Università degli Studi di Milano

Ph.D. Program in Computer Science

(XXXVIII Cycle)

Department of Computer Science

A thesis submitted for the degree of

Doctor of Philosophy

Distributed and Delayed Online Learning

Subject Area: INF/01

Author
Hao Qiu

Supervisor
Prof. Nicolò Cesa-Bianchi

Co-Supervisor
Prof. Wouter M. Koolen

PhD Coordinator
Prof. Roberto Sassi

Academic Year 2025–2026

To my parents

Abstract

This thesis investigates the design and analysis of distributed and delayed online learning algorithms. First, we introduce delayed online learning, where model updates rely on feedback arriving with variable delays. We study the online learning problem with curved losses and delayed feedback, designing algorithms that exploit loss curvature to achieve improved guarantees with delayed feedback. Furthermore, we demonstrate how intermediate observations can mitigate the deleterious effects of delayed feedback in settings with partial feedback (e.g., multi-armed bandits) by developing a meta-algorithm that achieves near-optimal regret with significantly reduced sensitivity to total delay. Second, we consider distributed online convex optimization over communication graphs, in which a network of agents cooperatively minimizes a global convex loss function expressed as the sum of local loss functions, using only neighbor-to-neighbor exchanges and local computation. We propose and rigorously analyze a distributed online convex optimization algorithm that accommodates both random communication and stochastic agent availability, where two agents communicate only when both are simultaneously active. We then present a unified algorithmic framework that simultaneously addresses network decentralization and delayed feedback in distributed online convex optimization. We design distributed online learning algorithms that adapt to unknown, time- and agent-varying delays while maintaining near-optimal regret guarantees. Finally, transitioning from adversarial to stochastic regimes, we extend this framework to distributed stochastic multi-armed bandit settings over random communication graphs. We derive improved regret bounds that combine the optimal centralized regret with a natural term depending on the graph's algebraic connectivity and edge probability.

Acknowledgments

My gratitude goes to my family for their love and support, to my supervisors Nicolò and Wouter, to all members of the LAILA and CWI machine learning groups, and to my coauthors over the past three years for their valued company. I would also like to thank the three reviewers of this thesis.

Preface

The contents of this dissertation are based on the following published papers or manuscripts under review, emerged from different collaborations with the respective co-authors during my PhD activities. Here is a comprehensive list divided into chapters:

- Chapter 2 is based on the published conference paper:

  > Hao Qiu, Emmanuel Esposito, and Mengxiao Zhang. Exploiting curvature in online convex optimization with delayed feedback. In *Forty-second International Conference on Machine Learning*, 2025a.

  In this paper, I have first co-authorship, and I was the main contributor to all the results therein.

- Chapter 3 is based on the published conference paper:

  > 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.

  In this paper, I have second authorship, and I was the main contributor to all experimental results therein.

- Chapter 4 is based on the under-review manuscript:

  > Juliette Achddou, Nicolò Cesa-Bianchi, and Hao Qiu. Distributed online optimization with stochastic agent availability. *arXiv preprint arXiv:2411.16477*, 2024.

  In this work, authors are in alphabetical order, and all authors equally contributed to the results therein.

- Chapter 5 is based on the under-review manuscript:

  > Hao Qiu, Mengxiao Zhang, and Juliette Achddou. Decentralized online convex optimization with unknown feedback delays. *preprint*, 2025b. To appear at *the 40th Annual AAAI Conference on Artificial Intelligence*.

  In this work, I have first authorship, and I was the main contributor to all the results therein.

- Chapter 6 is based on a forthcoming conference paper:

  > Jingyuan Liu, Hao Qiu, Lin Yang, and Mengfan Xu. Distributed multi-agent bandits over Erdős–Rényi random networks. *preprint*, 2025. To appear at *the 39th Annual Conference on Neural Information Processing Systems*.

In this paper, I have first co-authorship, and I was the main contributor to all the results therein.

The following published conference papers or manuscripts under review have also been the result of my activities during the PhD program:

- Dirk Van Der Hoeven, Ciara Pike-Burke, Hao Qiu, and Nicolò Cesa-Bianchi. Trading-off payments and accuracy in online classification with paid stochastic experts. In *International Conference on Machine Learning*, pages 34809–34830. PMLR, 2023.

Contents

Acknowledgments **iii**

Preface **v**

1 Introduction **1**
1.1 Online Learning . . . 1
1.2 Delayed Online Learning . . . 3
1.3 Distributed Online Learning . . . 4
1.4 Outline of the Thesis . . . 5

2 Exploiting Curvature in Online Convex Optimization with Delayed Feedback **9**
2.1 Introduction . . . 9
2.1.1 Related works . . . 11
2.2 Problem setting . . . 12
2.3 Delayed OCO with Strongly Convex Losses . . . 13
2.3.1 Regret Analysis . . . 14
2.4 Delayed OCO with Exp-concave Losses . . . 15
2.4.1 Regret Analysis . . . 16
2.5 Online Linear Regression with Delayed Labels . . . 18
2.6 Experiments . . . 20

3 Exploiting Intermediate Feedback in Multi-Armed Bandits with Delayed Feedback **23**
3.1 Introduction . . . 23
3.1.1 Related Works . . . 25
3.2 Problem Setting . . . 25
3.3 A Reduction to Standard Delayed Feedback . . . 26
3.4 Regret Analysis . . . 28
3.5 Lower Bounds . . . 32
3.6 Experiments . . . 34

4 Distributed Online Convex Optimization under Stochastic Agent Availability and Random Networks **39**
4.1 Introduction . . . 39
4.1.1 Related Works . . . 40
4.2 Problem Setting . . . 41
4.3 The `Gossip-FTRL` Algorithm . . . 43
4.4 Upper Bounds . . . 43
4.5 Lower Bound . . . 45
4.6 The Gossip Matrix . . . 46
4.7 Random Edges . . . 48
4.8 Experiments . . . 48

