either infeasible due to communication constraints or undesirable due to privacy concerns.

While immediate feedback is ideal, in practical distributed systems, local delays are ubiquitous and stem from factors such as fluctuating connectivity reliability, varying processing and computation times across heterogeneous devices, queuing latency in congested network links, or even delays introduced by human-in-the-loop feedback.

These delays can significantly degrade learning performance and raise fundamental challenges for algorithm design. While the impact of delays has been extensively studied in centralized online learning settings (Weinberger and Ordentlich 2002; Joulani, Gyorgy, and Szepesvári 2013), the interplay between decentralization and delayed feedback introduces unique complexities that remain less understood. Several works have considered delays in decentralized settings, but most assume either bounded time-invariant (Cao and Basar 2022) or known delays (Nguyen, Kim Thang, and Trystram 2024), which fail to capture the uncertainty and variability encountered in real-world systems. For example, in sensor networks, each node may incur delays both when acquiring measurements and when processing data (Rabbat and Nowak 2004; Olfati-Saber 2007). Recently, Nguyen, Kim Thang, and Trystram (2024) made progress by proposing a decentralized algorithm that handles arbitrary delays in D-OCO. However, their approach suffers from two limitations: (i) it requires prior knowledge of the total delay to set the learning rate appropriately, which is usually unavailable in practice, and (ii) even with this knowledge, their regret bounds suffer from suboptimal dependencies on both the total delay and network-dependent parameters. This raises a fundamental question:

Can we design decentralized online learning algorithms that adapt to unknown, time- and agent-varying delays while maintaining near-optimal regret guarantees?

In this paper, we answer this question affirmatively by developing novel decentralized online learning algorithms that achieve improved regret bounds under unknown, agent- and time-varying feedback delays. Specifically,

- For general convex losses, we derive an algorithm that achieves a regret bound of $\widetilde{\mathcal{O}}\big(N\sqrt{d_{\text{tot}}} + \frac{N\sqrt{T}}{(1-\sigma_2)^{1/4}}\big)$, where d_{tot} denotes the average total delay across agents,

N is the number of agents, T is the time horizon, and $1-\sigma_2$ is the spectral gap of the communication network.[1] Our algorithm is inspired by the recent advance in D-OCO (Wan et al. 2024a) but with an important adaptive learning rate mechanism combined with a decentralized communication protocol, where agents use gossip-based strategies to locally estimate delays without centralized coordination or prior knowledge of the total delay. Comparing to the results in Nguyen, Kim Thang, and Trystram (2024) whose regret bound is no better than $\mathcal{O}\left(\frac{N^2}{(1-\sigma_2)^2}\sqrt{d_{\text{tot}}}+\frac{\sqrt{N^3T}}{1-\sigma_2}\right)$, our result not only improves upon the regret bound dependency on N and σ_2 but also eliminates the need for prior knowledge of delays.[2] We further complement with a $\Omega(N\sqrt{d_{\text{tot}}}+N\sqrt{T}/(1-\sigma_2)^{1/4})$ lower bound, demonstrating that our algorithm's dependence on N, d_{tot}, T, and $1-\sigma_2$ is tight up to logarithmic factors.

- We then consider the case where the loss functions are all strongly convex, and extend our framework to derive regret bounds of $\mathcal{O}\left(\frac{N}{\alpha}\delta_{\max}\ln T+\frac{N\ln N\ln T}{\alpha\sqrt{1-\sigma_2}}\right)$, where α is the strong convexity parameter and $\delta_{\max}$ is the maximum number of missing observations averaged over agents, showing that strong convexity enables improved regret guarantee under D-OCO with delayed feedback. We remark again that our algorithm does not require the knowledge of the total delay.
- Finally, we implement extensive experiments on various network structures and loss functions, demonstrating superior empirical performances of our proposed algorithms comparing to existing baselines.

Related Works

Decentralized online convex optimization D-OCO is a framework in which multiple agents cooperatively solve an online optimization problem over a network, without relying on a central coordinator. Early foundational work in decentralized optimization focused on offline settings, leveraging techniques from gossip algorithms — originally used to achieve consensus to enable distributed optimization (Boyd et al. 2011; Nedic and Ozdaglar 2009). The first formal treatment of the online counterpart was given by (Hosseini, Chapman, and Mesbahi 2013), who analyzed a dual averaging algorithm and established sublinear regret guarantees. Specifically, they showed that a regret bound of $\mathcal{O}(N^{5/4}\sqrt{T}/(1-\sigma_2)^{1/2})$ is achievable, where σ_2 is the second highest singular value of the communication matrix W, whose definition is shown in later sections. Since then, various algorithmic approaches have been developed, including decentralized mirror descent (Shahrampour and Jadbabaie 2018), for which a similar regret rate is provable and accelerated gossiping for D-OCO (Wan et al. 2024a). The method from (Wan et al. 2024a) notably improves the previous regret bound by a factor of $(1-\sigma_2(W))^{-1/4}N^{1/4}/\sqrt{\log(N)}$. The D-OCO framework has seen various extensions, including work on settings with dynamic networks (Hosseini, Chapman, and Mesbahi 2016; Lei et al. 2020). For a comprehensive overview of such developments, we refer the reader to the recent monograph by Yuan et al. (2024).

Online learning with delayed feedbacks Our work is closely related to the literature on online learning with delayed feedback, initiated by Weinberger and Ordentlich (2002). They considered the setting with uniform, known per-round delays and proposed a general reduction to non-delayed online learning. Subsequent studies extended these results to handle non-uniform delays (Joulani, Gyorgy, and Szepesvári 2013). Various aspects of delayed feedback have been explored, including adaptive regret guarantees (Joulani, György, and Szepesvári 2016), diverse delay structures (Gatmiry and Schneider 2024; Bar-On and Mansour 2025; Ryabchenko, Attias, and Roy 2025), and limited-feedback scenarios (Cesa-Bianchi et al. 2016; Cella and Cesa-Bianchi 2020; Zimmert and Seldin 2020; Lancewicki, Rosenberg, and Mansour 2022; Van der Hoeven et al. 2023).