5 Distributed Online Convex Optimization with Delayed Feedback **53**
5.1 Introduction . . . 53
5.1.1 Related Works . . . 54
5.2 Preliminary . . . 55
5.3 DOCO with General Convex Loss Functions . . . 57
5.3.1 Non-Adaptive Algorithm with Known Total Delay . . . 57
5.3.2 Proof Sketch . . . 59
5.3.3 Adaptive Algorithm with Unknown Total Delay . . . 60
5.3.4 Lower bound . . . 61
5.4 DOCO with Strongly-Convex Loss Functions . . . 62
5.5 Experiments . . . 64

6 Distributed Stochastic Multi-Armed Bandits Over Random Networks **65**
6.1 Introduction . . . 65
6.1.1 Related Works . . . 67
6.2 Setting and Notations . . . 68
6.3 Algorithm: `Gossip Successive Elimination` . . . 69
6.4 Regret Analyses . . . 71
6.4.1 Upper Bound . . . 71
6.4.2 Lower Bound . . . 73
6.5 Experiments . . . 74

7 Summary and Discussion **77**

Bibliography **78**

Appendices **91**

A Proof Details for Chapter 2 **91**
A.1 Auxiliary results . . . 91
A.1.1 General results for the regret analysis . . . 91
A.1.2 Results for delay-related quantities . . . 92
A.2 Omitted details in Section 2.3 . . . 94
A.3 Omitted details in Section 2.4 . . . 97
A.4 Omitted details in Section 2.5 . . . 103
A.5 Online mirror descent for delayed OCO with strongly convex losses . . . 109
A.6 Additional experiments . . . 114

B Proof Details for Chapter 3 **117**
B.1 Auxiliary Results . . . 117
B.2 Omitted Details in Section 3.4 . . . 119
B.2.1 Total Effective Delay Bound . . . 119
B.2.2 Improved Regret for `DAda-Exp3` for Fixed δ . . . 119
B.2.3 Reduction to `DAda-Exp3` via `MetaBIO` . . . 120
B.2.4 Regret of `MetaBIO` . . . 123
B.2.5 Regret of `AdaMetaBIO` . . . 125
B.2.6 Expected Regret Analysis of `AdaMetaBIO` with `Tsallis-INF` . . . 127
B.3 Omitted Details in Section 3.5 . . . 129
B.4 Action-State Mappings and Loss Means Used in the Experiments . . . 131

C Proof Details for Chapter 4 **133**
C.1 Additional related works . . . 133
C.2 Additional remark on Algorithm 4.1 and Notation . . . 134
C.3 Preliminary results . . . 134
C.4 Omitted details in Section 4.4 . . . 138
C.4.1 Regret upper bounds in Expectation . . . 138
C.4.2 High-probability Upper Bounds . . . 144
C.4.3 High-probability bound on the network regret . . . 145
C.4.4 High-probability bound on the individual regret . . . 146
C.5 Omitted details in Section 4.5 . . . 147
C.6 Omitted details in Section 4.6 . . . 151
C.6.1 Proof of Corollary 4.2 . . . 152
C.6.2 Special cases . . . 155
C.7 Omitted details in Section 4.7 . . . 156

D Proof Details for Chapter 5 **159**
D.1 Preliminary Results . . . 159
D.1.1 General properties of FTRL . . . 159
D.1.2 Basic analysis facts . . . 159
D.1.3 Facts on the delay . . . 159
D.2 Omitted Details in Section 5.3 . . . 161
D.2.1 Non-Adaptive Algorithm with Known Total Delay . . . 161
D.2.2 Properties induced by the gossiping mechanism . . . 166
D.2.3 Adaptive Algorithm with Unknown Total Delay . . . 171
D.2.4 Lower Bound for the general convex case . . . 181
D.3 Omitted Details in Section 5.4 . . . 184
D.3.1 Lower bound for the strongly convex case . . . 191

E Proof Details for Chapter 6 **195**
E.1 Auxiliary results . . . 195
E.2 Omitted details in Section 6.4 . . . 198
E.3 Estimation of unknown link probability . . . 204

Chapter 1

Introduction

1.1 Online Learning

Online learning (Littlestone, 1990, Vovk, 1990) treats optimization as an inherently sequential process in which data or loss functions arrive over time. It has become a key tool across machine learning at scale, powering systems for contextual recommendation (Li et al., 2010), and market decision problems such as online portfolio selection and dynamic pricing (Cover, 1991, Kleinberg and Leighton, 2003, Besbes and Zeevi, 2009). Online and bandit formulations are also used in adaptive medical trials and A/B testing, where allocation policies trade off exploration and exploitation (Thompson, 1933, Scott, 2015) as well as in web advertising and sponsored search (Chapelle and Li, 2011). In forecasting applications, online aggregation and expert-advice methods have been successful for weather prediction (Flaspohler et al., 2021) and electricity-consumption/load forecasting at utility scale (Gaillard et al., 2016). In online learning setting, a learner repeatedly makes decisions with only past information while an environment (often modeled as an adversary) selects losses; the central objective is to design algorithms with provably low regret.

We begin with the general and powerful framework of Online Convex Optimization (OCO)(Zinkevich, 2003). At each round $t \in [T]$, the learner decides on a prediction $x_t \in \mathcal{X}$, where $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed, convex set. An adversary then reveals a convex loss function $\ell_t : \mathcal{X} \to \mathbb{R}$, and the learner incurs loss $\ell_t(x_t)$. The learner's performance is measured by the regret against the best fixed comparator in hindsight:

$$\mathrm{Reg}_T = \max_{u \in \mathcal{X}} \sum_{t=1}^{T} \big(\ell_t(x_t) - \ell_t(u)\big) \, .$$

The learner's goal is to guarantee that Reg_T is sub-linear in T for any sequence of losses. Let $g_t = \nabla \ell_t(x_t)$ be gradient of the current loss at the current iterate. By convexity,

$$\ell_t\left(x_t\right) - \ell_t(u) \leq \langle g_t, x_t - u \rangle \qquad \text{for all } u \in \mathcal{X},$$

and hence, writing $x^\star \in \arg\min_{u \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(u)$,

$$\mathrm{Reg}_T \leq \sum_{t=1}^{T} \langle g_t, x_t - x^\star \rangle \, .$$

This linearization reduces the analysis to online linear optimization in the observed gradients.