D-OCO with delayed feedbacks In D-OCO with local feedback delays, agents receive the gradient of their decision after a certain lag. For settings involving time-invariant but agent-specific delays, Cao and Basar (2022) proposed an online decentralized gradient descent algorithm, accommodating such delays for both convex and strongly convex loss functions. Meanwhile, Mao et al. (2025) studied online distributed convex optimization under delayed feedback within unbalanced, time-varying communication graphs. Additionally, Xiong et al. (2023a,b) considered D-OCO and its bandit counterpart with event-triggered communications and delayed feedback. For the more challenging setting with time- and agent-varying delays, Nguyen, Kim Thang, and Trystram (2024) introduced a projection-free approach; however, their method relies on prior knowledge of the cumulative delay to appropriately set the learning rate. Beyond local feedback delays, communication delay is also considered in the literature. For example, Tsianos and Rabbat (2012) analyzed distributed optimization under fixed communication delays.

Preliminary

Throughout this paper, we denote the set $\{1,2,\ldots,m\}$ for some positive integer m by $[m]$ and let $\mathbf{1}$ be an all-one vector in an appropriate dimension. For a vector $v\in\mathbb{R}^m$, denote its i-th entry by $v(i)$ and for a matrix $M\in\mathbb{R}^{m\times n}$, denote its (i,j)-th entry by $M(i,j)$. In this section, we introduce the preliminary of our problem.

Protocol In our model of decentralized online convex optimization, agents are organized in a communication network defined by a connected and undirected graph $G=(V,E)$. The node set $V=[N]$ corresponds to the N agents, and E denotes the set of edges indicating permissible communication among agents. We use V and $[N]$ interchangeably throughout the paper. Each agent $u\in V$ is associated

[1] A formal definition of the communication network is introduced in Section Preliminary. We use $\widetilde{\mathcal{O}}(\cdot)$ to hide logarithmic factors of N and T.

[2] We also remark that Nguyen, Kim Thang, and Trystram (2024) requires β-smoothness for the loss functions for all agents, which is not assumed in our work.

with an arbitrary and unknown sequence of *local loss functions* $f_1(u,\cdot), f_2(u,\cdot), \dots f_T(u,\cdot)$ decided by an adversary, where $f_t(u,\cdot) : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}$ for $t \in [T]$ has a bounded feasible domain and is L-Lipschitz with respect to ℓ_2 norm.

Assumption 1 (Bounded domain)**.** *The common decision space $\mathcal{X} \subseteq \mathbb{R}^n$ is convex and closed. Let $D = \sup_{x,y\in\mathcal{X}} \|x-y\|_2$ be the diameter of $\mathcal{X}$ and $\mathbf{0} \in \mathcal{X}$.*

Assumption 2 (Lipschitzness)**.** *For every $t \in [T]$, we assume that $f_t(u,\cdot)$ is convex and L-Lipschitz with respect to $\|\cdot\|_2$ for all $u \in V$.*

The learning protocol of D-OCO with time- and agent-varying feedback delays is defined as follows. The interaction between the agents and the environment proceeds in T rounds. At each round t, each agent $u \in V$ selects an action $x_t(u) \in \mathcal{X}$ simultaneously and suffers a loss $f_t(u, x_t(u))$. For each agent u, instead of observing the gradient $\nabla f_t(u, x_t(u))$ immediately in the standard OCO setting, agent u observes this gradient information at the end of round $t + d_t(u)$. Without loss of generality, we assume that $t + d_t(u) \le T$, for all $u \in V,\ t \in [T]$ since any feedback received at round T will never be used in the learning process. In addition, here we consider the *anonymous delayed feedback* setting where the agent does not know the time stamp of the received gradient. After receiving feedback, each agent shares the information it received with its neighbors in G. Each agent's goal is to minimize their regret defined as follows, which is in terms of the *global loss function* $\sum_{v\in V} f_t(v,x)$:

$$\mathrm{Reg}_T(u) \triangleq \max_{x\in\mathcal{X}} \Big(\sum_{t=1}^{T} \sum_{v\in V} (f_t(v, x_t(u)) - f_t(v,x)) \Big). \tag{1}$$

We also define $\mathrm{Reg}_T \triangleq \max_{u\in V} \mathrm{Reg}_T(u)$.

It remains to introduce how agents communicate their information with each other in this network. Specifically, following previous works of D-OCO (Yan et al. 2012; Hosseini, Chapman, and Mesbahi 2013; Wan et al. 2024a), we consider a gossip mechanism, or more specifically, an accelerated one defined as follows. This mechanism is defined by a communication matrix W constructed based on G.

Definition 3. *A matrix $W \in [0,1]^{N\times N}$ is a valid communication matrix with respect to $G = (V,E)$ if W satisfies that (i) $W(u,v) = 0$ if $u \ne v$ and $(u,v) \notin E$; W is symmetric and doubly-stochastic meaning that (ii) $W(u,v) \ge 0,\ \forall u,v \in V$; (iii) $W(u,v) = W(v,u),\ \forall u,v \in V$ (iv) $\sum_{v\in V} W(u,v) = 1,\ \forall u \in V$. Consequently, a valid communication W is positive semi-definite with $0 \le \sigma_2(W) < 1$ where $\sigma_2(W)$ is the second-largest eigenvalue of W.*

A typical construction of this matrix is as follows:

$$W = I_N - c \cdot \mathrm{Lap}(G), \tag{2}$$

where $I_N \in \mathbb{R}^{N\times N}$ denotes the identity matrix and $\mathrm{Lap}(G)$ denotes the Laplacian of the graph G with $\mathrm{Lap}(G)(i,i) = \deg(i)$ for all $i \in V$, $\mathrm{Lap}(G)(i,j) = -1$ if $i \ne j, (i,j) \in E$, and $\mathrm{Lap}(G)(i,j) = 0$ if $i \ne j, (i,j) \notin E$. c is a certain constant such that $0 < c \le 1/\sigma_1(\mathrm{Lap}(G))$, with $\sigma_1(\mathrm{Lap}(G))$ being the largest eigenvalue of the Laplacian $\mathrm{Lap}(G)$. In particular, building row $W(u,\cdot)$ defined in Equation (2) only requires knowing agent u's direct neighbors.

Based on this communication matrix W, whose u-th row is given to each agent u at the beginning of the learning process, the gossip communication process is defined as follows. Suppose there are N vectors $\{x(u)\}_{u\in V}$ for each agent where $x(u) \in \mathbb{R}^n$ represents the information agent u wants to communicate. In the context of D-OCO, this information can correspond to various quantities such as predictions (Shahrampour and Jadbabaie 2018) or loss gradients (Hosseini, Chapman, and Mesbahi 2013). In order to approximate the averaged vector $\bar{x} = \frac{1}{N}\sum_{u\in V} x(u)$, Liu and Morse (2011) considers the following accelerated gossip process:

$$x^{k+1}(u) = (1+\theta) \sum_{v\in\mathcal{N}_u} W(u,v) x^k(v) - \theta x^{k-1}(u), \tag{3}$$

for $k \ge 0$ where $x^0(u) = x^{-1}(u) = x(u)$ for all $u \in V$, $\mathcal{N}_u = \{v : (u,v) \in E\} \cup \{u\}$ the set of neighbors of u according to G, and $\theta > 0$ is the mixing coefficient. Let $X^k \in \mathbb{R}^{N\times n}$ be a concatenation of $\{x^k(u)\}_{u\in V}$ and $\bar{X} = \bar{x}\mathbf{1}^\top$. Ye et al. (2023) shows that X^k converges to $\bar{X}$ in a linear rate.

Proposition 4 (Proposition 1 in Ye et al. (2023))**.** *The iterations of* (3) *with $\theta = \left(1 + \sqrt{1-\sigma_2^2(W)}\right)^{-1}$ ensure that*

$$\left\|X^k - \bar{X}\right\|_F \le \sqrt{14} b^k \left\|X^0 - \bar{X}\right\|_F$$

for any $k \in \mathbb{N}$, where $b = \left(1 - (1 - 1/\sqrt{2})\sqrt{1-\sigma_2(W)}\right)$ and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

Other Notations Let $\mathbf{0}$ be an all-zero vector in an appropriate dimension. For each agent $u \in V$, define set $o_t(u) = \{\tau \in \mathbb{N} : \tau + d_\tau(u) < t\} \subseteq [t-1]$ to be the set of rounds for agent u whose gradients are observed before round t, and let $m_t(u) = [t-1] \setminus o_t(u)$ be the set of rounds for agent u whose observation is yet to be received at the beginning of round t. Define $\delta_{\max} = \max_{t\in[T]} \frac{1}{N}\sum_{u\in V} |m_t(u)|$ to be the maximum number of per-round missing observations averaged over all agents and $d_{\text{tot}} = \frac{1}{N}\sum_{t\in[T]}\sum_{u\in V} d_t(u)$ to be the total delay averaged over all agents.

D-OCO with General Convex Loss Functions

In this section, we study the setting where the loss functions for each agent at each round are convex. We first consider the case where the total delay d_{tot} is known and propose an algorithm that achieves an $\widetilde{\mathcal{O}}\left(N\sqrt{d_{\text{tot}}} + \frac{N\sqrt{T}}{(1-\sigma_2(W))^{1/4}}\right)$ regret guarantee. We then extend this approach to the more realistic case where d_{tot} is unknown, using a specific adaptive learning rate tuning. Finally, we provide a lower bound of $\Omega\left(N\sqrt{d_{\text{tot}}} + \frac{N\sqrt{T}}{(1-\sigma_2(W))^{1/4}}\right)$, showing that our upper bound is tight in its dependence on T , d_{tot}, and $1 - \sigma_2(W)$.

Algorithm 1: Accelerated Decentralized Follow the Regularized Leader with Delayed Feedback (AD-FTRL-DF) for Agent u.

Initialize: $x_1(u) = z_1^{-1}(u) = z_1^0(u) = \mathbf{0}$.
for $s = 1, 2, \ldots, T/B$ **do**
Define $\mathcal{T}_s = \{(s-1)B+1, \ldots, sB\}$
for $t \in \mathcal{T}_s$ **do**.
Play $x_s(u)$ and set $k \leftarrow t - (s-1)B - 1$.
Update $z_s^{k+1}(u)$ using accelerated gossiping:

$$z_s^{k+1}(u) = (1+\theta)\sum_{v\in V} W(u,v) z_s^k(v) - \theta z_s^{k-1}(u). \tag{4}$$

Send $z_s^{k+1}(u)$ to every neighbor $v \in \mathcal{N}_u$.
end for
Compute $x_{s+1}(u)$ for next block as follows:

$$x_{s+1}(u) = \operatorname*{argmin}_{x\in\mathcal{X}} \langle z_s^B(u), x\rangle + \frac{1}{\eta_{s+1}(u)}\|x\|_2^2. \tag{5}$$

Aggregate gradients observed during the block:

$$y_s(u) = \sum_{\tau \in o_{sB+1}(u)\setminus o_{(s-1)B+1}(u)} g_\tau(u),$$

with $g_\tau(u) \triangleq \nabla f_\tau(x_{s(\tau)}(u))$, $s(\tau)$ is the block τ lies in.
Compute $z_{s+1}^{-1}(u)$ and $z_{s+1}^0(u)$ for next block:

$$z_{s+1}^{-1}(u) = z_s^{B-1}(u) + y_s(u),$$
$$z_{s+1}^{0}(u) = z_s^{B}(u) + y_s(u).$$

end for

Non-Adaptive Algorithm with Known Total Delay

When the total delay is known, our algorithm is built upon the algorithm proposed in Wan et al. (2024a), whose idea is to incorporate the accelerated gossiping process into a blocking update mechanism to estimate the gradient of the global loss function. Specifically, the algorithm operates in blocks of size B. Without loss of generality, we assume that T/B is an integer such that each block contains exact B time steps. Following Wan et al. (2024a), within each block $s \in [T/B]$, every agent u uses a fixed decision $x_s(u)$ and iteratively updates an auxiliary variable $z_s^{k+1}(u)$ using the accelerated gossip procedure defined in Equation (4). From a high level, $z_s^{k+1}(u)$ aims to approximate the gradient of the global loss function collected from all previous epochs. The parameters θ and B are chosen based on the spectral gap of the communication matrix W, specifically:

$$\theta = \frac{1}{1+\sqrt{1-\sigma_2^2(W)}}, \quad B = \left\lceil \frac{\sqrt{2}\ln(N\sqrt{14N})}{(\sqrt{2}-1)\sqrt{1-\sigma_2(W)}} \right\rceil. \tag{6}$$

After completing all iterations within block s, each agent updates her decision for the next block by solving a Follow-the-Regularized-Leader problem Equation (5) with learning rate $\eta_s(u)$. Then, different from Wan et al. (2024a) which aggregates the received gradient within this block, due to the feedback delay, we compute $y_s(u)$ which only aggregates all gradients $g_\tau(u)$ received during block s. This is formalized through the difference set $o_{sB+1}(u) \setminus o_{(s-1)B+1}(u)$, which captures newly received gradients within the block. Finally, we compute the first two iterates of the subsequent block using the prior iterates and the aggregated gradient $y_s(u)$. In the absence of delay, our algorithm exactly recovers the algorithm proposed in Wan et al. (2024a).