A generic and intuitive approach in OCO is Follow-the-Regularized-Leader (FTRL). At each time step, FTRL minimizes the cumulative linearized losses plus a regularization term. Let $\psi : \mathcal{X} \to \mathbb{R}$ be 1-strongly convex with respect to a norm $\|\cdot\|$; then, FTRL with a learning rate $\eta > 0$ updates

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \langle \sum_{s=1}^{t} g_s, x \rangle + \frac{1}{\eta}\psi(x),$$

A standard analysis (see Corollary 7.7 in Orabona (2025)) yields the regret bound

$$\mathrm{Reg}_T \leq \frac{\psi\left(x^\star\right) - \psi\left(x_1\right)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T} \|g_t\|_*^2,$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. Optimizing the bound over η and invoking the common assumptions $\|g_t\|_* \leq L$ (Lipschitz losses) and $\|x - y\| \leq D$ for all $x, y \in \mathcal{X}$ (bounded diameter) gives $\mathrm{Reg}_T = \mathcal{O}\left(DL\sqrt{T}\right)$. Through appropriate choices of ψ, FTRL subsumes several classic algorithms-including Hedge/multiplicative weights and is closely related to Online Mirror Descent (see Chapter 6 in Orabona (2025))-while providing a unified route to sublinear regret guarantees. For comprehensive treatments of OCO, we refer the reader to the monographs of Shalev-Shwartz et al. (2012), Hazan (2016), and Orabona (2025).

Many real-world systems do not reveal full-information feedback. Instead, the learner only sees the loss of the chosen decision. Regarding the partial information problem, the adversarial multi-armed bandit (MAB) (Auer et al., 2002b) is a canonical framework. The learner's decision set consists of K arms. In each round $t \in [T]$, the learner chooses arm $I_t \in [K]$ and suffers loss $\ell_t\left(I_t\right) \in [0, 1]$. Crucially, only the loss $\ell_t\left(I_t\right)$ of the chosen arm is observed, and losses of unplayed arms remain hidden. The regret compares the learner to the best single arm in hindsight:

$$\mathrm{Reg}_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t\left(I_t\right)\right] - \min_{i \in [K]} \sum_{t=1}^{T} \ell_t(i).$$

Although this problem is directly subsumed by the OCO framework; it can still be embedded naturally into it. Firstly, we replace the (unavailable) full loss vector $\ell_t \in [0, 1]^K$ with a carefully constructed importance-weighted estimator based on the observed scalar feedback:

$$\widehat{\ell}_t(i) = \frac{\ell_t(i)}{x_t(i)} \mathbb{I}\left\{I_t = i\right\}, \quad \text{for all } i \in [K],$$

where x_t is the sampling distribution at time t and $\mathbb{I}\{\cdot\}$ is the indicator function. As for the decision set, it can be seen as the probability simplex $\Delta_K = \left\{x \in \mathbb{R}_{\geq 0}^K : \sum_{i=1}^K x(i) = 1\right\}$, such that upon choosing $x_t \in \Delta_K$ at round t, the learner plays $I_t \sim x_t$. This perspective suggests the following approach: run FTRL over Δ_K with the negative entropy regularizer $\psi(x) = \sum_{i=1}^K x(i) \log x(i)$:

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \langle \sum_{s=1}^{t} \widehat{\ell}_s, x \rangle + \frac{1}{\eta}\psi(x), \quad \eta > 0$$

This formulation recovers the EXP3 algorithm (Auer et al., 2002b), where at each round t, the distribution x_t is given by

$$x_t(i) \propto \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right).$$

Via a more refined analysis of FTRL using the local norm (see Lemma 7.16 in (Orabona, 2025)), we obtain

$$\sum_{t=1}^{T} \langle x_t - x^\star, \widehat{\ell}_t \rangle \leq \frac{\psi(x^\star)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{K} x_t(i) \widehat{\ell}_t(i)^2,$$

where the expectation of the left-hand side is equal to Reg_T. By upper bounding the expectation of the right-hand side, we obtain

$$\mathrm{Reg}_T \leq \frac{\log K}{\eta} + \frac{\eta K T}{2}.$$

Optimizing over η yields the regret bound $\mathcal{O}(\sqrt{KT \log K})$. Moreover, FTRL with the negative 1/2-Tsallis entropy $\psi(x) = 2\left(1 - \sum_{i=1}^{K} \sqrt{x(i)}\right)$ obtains the sharper bound $\mathcal{O}(\sqrt{KT})$, which is minimax-optimal up to constants, matching a lower bound by Auer et al. (2002b).

1.2 Delayed Online Learning

In many realistic systems, feedback about a decision does not arrive immediately. Examples include online advertising, where outcomes such as click-through rates or conversions are often delayed due to subsequent user interactions; medical trials, where the effects of a drug or treatment may only be observed after an initially unknown period; and financial investments, where returns or risk signals may materialize over days or weeks. This has prompted the investigation of online learning algorithms that are robust to such delays (Joulani et al., 2013, Cesa-Bianchi et al., 2016a, 2019, Zimmert and Seldin, 2020a, Masoudian et al., 2022a, 2023). To capture this phenomenon, we augment the standard online protocol with an arbitrary, possibly adversarial, delay sequence $\{d_t\}_{t=1}^{T}$ indicating when the feedback for round t becomes available, see Online Protocol 1.1.

Online Protocol 1.1: OCO with delayed feedback

for $t = 1, \ldots, T$ **do**
The learner selects an action $x_t \in \mathcal{X}$
The learner incurs loss $\ell_t(x_t)$ immediately, but only observes $\ell_t(x_t)$ at time $t + d_t$.