The pseudo code of our algorithm is formally shown in Algorithm 1 and the following theorem shows that our algorithm achieves $\mathcal{O}(N\sqrt{d_{\text{tot}}} + N\sqrt{T}/(1-\sigma_2(W))^{1/4})$ when $\eta_s(u)$ is fixed over all blocks and is dependent on d_{tot}.

Theorem 5. *Assume each agent $u \in V$ runs an instance of Algorithm 1 with a valid communication matrix W, parameters θ and B defined in Equation (6), and a fixed learning rate*

$$\eta_s(u) = \eta = \frac{D}{L\sqrt{d_{tot}+BT}}, \quad \forall\, s \in [T/B]. \tag{7}$$

Then, under Assumption 1 and 2, the regret is bounded as

$$\mathrm{Reg}_T = \mathcal{O}\Big(DLN\Big(\sqrt{d_{tot}} + \frac{\sqrt{T\ln N}}{(1-\sigma_2(W))^{1/4}}\Big)\Big).$$

Two remarks are as follows. First, note that Nguyen, Kim Thang, and Trystram (2024) considered the exact same case where d_{tot} is known and obtain a regret bound no better than $\mathcal{O}\left(\frac{N^2}{(1-\sigma_2)^2}\sqrt{d_{\text{tot}}} + \frac{N\sqrt{N}}{1-\sigma_2}\sqrt{T}\right)$. Comparing to their results, our result not only achieves a better dependency on the spectral gap $1-\sigma_2(W)$ and the number of agents N, but also shows that the effects of the delay and those of the network topology can be *decoupled*. Specifically, the portion of the regret that does not depend on the delay scales with $N/(1-\sigma_2(W))^{1/4}\sqrt{T}$ in Theorem 5 instead of $N\sqrt{N}/(1-\sigma_2(W))\sqrt{T}$ in their bound. For the delay related term, our bound *does not* depend on the spectral gap $1-\sigma_2(W)$ while theirs suffer from a suboptimal $1/(1-\sigma_2(W))^2$ dependency. Specifically, our result also improves the dependency on N upon the $\mathcal{O}(N\sqrt{d_{\text{tot}}} + \frac{N^{1.5}\sqrt{T}}{1-\sigma_2(W)})$ achieved by Cao and Basar (2022), where delays are time-invariant and agent-specific, i.e, $d_t(u) = d(u)$ for all $t \in [T]$. Moreover, our upper bound matches the lower bound up to logarithmic factors, as will be shown later. In addition, our bound also recovers the regret bound proven in Wan et al. (2024a) when $d(u) = 0$ for all $u \in V$.

Proof Sketch The full proof of Theorem 5 is deferred to the Appendix and we introduce the proof sketch in this section. With some calculation we decompose the regret for agent u as follows:

$$\mathrm{Reg}_T(u) \le \underbrace{\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in[N]} \langle g_t(v), \bar{x}_s - x^*\rangle}_{\spadesuit}$$

$$+ BL \underbrace{\sum_{s=1}^{T/B} \sum_{v \in [N]} \mathcal{O}\left(\|x_s(v) - \bar{x}_s\|_2 + \|x_s(u) - \bar{x}_s\|\right)}_{\clubsuit},$$

where

$$\bar{x}_s = \operatorname*{argmin}_{x \in \mathcal{X}} \{\langle \sum_{v \in V} \sum_{\tau \in o_{(s-1)B+1}(v)} g_\tau(v), x\rangle + \frac{N}{\eta}\|x\|_2^2\}$$

denotes the FTRL decision assuming that agent u receives all agents' gradients that have been observed up to time t and we use η to represent $\eta_s(u)$ since $\eta_s(u)$ is fixed over all agents and blocks. Intuitively, ♠ accounts for the regret incurred by the agent if she only suffers from the delayed feedback, while ♣ accounts for the regret incurred due to the communication among the network.

To bound ♠, following a classic analysis in online learning with delayed feedback, we further split ♠ into the regret of the decision assuming no feedback delay and the distance between the decisions with and without feedback delay. With some rather standard calculations, the first part can be bounded by $\mathcal{O}(ND^2/\eta + \eta BNL^2T)$ while the second term can be bounded by $\mathcal{O}(\eta NL^2(d_{\text{tot}} + BT))$.

To bound ♣, we analyze the effect of gossip-based averaging. While agents can not locally receive the true global gradient, using accelerated gossip, the disagreement between local and average quantities decays exponentially in B as shown by Proposition 4. Specifically, we show that for any agent $v \in [N]$, $\sum_{s=1}^{T/B} \|x_s(v) - \bar{x}_s\|_2$ is bounded by $\mathcal{O}(\eta TL)$, which is the main technical part of the proof and require an involved analysis. Finally, picking η optimally leads to our final bound.

Adaptive Algorithm with Unknown Total Delay

The main issue with the algorithm described above is that the learning rate choice $\eta_s(u)$ relies on the unknown total delay d_{tot}. To illustrate the difficulty of adaptively tuning the learning rate with respect to the total delay in D-OCO, consider the single-agent setting, where it is indeed possible to adjust the learning rate dynamically by tracking the cumulative number of the agent's own missing observations (McMahan and Streeter 2014; Gyorgy and Joulani 2021). In contrast, in the decentralized setting, each agent cannot directly observe the number of gradients missed by other agents, and thus cannot directly compute the global cumulative delay. However, note that $d_{\text{tot}} = \frac{1}{N}\sum_{u \in [N]}\sum_{t=1}^{T}|m_t(u)|$. Therefore, if each agent additionally communicates their own number of missing observations to others through a gossiping protocol, every agent can well estimate the total number of averaged missing observations, leading to an estimation of d_{tot}.

Specifically, each agent still runs an instance of Algorithm 1 to perform the decision update and track the average gradients under delay. In addition, each agent also runs an instance of Algorithm 2 in parallel to compute the learning rate by gossiping the number of their own missing observations with their neighbors. The algorithm is formally shown in Algorithm 2. From a high level, Algorithm 2 closely mirrors the accelerated gossip routine of Algorithm 1, but instead focuses on gossiping the cumulative number of missing observations. Concretely, Algorithm 2 still goes in blocks and updates the auxiliary variable ζ_s^k using the accelerated gossiping, which can be viewed as an approximation of the cumulative missing observations averaged till block $s-1$. The learning rate $\eta_{s+1}(u)$ is then computed by replacing the exact total delay d_{tot} used in Equation (7) by this local estimate till block $s-1$ as shown in Equation (8). At the end of the epoch s, similar to Algorithm 1, we update the first two iterates ζ_{s+1}^{-1} and ζ_{s+1}^{-1} of the subsequent block by adding the number of missing observations at the end of block s to ζ_s^{B-1} and ζ_s^B. This finishes our algorithm for adaptive learning rate tuning. Each agent u is then supposed to run Algorithm 1 alongside Algorithm 2 (with the same θ and B described in Equation (6)) to use $\eta_s(u)$ computed in Equation (8) to update $x_{s+1}(u)$. The following theorem shows that with this adaptive learning rate tuning, we achieve $\widetilde{\mathcal{O}}(N\sqrt{d_{\text{tot}}} + N\sqrt{T}/(1-\sigma_2(W))^{1/4})$ without knowing d_{tot}.

Algorithm 2: Accelerated Gossip Routine for the Adaptive Learning Rate for Agent u

---

Initialize: $\eta_1(u) = \frac{D}{L\sqrt{BT+3B^2}}$, $\zeta_1^{-1}(u) = \zeta_1^0(u) = 0$.

for $s = 1, 2, \ldots, T/B$ **do**

  for $t = (s-1)B+1, \ldots, sB$ **do**

    $k \leftarrow t - (s-1)B - 1$.

    Update $\zeta_s^{k+1}(u)$ using accelerated gossiping:

$$\zeta_s^{k+1}(u) = (1+\theta)\sum_{v \in V} W(u,v)\zeta_s^k(v) - \theta z_s^{k-1}(u).$$

    Send $\zeta_s^{k+1}(u)$ to every neighbor $v \in \mathcal{N}_u$.

  end for

  Count missing observations at the end of the block

$$q_s(u) = |m_{sB+1}(u)|.$$

  Update

$$\eta_{s+1}(u) = \frac{D}{L\sqrt{BT + B \cdot \zeta_s^B(u) + 3sB^2}}. \tag{8}$$

  Compute first iterates for next block:

$$\zeta_{s+1}^{-1}(u) = \zeta_s^{B-1}(u) + q_s(u),$$
$$\zeta_{s+1}^{0}(u) = \zeta_s^{B}(u) + q_s(u).$$

end for

Theorem 6. *Assuming each agent $u \in [N]$ runs an instance of Algorithm 2 with a valid communication matrix W and parameters θ and B defined in Equation (6) together with an instance of Algorithm 1 parametrized by the same W, θ and B and using $\eta_s(u)$ computed by Algorithm 2. Then, under Assumption 1 and 2, the regret is bounded as*

$$\text{Reg}_T = \widetilde{\mathcal{O}}\left(DLN\left(\sqrt{d_{tot}} + \frac{\sqrt{T}}{(1-\sigma_2(W))^{1/4}}\right)\right).$$

The proof of Theorem 6 is provided in the Appendix. We emphasize that our analysis is non-trivial, which includes

(i) a careful bounding on the gossip-based estimation error of the adaptive learning rate compared to the optimal rate defined with respect to d_{tot}, and (ii) a more involved analysis of the FTRL updates, particularly due to possibly non-decreasing learning rates $\eta_s(u)$.

Lower bound

Finally, we complement our obtained upper bounds with the following $\Omega(N\sqrt{T}/(1-\sigma_2(W))^{1/4}+N\sqrt{d_{\text{tot}}})$ lower bound.

Theorem 7. *Let d be the constant feedback delay suffered by all agents $u \in [N]$ in the network. Then, there exists a graph $G = ([N], E)$, with $N = 2(M+1)$ where M is an even integer, and a sequence of L-Lipschitz loss functions $\{f_1(1,\cdot),\ldots,f_1(N,\cdot)\},\ldots,\{f_T(1,\cdot),\ldots,f_T(N,\cdot)\}$ such that any algorithm has to suffer regret at least:*

$$\mathrm{Reg}_T \; = \Omega\left(DLN\left(\sqrt{T}/(1-\sigma_2(W))^{1/4}+\sqrt{dT}\right)\right),$$

where $W = I - \frac{1}{\sigma_1(Lap(G))}\cdot Lap(G)$.

Compared to this lower bound, our obtained upper bounds are optimal in the dependence on T, $1-\sigma_2(W)$, and d_{tot}, though there is still a gap of polynomial factors in the number of agents N. We provide a proof sketch here and the full proof is deferred to Appendix. Our proof is adapted from the construction in Wan et al. (2024a) which considered a carefully designed problem instance where the global loss is supported on one half of the graph, while the remaining half consists of agents with identically zero local loss functions. Focusing on an agent u in the latter group, we observe that its optimization problem effectively reduces to an instance of online linear optimization (OLO) with feedback delay. The total delay experienced by agent u in this setting consists of the constant delay d, combined with a graph-dependent communication delay due to the network structure. The remaining proof builds on standard lower bound analysis for centralized OLO with delayed feedback.

D-OCO with Strongly-Convex Loss Functions

In this section, we consider the case where all loss functions satisfy α-strongly convexity defined as follows.