Let $d_{\mathrm{tot}} = \sum_{t=1}^{T} d_t$ denote the total amount of delay. We now adapt FTRL to the delayed setting. The learner maintains the set of indices whose feedback has been received by the start of round t, $o_t = \{ s \in [t-1] : s + d_s < t \}$ and computes the updates using only the arrived information:

$$x_t = \arg\min_{x \in \mathcal{X}} \langle \sum_{s \in o_t} g_s, x \rangle + \frac{1}{\eta} \psi(x).$$

For convex losses, the delayed variant of FTRL with the optimized η satisfy $\mathrm{Reg}_T = \mathcal{O}(DL\sqrt{T + d_{\mathrm{tot}}})$ (Quanrud and Khashabi, 2015, Gyorgy and Joulani, 2021, Flaspohler et al., 2021), matching a known lower bound by Joulani et al. (2013).

In the case of multi-armed bandit with delayed feedback, FTRL with a hybrid regularizer can achieve regret $\mathcal{O}(\sqrt{KT} + \sqrt{d_{\mathrm{tot}} \ln K})$ (Zimmert and Seldin, 2020a), which matches a lower bound by

Cesa-Bianchi et al. (2016a). These results illustrate that the presence of delays negatively affects the regret in an unavoidable manner.

1.3 Distributed Online Learning

Distributed online convex optimization (DOCO) (Yan et al., 2013, Hosseini et al., 2013) arises in emerging applications such as smart grids (Wang et al., 2014), sensor networks (Li et al., 2002), and robotics (Shorinwa et al., 2024), where both data and decision-making are distributed across physically separated subsystems represented by agents. These agents are interconnected through a communication network and collectively aim to minimize global losses. DOCO integrates the strengths of OCO and distributed optimization (Nedic and Ozdaglar, 2009b, Duchi et al., 2011), yielding robust and scalable optimization frameworks for complex multi-agent systems (see Online Protocol 1.2). In its simplest form, the communication network is modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a finite set of indices representing agents, and $\{u, v\} \in \mathcal{E}$ indicates that agents u and v can exchange information. At each time step $t \in [T]$, each agent $u \in \mathcal{V} = \{1, \dots, N\}$ observes a local loss $\ell_t(u, x_t(u))$ and incurs a global loss $\sum_{v\in\mathcal{V}} \ell_t(v, x_t(u))$. Each agent then selects a decision $x_t(u) \in \mathcal{X}$ while exchanging information with its neighbors over the communication graph $\mathcal{G}$. The regret of agent $u \in \mathcal{V}$ over T rounds is defined as

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

Online Protocol 1.2: Distributed OCO for each agent $u \in \mathcal{V}$

for $t = 1, \dots, T$ **do**
The agent selects an action $x_t(u) \in \mathcal{X}$
The agent incurs a global loss $\sum_{v\in\mathcal{V}} \ell_t(v, x_t(u))$ and observes a local loss $\ell_t(u, x_t(u))$
The agent u receives neighbors' messages over $\mathcal{G}$ (e.g., decisions $x_t(v)$ or local summaries) for all $v \in \mathcal{N}_u$, $\mathcal{N}_u \triangleq \{v \in \mathcal{V} : \{u, v\} \in \mathcal{E}\}$ be the neighbor set of agent u.

This quantity captures the performance gap between the sequence of decisions made by agent u and the best fixed decision in hindsight. Since distributed systems must balance local autonomy with global objectives, the network-level performance is naturally evaluated through the worst-case regret across all agents:

$$\mathrm{Reg}_T = \max_{u\in\mathcal{V}} \mathrm{Reg}_T(u).$$

Minimizing Reg_T ensures that no single agent suffers disproportionately large losses compared to others, thereby promoting fairness and robustness in distributed decision-making. This system-wide perspective makes DOCO particularly suitable for large-scale, heterogeneous networks where individual agents may encounter different environments or loss functions, yet the collective outcome must remain reliable.

The primary challenge in DOCO is reconciling the robustness inherent in online learning with the scalability requirements of distributed multi-agent systems. Existing online learning algorithms must be carefully reformulated to accommodate decentralized settings, giving rise to distributed online learning algorithms. The attainable global performance—typically quantified by regret—is

fundamentally influenced by factors such as the choice of learning rate, the underlying network topology, and the communication complexity. These interdependencies pose significant challenges for rigorous performance evaluation and theoretical analysis.

Let $W \in \mathbb{R}^{N\times N}$ be a symmetric, doubly-stochastic matrix consistent with the communication graph $\mathcal{G}$ (i.e., $W(u,v) > 0$ only if $u = v$ or $\{u,v\} \in \mathcal{E}$). Define the spectral gap as $1-\sigma_2(W) \in (0,1]$, where $\sigma_2(W)$ denotes the second-largest eigenvalue. A larger spectral gap indicates faster mixing of information across the network. A distributed variant of FTRL algorithm (Hosseini et al., 2013) operates as follows. Each agent $u \in \mathcal{V}$ maintains a local accumulator $z_t(u)$ of (sub)gradients. At each round, the agent first performs a mixing (or gossip) step with its neighbors, then adds its local gradient:

$$z_{t+1}(u) = \sum_{v\in\mathcal{N}_u} W(u,v) z_t(v) + \nabla \ell_t\left(u, x_t(u)\right),$$
$$x_{t+1}(u) = \arg\min_{x\in\mathcal{X}} \langle z_{t+1}(u), x\rangle + \tfrac{1}{\eta}\psi(x).$$

In contrast to the omniscient centralized FTRL update, where the cumulative gradient

$$\frac{1}{N}\sum_{s=1}^{t}\sum_{v\in\mathcal{N}_v} \nabla \ell_s(v, x_s(v))$$

is directly accessible, the decentralized version replaces it with the locally maintained accumulator $z_{t+1}(u)$. This variable is computed by averaging (via the gossip step) the accumulators of neighboring agents and then incorporating the local gradient $\nabla\ell_t(u, x_t(u))$. For convex losses, this decentralized variant of FTRL achieves a regret bound of $\mathrm{Reg}_T = \mathcal{O}\left(N^{5/4}\left(1-\sigma_2(W)\right)^{-1/2}\sqrt{T}\right)$, for each agent $u \in \mathcal{V}$. Moreover, Wan et al. (2024b) obtained a improved regret bounds $\widetilde{\mathcal{O}}\left(N(1-\sigma_2(W))^{-1/4}\sqrt{T}\right)$ and provide a nearly matching lower bound $\Omega\left(N(1-\sigma_2(W))^{-1/4}\sqrt{T}\right)$.