Assumption 8 (strong convexity)**.** *For every $t \le T$ and $v \in V$, we assume that $f_t(v,\cdot)$ is α-strongly convex: $\forall x, y \in \mathcal{X}$*

$$f_t(v,y) \ge f_t(v,x) + \langle \nabla f_t(v,x), y-x\rangle + \frac{\alpha}{2}\|y-x\|_2^2.$$

In order to show an improved regret bound when losses are strongly convex in D-OCO with feedback delay, following the algorithm proposed in Wan et al. (2024a) for strongly convex functions, we propose our algorithm AD-FTRL-DF-SC outlined in Algorithm 3. Compared to AD-FTRL-DF shown in Algorithm 1, there are two key differences. First, the cumulative gradient $y_s(u)$ are replaced by $y_s^+(u)$, which includes an additional $-\alpha B x_s(u)$ term (Equation (9)); second, we do not need to apply a gossip-based communication among agents to tune the learning rate adaptively but only need $\eta_{s+1}(u) = \frac{2}{\alpha s B}$ for all $u \in [N]$. The following theorem shows that Algorithm 3 achieves $\mathcal{O}((N\delta_{\max} + N\ln N/\sqrt{1-\sigma_2(W)})(\ln T/\alpha))$ regret.

Algorithm 3: Accelerated Decentralized Follow the Regularized Leader with Delayed Feedback under Strong Convexity (AD-FTRL-DF-SC) for Agent u.

Initialize: $x_1(u) = z_1^{-1}(u) = z_1^0(u) = \mathbf{0}$
  for $s = 1, 2, \ldots, T/B$ **do**
    $\eta_{s+1} = \frac{2}{\alpha s B}$
    for $t = (s-1)B+1, \ldots, sB$ **do**
      Play $x_s(u)$ and set $k \leftarrow t-(s-1)B-1$.
      Update $z_s^{k+1}(u)$ using accelerated gossiping:

$$z_s^{k+1}(u) = (1+\theta)\sum_{v\in V} W(u,v) z_s^k(v) - \theta z_s^{k-1}(u).$$

      Send $z_s^{k+1}(u)$ and $x_s(u)$ to every $v \in \mathcal{N}_u$.
    end for
    Compute action $x_{s+1}(u)$:

$$x_{s+1}(u) = \operatorname*{argmin}_{x\in\mathcal{X}} \langle z_s^B(u), x\rangle + \frac{1}{\eta_{s+1}}\|x\|_2^2.$$

    Compute augmented aggregated gradients $y_s^+(u)$:

$$y_s^+(u) = \sum_{\tau \in o_{sB+1}(u)\setminus o_{(s-1)B+1}(u)} g_\tau(u) - \alpha B x_s(u). \tag{9}$$

    Compute first iterates for next block:

$$z_{s+1}^{-1}(u) = z_s^{B-1}(u) + y_s^+(u),$$
$$z_{s+1}^{0}(u) = z_s^{B}(u) + y_s^+(u).$$

  end for

Theorem 9. *Assume each agent $u \in V$ runs an instance of AD-FTRL-DF-SC with a valid communication matrix W and parameters θ and B defined in* (6)*. Then, under Assumption 1, 2 and 8, the global regret is bounded as*

$$\mathcal{O}\left(\frac{N(\alpha DL + L^2)}{\alpha}\left(\delta_{\max} + \frac{\ln(N)}{\sqrt{1-\sigma_2(W)}}\right)\ln(T)\right),$$

where $\delta_{\max} = \max_{t\in[T]} \frac{1}{N}\sum_{u\in[N]} |m_t(u)|$. Moreover, when $d_t(u) = d(u)$ for all $t \in [T]$, define $\bar{d} \triangleq \frac{1}{N}\sum_{v\in V} d(v)$ and the global regret is bounded as

$$\mathcal{O}\left(\frac{N(\alpha DL + L^2)}{\alpha}\left(\bar{d} + \frac{\ln(N)}{\sqrt{1-\sigma_2(W)}}\right)\ln(T)\right).$$

The full proof is deferred to Appendix. To our knowledge, there are no previous results for D-OCO under strongly convex losses with time- and agent-varying delays. Several remarks are as follows. First, to interpret the delay-dependent term $\delta_{\max}$, it is not hard to see that $\delta_{\max} \le \frac{1}{N}\sum_{n\in[N]} \max_{t\in[T]} d_t(u)$, which is the maximum delay averaged over all agents. Following Qiu, Esposito, and Zhang (2025), we can also show that $\delta_{\max} \le \sqrt{N d_{\text{tot}}}$. Second, reducing to the case where the delay is time-invariant, we achieve an improved bound compared to

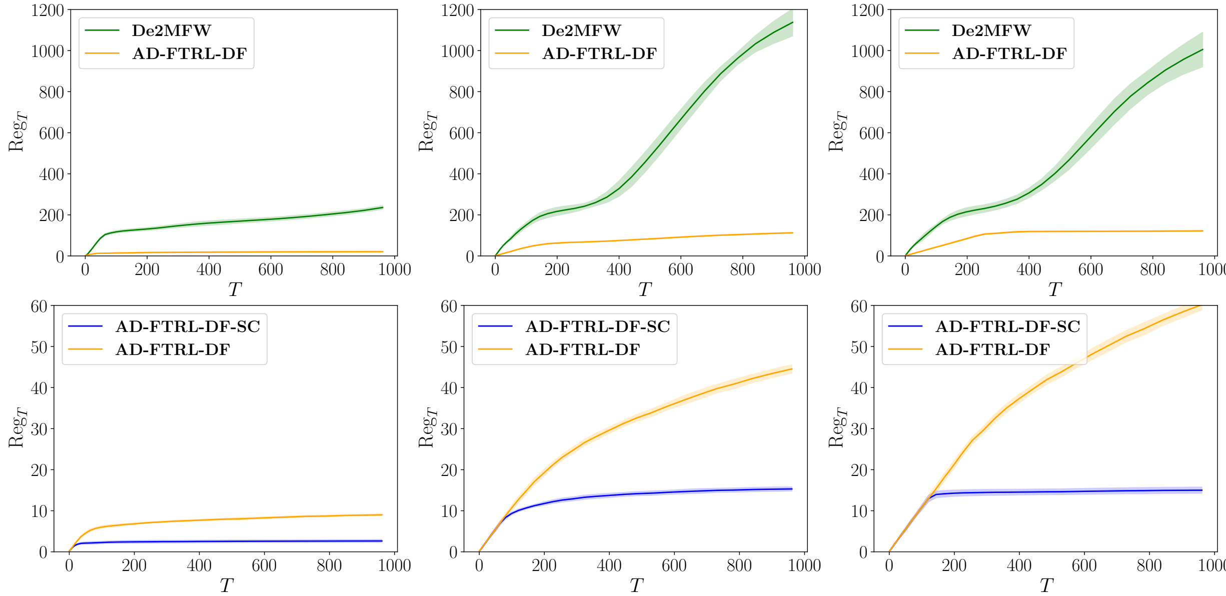

Figure 1: Comparison with relevant baselines across three network topologies—complete (left), grid (middle), and cycle (right)—under convex losses (top row) and strongly convex losses (bottom row).

Cao and Basar (2022), which obtained a regret bound of $\mathcal{O}(\frac{N\bar{d}}{\alpha}\ln T + \frac{N\sqrt{N}}{1-\sigma_2}\frac{\ln T}{\alpha})$. We also recover the bound proven in Wan et al. (2024a) when $d(u) = 0$ for all $u \in [N]$. Finally, in Appendix, we also provide a lower bound of $\Omega\left((d + 1/(1-\sigma_2(W))^{1/2}) \cdot N\alpha \ln(T/d)\right)$ when $d_t(u) = d$ for all $t \in [T]$ and $u \in [N]$, and all loss functions are αD-Lipschitz and α-strongly convex, showing that our upper bound is tight with respect to T, $\delta_{\max}$ (since $\delta_{\max} = d$ in this case), N and $1 - \sigma_2(W)$ up to logarithmic factors.

Numerical Experiments

In this section, we evaluate the performance of our proposed algorithms in the delayed D-OCO setting, using two representative sets of loss functions that capture the convex and strongly convex regimes, respectively.

Setting. To show the algorithms' performances under the general convex loss case, following the experiment setup used in Yuan, Proutiere, and Shi (2020), we define the local losses for all agents $v \in V$ as

$$f_t(v, x) = \tfrac{1}{2}\left(\langle w_t(v), x\rangle - y_t(v)\right)^2, \tag{10}$$

where each feature vector $w_t(v)$ has independent coordinates drawn uniformly from $[-1, 1]$. Labels are generated as follows: for $1 \le v \le N/2$, $y_t(v) = \varepsilon_t(v)$, and for the remaining agents, we have $y_t(v) = \langle w_t(v), \mathbf{1}\rangle + \varepsilon_t(v)$ with $\varepsilon_t(v)$ being zero-mean, unit-variance Gaussian noise clipped to $[-1, 1]$. For strongly convex losses, we augment each local loss with an ℓ_2-regularizer:

$$f_t(v, x) = \tfrac{1}{2}\left(\langle w_t(v), x\rangle - y_t(v)\right)^2 + \tfrac{1}{2}\|x\|_2^2. \tag{11}$$

We evaluate the performance of our algorithms and baselines on three network topologies with $N = 36$ nodes — the complete graph, in which all agents are connected to one another; the grid, in which agents are organized in a two-dimensional lattice and communicate with their immediate horizontal and vertical neighbors; and the cycle, where each agent v is connected to $v - 1$ and $v + 1$. We use Equation (2) with $c = 1/N$ to set the communication matrix W. Therefore, $1/(1 - \sigma_2(W))^{1/4}$ associated to each of the above topologies is respectively 1, 3.40 and 5.87. Each local delay $d_t(v)$ is independently and uniformly drawn from $\{0, 1, \ldots, 50\}$. All experiments are conducted over $T = 1000$ rounds, and each result is averaged over 20 independent trials. We set the agents' decision space to be $\mathcal{X} = \{x \in \mathbb{R}^{10}, \|x\|_2 \le 2\}$.

Baselines. For the general convex loss setting, we compare our algorithm AD-FTRL-DF (Algorithm 1) with adaptive learning rate tuning (Algorithm 2) against De2MFW (Nguyen, Kim Thang, and Trystram 2024). In the strongly convex loss setting, we compare our algorithm AD-FTRL-DF-SC (Algorithm 3) against AD-FTRL-DF (Algorithm 1) with adaptive learning rate tuning.

Results. Figure 1 shows the regret curve (with the shaded area the standard deviation over 20 trials) of our algorithms and the above baselines with losses defined in Equation (10) and Equation (11) for all three topologies. From the plots, we observe that for the losses defined in Equation (10), AD-FTRL-DF with an adaptive learning rate substantially outperforms De2MFW across all network topologies. In the strongly convex loss case, AD-FTRL-DF-SC achieves consistently lower regret than the baseline AD-FTRL-DF, which matches our theoretical guarantees. Comparing among different network topologies, for both convexity regimes, the regret is significantly higher with the grid and cycle graph compared to the one with the complete graph. This is consistent with the regret dependence on the reciprocal of a power of the spectral gap since the associated spectral gap for complete graph is smaller than that for grid and cycle graph.

Acknowledgments

Hao Qiiu acknowledges the financial support from the EU Horizon CL4-2021-HUMAN-01 research and innovation action under grant agreement 101070617, project ELSA (European Lighthouse on Secure and Safe AI)

References

Akbari, M.; Gharesifard, B.; and Linder, T. 2015. Distributed online convex optimization on time-varying directed graphs. *IEEE Transactions on Control of Network Systems*, 4(3): 417–428.

Bar-On, Y.; and Mansour, Y. 2025. Non-stochastic Bandits With Evolving Observations. In Kamath, G.; and Loh, P.-L., eds., *Proceedings of The 36th International Conference on Algorithmic Learning Theory*, volume 272 of *Proceedings of Machine Learning Research*, 204–227. PMLR.

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J.; et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1): 1–122.

Cao, X.; and Basar, T. 2022. Decentralized Online Convex Optimization With Feedback Delays. *IEEE Trans. Autom. Control.*, 67(6): 2889–2904.

Cella, L.; and Cesa-Bianchi, N. 2020. Stochastic bandits with delay-dependent payoffs. In *International Conference on Artificial Intelligence and Statistics*, 1168–1177. PMLR.

Cesa-Bianchi, N.; Gentile, C.; Mansour, Y.; and Minora, A. 2016. Delay and Cooperation in Nonstochastic Bandits. In Feldman, V.; Rakhlin, A.; and Shamir, O., eds., *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, 605–622. Columbia University, New York, New York, USA: PMLR.

Gatmiry, K.; and Schneider, J. 2024. Adversarial Online Learning with Temporal Feedback Graphs. In Agrawal, S.; and Roth, A., eds., *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, 4548–4572. PMLR.

Gyorgy, A.; and Joulani, P. 2021. Adapting to delays and data in adversarial multi-armed bandits. In *International Conference on Machine Learning*, 3988–3997. PMLR.