1.4 Outline of the Thesis

This thesis is organized as follows. Chapter 2 and Chapter 3 study online learning with delayed feedback, while Chapter 4 focuses on distributed online learning. Chapter 5 presents a unified algorithmic framework for handling delayed feedback in distributed online convex optimization. Finally, Chapter 6* moves from the adversarial to the stochastic regime and investigates distributed stochastic bandits.

Delayed Online Convex Optimization with Curvature In Chapter 2, we study the OCO with curved losses and delayed feedback. The main objective is to exploit the loss curvature to improve regret guarantees even with delayed feedback. For strongly convex losses, we propose a variant of FTRL that obtains regret of order $\min\{\sigma_{\max}\ln T, \sqrt{d_{\mathrm{tot}}}\} + \ln T$, where $\sigma_{\max}$ is the maximum number of missing observations. We then consider exp-concave losses and extend the Online Newton Step algorithm to handle delays with an adaptive learning rate tuning, achieving regret $\min\{d_{\max} n \ln T, \sqrt{d_{\mathrm{tot}}}\} + n\ln T$ where n is the dimension. We further consider the problem

*Note that in this chapter, we adopt slightly different notation to align with the conventions commonly used in the stochastic bandit literature.

of unconstrained online linear regression and achieve a similar guarantee by designing a variant of the Vovk-Azoury-Warmuth forecaster with a clipping trick.

Delayed Bandits with Intermediate Observation In Chapter 3, we study a multi-armed bandit problem with delayed bandit feedback and intermediate observations. In this model, an intermediate observation is any element from a finite state space $\mathcal{S}$ and is observed immediately after taking an action, whereas the loss is observed after an adversarially chosen delay. The main objective is to understand when intermediate observations help reduce the effect of the total delay on the regret. We show that the regime of the mapping of states to losses determines the complexity of the problem, irrespective of whether the mapping of actions to states is stochastic or adversarial. If the state-loss mapping is adversarial, then we prove that intermediate observations cannot help. Otherwise, if the same mapping is stochastic and uniform delays d, we design an algorithm whose regret grows at rate $\sqrt{\big(K + \min\{|\mathcal{S}|, d\}\big)T}$ without logarithmic factors, implying that intermediate observations can reduce the negative effect of the total delay if their number $|\mathcal{S}|$ is sufficiently small. We also provide refined high-probability regret bounds for non-uniform delays.

Distributed Online Convex Optimization with Stochastic Agent Availability In Chapter 4, we investigate a variant of DOCO where agents $v \in \mathcal{V}$ are active with a known probability p_v at each time step, and communication between neighboring agents can only take place if they are both active. We propose a distributed variant of FTRL algorithm and analyze its individual regret, defined for each agent as the cumulative regret with respect to the global loss function, restricted to the time steps when the agent is active. Our analysis shows that, for any connected communication graph $\mathcal{G}$ over N agents, the expected individual regret of our FTRL variant after T steps is at most of order $(\kappa/p^{3/4})N^{1/4}\sqrt{T}$ for any agent, when $p_v = p$ for all agents and where κ is the condition number of the Laplacian of $\mathcal{G}$. Moreover, we show a regret lower bound that implies that our bounds are not significantly improvable.

Distributed Online Convex Optimization with Feedback Delays In Chapter 5, we study DOCO under unknown, time- and agent-varying feedback delays. The main objective is how to design decentralized online learning algorithms that adapt to unknown, time- and agent-varying delays while maintaining near-optimal regret guarantees. We propose a novel algorithm based on FTRL that achieves an improved regret bound of $\widetilde{\mathcal{O}}\Big(\sqrt{N^3 d_{\text{tot}}} + \sqrt{\frac{N^3 T}{\sqrt{1-\sigma_2}}}\Big)$, 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 show that the dependence on T, d_{tot} and $1 - \sigma_2$ is tight by providing a matching lower bound. Our approach crucially incorporates an adaptive learning rate mechanism via a distributed 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 sharper regret bounds.

Distributed Stochastic Bandits over Random Communication Networks In Chapter 6, we study the distributed multi-agent multi-armed bandit problem with heterogeneous rewards over random communication graphs. Uniquely, at each time step t agents communicate over a time-varying random graph $\mathcal{G}_t$ generated by applying the Erdős–Rényi model to a fixed connected

base graph $\mathcal{G}$, where each potential edge in $\mathcal{G}$ is randomly and independently present with probability p. Additionally, each agent's arm rewards follow time-invariant distributions, and the reward distribution for the same arm may differ across agents. The goal is to minimize the cumulative expected regret relative to the global mean reward of each arm, defined as the average of that arm's mean rewards across all agents. To this end, we propose a fully distributed algorithm that integrates the arm elimination strategy with the random gossip algorithm. We analyze the regret of our algorithm and also provide a lower bound. We theoretically show that the regret upper bound is of order $\log T$ and is highly interpretable, where T is the time horizon. It includes the optimal centralized regret $\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log T}{\Delta_k}\right)$ and an additional term $\mathcal{O}\left(\frac{N^2 \log T}{p\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))} + \frac{KN^2 \log T}{p}\right)$ where N and K denotes the total number of agents and arms, respectively. This term reflects the impact of $\mathcal{G}$'s algebraic connectivity $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ and the link probability p, and thus highlights a fundamental trade-off between communication efficiency and regret.

Chapter 2

Exploiting Curvature in Online Convex Optimization with Delayed Feedback

2.1 Introduction

As mentioned in the introductory chapter, feedback in many real-world applications is not immediately available after the learner's decision but is instead subject to a delay. Another crucial element in OCO is given by properties of the loss functions such as the curvature. It is indeed often the case that losses have additional curvature properties such as strong convexity or exp-concavity. For example, exp-concave losses are prevalent in portfolio management (Cover, 1991), in which the learner (investor) needs to distribute her wealth over a set of financial instruments in order to maximize her return. When the loss functions have a certain curvature, previous works (Hazan et al., 2007) have shown that a significantly better regret guarantee can be achieved (i.e., the so-called fast rates). However, this type of assumption received little attention when assuming that the feedback suffers some delay. Therefore, we are interested in investigating the following question:

Can we design algorithms that exploit the loss curvature to obtain improved guarantees even with delayed feedback?

There is a line of works studying OCO with delayed feedback. For general convex functions, Quanrud and Khashabi (2015) provided an algorithm called Delayed Online Gradient Descent (DOGD) and achieves a regret of $\mathcal{O}(\sqrt{T + d_{\text{tot}}})$ where T is the time horizon and d_{tot} is the total delay. Subsequently, Wan et al. (2022a), Wu et al. (2024) focused on strongly convex losses, introducing DOGD-SC and SDMD-RSC, which achieve a regret bound of $\mathcal{O}((d_{\max} + 1) \ln T)$, where $d_{\max}$ represents the maximum delay for any single round of feedback. However, the $\mathcal{O}((d_{\max} + 1) \ln T)$ regret bound can sometimes be much worse than $\mathcal{O}(\sqrt{T + d_{\text{tot}}})$. This occurs in scenarios when even a single round of feedback is delayed by $\Theta(T)$ rounds (e.g., missing feedback), undermining the benefits of having both regret guarantees under stronger curvature assumptions. Furthermore, to the best of our knowledge, no prior work has investigated whether improved regret guarantees are achievable for exp-concave losses under delayed feedback, leaving an important gap in the literature.

Contribution. To address these gaps, we propose a suite of algorithms and offer a comprehensive analysis for OCO with delayed feedback under both strongly convex and exp-concave losses, and we include a special case of (unconstrained) online linear regression with delays.

Loss type	Regret bound		
	Quanrud and Khashabi (2015)	Wan et al. (2022a); Wu et al. (2024)	Our work
Strongly convex	$\sqrt{d_{\text{tot}} + T}$	$(d_{\max} + 1) \ln T$	$\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\} + \ln T$
Exp-concave	$\sqrt{d_{\text{tot}} + T}$	N/A	$\min\{d_{\max} n \ln T, \sqrt{d_{\text{tot}}}\} + n \ln T$
Online linear regression	N/A	N/A	$\min\{d_{\max} n \ln T, \sqrt{d_{\text{tot}}}\} + n \ln T$

Table 2.1: Main results and comparisons with prior work. Here T is the number of rounds, n is the dimension of the feasible domain, $d_{\max}$ is the maximum delay, $\sigma_{\max} \leq d_{\max}$ is the maximum number of missing observations, and d_{tot} is the total delay. In Table 2.1, we omit the dependency on the curvature parameters, Lipschitz parameters, the norm of the comparator and domain diameter for conciseness. The detailed dependencies are explicitly shown in the respective theorem statements.

The main contributions of this work can be summarized as follows (see also Table 2.1):

- We first consider the class of strongly convex losses in Section 2.3. Specifically, we propose an algorithm based on the follow-the-regularized-leader framework and obtain a $\mathcal{O}\left(\min\left\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\right\} + \ln T\right)$ regret, where $\sigma_{\max}$ is the maximum number of missing observations over rounds. Compared with the results obtained by Wan et al. (2022a) and Wu et al. (2024), our results have several advantages. First, since $\sigma_{\max}$ is always no larger than $d_{\max}$ and can be significantly smaller than it, our $\sigma_{\max} \ln T$ bound improves upon the $d_{\max} \ln T$ bound in Wan et al. (2022a) and Wu et al. (2024). Second, we prove that our algorithm *simultaneously* achieves a $\mathcal{O}\left(\sqrt{d_{\text{tot}}} + \ln T\right)$ regret bound, making our algorithm no worse than the bound achieved by DOGD (Quanrud and Khashabi, 2015) either. Third, compared with the regret bounds obtained in Wan et al. (2022a) and Wu et al. (2024), our regret guarantee *does not depend on* the diameter of the action domain and recovers the one proven in Hazan et al. (2007) when there is no delay. Additionally, we provide a novel and improved analysis of the OMD-based algorithm originally proposed by Wu et al. (2024) in Appendix A.5, obtaining a regret bound that is again independent of the diameter of the action domain.
- In Section 2.4, we consider exp-concave losses, a broader function class compared to the strongly convex one. Specifically, we propose an algorithm based on the Online Newton Step (ONS) method that achieves a $\mathcal{O}\left(\min\left\{d_{\max} n \ln T, \sqrt{d_{\text{tot}}}\right\} + n \ln T\right)$ regret bound. To the best of our knowledge, this is the first algorithm to achieve logarithmic regret for exp-concave losses under delayed feedback, answering an open question proposed in Wan et al. (2022a). While both the bounds $d_{\max} n \ln T$ and $\sqrt{d_{\text{tot}}}$ can be achieved using a simple learning rate within the ONS framework, it is essential to use a delay-adaptive learning rate tuning scheme to achieve the best of these two guarantees within our analysis.
- In Section 2.5, we investigate online linear regression (OLR) problem, where the feasible domain is *unconstrained*, i.e., it corresponds to the entire n-dimensional Euclidean space $\mathbb{R}^n$. Leveraging the specific structure in OLR, we develop an algorithm based on the Vovk-Azoury-Warmuth forecaster, achieving a regret bound of $\mathcal{O}\left(\|u\|_2^2(\min\left\{d_{\max} n \ln T, \sqrt{d_{\text{tot}}}\right\} + n \ln T)\right)$ without requiring any prior knowledge of neither the comparator $u \in \mathbb{R}^n$ nor the data. This result is achieved by incorporating a carefully designed clipping technique and, once again, employing an adaptive tuning of the learning rate.