Hazan, E.; and Kale, S. 2014. Beyond the regret minimization barrier: optimal algorithms for stochastic strongly-convex optimization. *The Journal of Machine Learning Research*, 15(1): 2489–2512.

Hosseini, S.; Chapman, A.; and Mesbahi, M. 2013. Online distributed optimization via dual averaging. In *52nd IEEE Conference on Decision and Control*, 1484–1489. IEEE.

Hosseini, S.; Chapman, A.; and Mesbahi, M. 2016. Online distributed convex optimization on dynamic networks. *IEEE Transactions on Automatic Control*, 61(11): 3545–3550.

Joulani, P.; Gyorgy, A.; and Szepesvári, C. 2013. Online learning under delayed feedback. In *International conference on machine learning*, 1453–1461. PMLR.

Joulani, P.; György, A.; and Szepesvári, C. 2016. Delay-Tolerant Online Convex Optimization: Unified Analysis and Adaptive-Gradient Algorithms. In Schuurmans, D.; and Wellman, M. P., eds., *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, 1744–1750. AAAI Press.

Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2): 1–210.

Lancewicki, T.; Rosenberg, A.; and Mansour, Y. 2022. Learning adversarial markov decision processes with delayed feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7281–7289.

Lei, J.; Yi, P.; Hong, Y.; Chen, J.; and Shi, G. 2020. Online convex optimization over Erdos-Rényi random networks. *Advances in neural information processing systems*, 33: 15591–15601.

Lesage-Landry, A.; and Callaway, D. S. 2020. Dynamic and distributed online convex optimization for demand response of commercial buildings. *IEEE Control Systems Letters*, 4(3): 632–637.

Liu, J.; and Morse, A. S. 2011. Accelerated linear iterations for distributed averaging. *Annual Reviews in Control*, 35(2): 160–165.

Liu, J.; and Wu, J. 2018. *Multiagent robotic systems*. CRC press.

Mao, S.; Du, W.; Tian, Y.-C.; Gu, J.; and Tang, Y. 2025. Online Distributed Convex Optimization for Unbalanced Varying Graphs With Delayed Feedback. *IEEE Transactions on Circuits and Systems I: Regular Papers*.

McMahan, B.; and Streeter, M. 2014. Delay-tolerant algorithms for asynchronous distributed online learning. *Advances in Neural Information Processing Systems*, 27.

Nedic, A.; and Ozdaglar, A. 2009. Distributed Subgradient Methods for Multi-Agent Optimization. *IEEE Transactions on Automatic Control*, 54(1): 48–61.

Nguyen, T.-A.; Kim Thang, N.; and Trystram, D. 2024. Handling Delayed Feedback in Distributed Online Optimization: A Projection-Free Approach. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 197–211. Springer.

Olfati-Saber, R. 2007. Distributed Kalman filtering for sensor networks. In *2007 46th IEEE conference on decision and control*, 5492–5498. IEEE.

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

Qiu, H.; Esposito, E.; and Zhang, M. 2025. Exploiting Curvature in Online Convex Optimization with Delayed Feedback. In *International Conference on Machine Learning*. PMLR.

Rabbat, M.; and Nowak, R. 2004. Distributed optimization in sensor networks. In *Proceedings of the 3rd international symposium on Information processing in sensor networks*, 20–27.

Ryabchenko, A.; Attias, I.; and Roy, D. M. 2025. Capacity-Constrained Online Learning with Delays: Scheduling

Frameworks and Regret Trade-offs. *arXiv preprint arXiv:2503.19856v1*.

Shahrampour, S.; and Jadbabaie, A. 2018. Distributed Online Optimization in Dynamic Environments Using Mirror Descent. *IEEE Transactions on Automatic Control*, 63(3): 714–725.

Spielman, D. 2019. Spectral and algebraic graph theory. *Yale lecture notes, draft of December*, 4: 47.

Tsianos, K. I.; and Rabbat, M. G. 2012. Distributed dual averaging for convex optimization under communication delays. In *2012 American Control Conference (ACC)*, 1067–1072. IEEE.

Van der Hoeven, D.; Zierahn, L.; Lancewicki, T.; Rosenberg, A.; and Cesa-Bianchi, N. 2023. A Unified Analysis of Nonstochastic Delayed Feedback for Combinatorial Semi-Bandits, Linear Bandits, and MDPs. In Neu, G.; and Rosasco, L., eds., *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, 1285–1321. PMLR.

Wan, Y.; Wei, T.; Song, M.; and Zhang, L. 2024a. Nearly optimal regret for decentralized online convex optimization. In *The Thirty Seventh Annual Conference on Learning Theory*, 4862–4888. PMLR.

Wan, Y.; Wei, T.; Xue, B.; Song, M.; and Zhang, L. 2024b. Optimal and Efficient Algorithms for Decentralized Online Convex Optimization. arXiv:2402.09173.

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

Xiong, M.; Ho, D. W.; Zhang, B.; Yuan, D.; and Xu, S. 2023a. Distributed online mirror descent with delayed subgradient and event-triggered communications. *IEEE Transactions on Network Science and Engineering*, 11(2): 1702–1715.

Xiong, M.; Zhang, B.; Yuan, D.; Zhang, Y.; and Chen, J. 2023b. Event-triggered distributed online convex optimization with delayed bandit feedback. *Applied Mathematics and Computation*, 445: 127865.

Yan, F.; Sundaram, S.; Vishwanathan, S.; and Qi, Y. 2012. Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties. *IEEE Transactions on Knowledge and Data Engineering*, 25(11): 2483–2493.

Ye, H.; Luo, L.; Zhou, Z.; and Zhang, T. 2023. Multi-consensus decentralized accelerated gradient descent. *Journal of machine learning research*, 24(306): 1–50.

Yuan, D.; Proutiere, A.; and Shi, G. 2020. Distributed online linear regressions. *IEEE Transactions on Information Theory*, 67(1): 616–639.

Yuan, D.; Proutiere, A.; Shi, G.; et al. 2024. Multi-agent online optimization. *Foundations and Trends® in Optimization*, 7(2-3): 81–263.

Zimmert, J.; and Seldin, Y. 2020. An Optimal Algorithm for Adversarial Bandits with Arbitrary Delays. In Chiappa, S.; and Calandra, R., eds., *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, 3285–3294. PMLR.