- Finally, in Section 2.6, we implement all our proposed algorithms and conduct experiments to validate our theoretical results across multiple delayed settings and loss functions with different curvature properties. We also compare our methods with existing approaches to demonstrate their effectiveness.

2.1.1 Related works

Online learning with curved losses. While Abernethy et al. (2008) have shown that $\Theta(\sqrt{T})$ is the minimax regret for OCO, if the loss functions further enjoy curvature, the minimax regret can be improved. Hazan et al. (2007) show that OGD with a specific choice of learning rate achieves $\mathcal{O}(\frac{L^2}{\lambda}\ln T)$ regret for strongly convex losses where L is the maximum ℓ_2 norm of any loss gradient and λ is the strong convexity parameter.* This upper bound is also minimax optimal as proven in Abernethy et al. (2008). For exp-concave losses, Hazan et al. (2007) proposed Online Newton Step (ONS) achieving $\mathcal{O}((\frac{1}{\alpha}+LD)\ln T)$ regret where α is the exp-concavity parameter and D is the diameter of the feasible domain. Hazan et al. (2007) also proposed Exponential Weight Online Optimization (EWOO), achieving diameter and gradient scale independent guarantees. However, the algorithm is less practical due to its sampling complexity. For OLR, Vovk (2001) and Azoury and Warmuth (2001) independently introduced the Vovk-Azoury-Warmuth (VAW) forecaster achieving $\mathcal{O}(\ln T)$ regret without requiring prior knowledge of the data and the comparator. For a more detailed survey on OCO, we recommend the reader to Hazan (2016) and Orabona (2025).

Online learning with delayed feedback. Weinberger and Ordentlich (2002) initiated the study of online learning with delayed feedback, proposing an algorithm achieving $d \cdot \mathrm{Reg}_T(T/d)$ where d is the *fixed and known per-round delay* and $\mathrm{Reg}_T(T)$ is the regret upper bound for some base algorithm that assumes no delay in the feedback. Specifically, their meta-algorithm runs $d+1$ independent copies of the base algorithm on disjoint time lines in a round-robin fashion. However, this meta-algorithm is computationally expensive and does not show good empirical performances. Subsequently, Langford et al. (2009) proposed a practical algorithm by simply performing the gradient descent step using the observed gradients at each round, and achieved $\mathcal{O}(\sqrt{dT})$ and $\mathcal{O}(d\ln T)$ regret bounds for convex and strongly convex functions, respectively.

When delay is not uniform, Joulani et al. (2013) proposed BOLD (Black-box Online Learning with Delays) extending the method of Weinberger and Ordentlich (2002) and achieve $d_{\max} \cdot R(T/d_{\max})$ regret, but the algorithm still maintains multiple instances of base algorithms, which could be prohibitive in terms of computational costs. For convex functions, Quanrud and Khashabi (2015) achieved $\mathcal{O}(\sqrt{d_{\mathrm{tot}}})$ where d_{tot} is the total delay accumulated over T rounds. Wan et al. (2022b, 2023) proposed a first Frank-Wolfe-type online algorithm to handle delayed feedback and obtain a regret bound of $\mathcal{O}(T^{3/4}+d_{\mathrm{tot}}T^{-3/4})$ for general convex loss and $\mathcal{O}(T^{2/3}+d_{\max}\ln T)$ under strong convexity. There is also an interesting line of works whose focus is to obtain adaptive regret guarantees with delayed feedback (McMahan and Streeter, 2014a, Joulani et al., 2016b, Flaspohler et al., 2021) or variants of delayed feedback (Gatmiry and Schneider, 2024, Bar-On and Mansour, 2025, Ryabchenko et al., 2025).

Two most related works to ours are Wan et al. (2022a) and Wu et al. (2024), which consider strongly convex losses together with delays. Specifically, Wan et al. (2022a) first proposed DOGD-SC

*The definitions of these parameters are deferred to Section 2.2.

for strongly convex losses, and establish a regret bound of $\mathcal{O}(\frac{LD+L^2}{\lambda} d_{\max} \ln T)$. Subsequently, Wu et al. (2024) proposed SDMD-RSC and obtained a $\mathcal{O}(\frac{d_{\max} L^2}{\lambda^2} + \frac{L^2+D}{\lambda} d_{\max} \ln T)$ regret bound.[†]

Beyond full-gradient feedback, there exists a growing interest in developing algorithms with delayed bandit feedback for a range of problems, including multi-armed bandits (Cesa-Bianchi et al., 2016b, Cella and Cesa-Bianchi, 2020, Zimmert and Seldin, 2020b, Masoudian et al., 2022b, van der Hoeven and Cesa-Bianchi, 2022a, Esposito et al., 2023, van der Hoeven et al., 2023a, Masoudian et al., 2024b, Schlisselberg et al., 2025, Zhang et al., 2025), Markov decision processes (Lancewicki et al., 2022, Jin et al., 2022, van der Hoeven et al., 2023a), and online convex optimization (Héliou et al., 2020, Bistritz et al., 2022b, Wan et al., 2024d).

2.2 Problem setting

Let $T \in \mathbb{N}$ be the time horizon and $n \in \mathbb{N}$ be the dimension. Denote by $\mathcal{X} \subset \mathbb{R}^n$ the domain, which we assume to be closed and non-empty. In each round $t \in [T]$, the learner selects a point $x_t \in \mathcal{X}$ as its decision and incurs a loss $\ell_t(x_t)$ given by some unknown function $\ell_t \colon \mathcal{X} \to \mathbb{R}$ that we assume to be convex and differentiable. Normally, in the standard OCO setting, the learner would then immediately observe the gradient $g_t = \nabla \ell_t(x_t)$. On the other hand, here we consider the delayed feedback scenario in which such a gradient g_t is only observed at round $t + d_t$ with some *unknown arbitrary delay* $d_t \geq 0$. We assume $t + d_t \leq T$ for all $t \in [T]$ without loss of generality Joulani et al. (2013, 2016b) because the feedback of any round t with $t + d_t \geq T$ cannot be used the learner. The performance of the learner is then measured via the regret, which is defined as follows:

$$\mathrm{Reg}_T = \max_{u \in \mathcal{X}} \mathrm{Reg}_T(u) = \max_{u \in \mathcal{X}} \sum_{t=1}^{T} \big(\ell_t(x_t) - \ell_t(u)\big) \,.$$

For convenience, we define $o_t = \{\tau \in \mathbb{N} : \tau + d_\tau < t\} \subseteq [t-1]$ to be the set of rounds whose gradients are observed before round t, and let $m_t = [t-1] \setminus o_t$ be the set of rounds whose observation is yet to be received at the beginning of round t. Define $\sigma_{\max} = \max_{t \in [T]} |m_t|$ to be the maximum number of missing observations over T rounds, $d_{\max} = \max_{t \in [T]} d_t$ to be the maximum delay, and $d_{\mathrm{tot}} = \sum_t d_t$ to be the total delay. Also define $d_{\max}^{\leq t} = \max_{\tau \leq t} \min\{d_\tau, t - \tau\}$ as the maximum delay that has been perceived up to round t.

Before presenting our main results, we must first introduce some definitions about the curvature of the loss functions.

Definition 2.1. *A function $f \colon \mathcal{X} \to \mathbb{R}$ is* λ-strongly convex *with respect to $\|\cdot\|$ for $\lambda > 0$ if, for all $x, y \in \mathcal{X}$, $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\lambda}{2} \|y - x\|_2^2$.*

Definition 2.2. *A function $f \colon \mathcal{X} \to \mathbb{R}$ is* α-exp-concave *for $\alpha > 0$ if $x \mapsto \exp(-\alpha f(x))$ is concave over $\mathcal{X}$.*

We finally introduce some standard boundedness assumptions relating to the gradients and the domain.

Assumption 2.1. *For every $t \in [T]$, the gradient of ℓ_t has norm bounded by $L \geq 0$, i.e., $\max_{x \in \mathcal{X}} \|\nabla \ell_t(x)\|_2 \leq L$.*

[†]Wu et al. (2024) also considers the class of relative strongly convex loss functions.

Algorithm 2.1: Delayed FTRL for strongly convex functions

input: strong convexity parameter $\lambda > 0$
initialize: $x_1 \in \mathcal{X}$
for $t = 1, 2, \ldots$ **do**
    Play x_t; receive $g_\tau = \nabla \ell_\tau(x_\tau)$ for all $\tau \in o_{t+1} \setminus o_t$
    Update

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \sum_{\tau \in o_{t+1}} \langle g_\tau, x \rangle + \frac{\lambda}{2} \sum_{s \le t} \|x - x_s\|_2^2 \tag{2.1}$$

Assumption 2.2. *The diameter of $\mathcal{X}$ is bounded by $D \ge 0$, i.e., $\max_{x,y \in \mathcal{X}} \|x - y\|_2 \le D$. We also assume $\mathbf{0} \in \mathcal{X}$.*

Other notations. For a positive semidefinite matrix $A \in \mathbb{R}^{n \times n}$ and $x \in \mathbb{R}^d$, we denote $\|x\|_A = \sqrt{x^\top A x}$ to be the Mahalanobis norm induced by A and, if A is positive definite, let $\|x\|_{A^{-1}} = \sqrt{x^\top A^{-1} x}$ be the dual norm. We denote $\mathbf{1}$ as the all-one vector in an appropriate dimension.

2.3 Delayed OCO with Strongly Convex Losses

In this section, we consider the problem of delayed OCO with strongly convex losses and propose Algorithm 2.1, which is built upon the follow-the-regularized-leader (FTRL) algorithm. Specifically, after receiving the gradients g_τ for all $\tau \in o_{t+1} \setminus o_t$ at the end of round t, we compute the updated decision x_{t+1} as shown in Equation (2.1), which is the minimizer of the cumulative linearized loss with respect to all the currently observed gradients, plus a squared ℓ_2-regularization term with respect to *all the past decisions*. The following theorem shows that Algorithm 2.1 achieves $\mathcal{O}\big(\frac{L^2}{\lambda}\left(\ln T + \min\left\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\right\}\right)\big)$ regret bound without any diameter assumption on the domain.

Theorem 2.1. *Assume that $\ell_1, \ldots, \ell_T$ are λ-strongly convex with respect to the Euclidean norm $\|\cdot\|_2$. Then, under Assumption 2.1, Algorithm 2.1 guarantees that*

$$\mathrm{Reg}_T = \mathcal{O}\left(\frac{L^2}{\lambda}\left(\ln T + \min\left\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\right\}\right)\right).$$

Theorem 2.1 highlights two advantages over previous works. From the perspective of the delay-related term, while both DOGD-SC (Wan et al., 2022a) and SDMD-RSC (Wu et al., 2024) achieve a $\mathcal{O}\left(d_{\max} \ln T\right)$ regret bound, the terms $\sigma_{\max}$ and $\sqrt{d_{\text{tot}}}$ in our regret bound can be substantially smaller than $d_{\max}$, with $\sigma_{\max} \le d_{\max}$ always being true (Masoudian et al., 2022b).[‡] Second, while both DOGD-SC and SDMD-RSC have polynomial dependence on the diameter D of the action set $\mathcal{X}$, we remark that our bound *does not* depend on D and recovers the optimal $\mathcal{O}\big(\frac{L^2}{\lambda} \ln T\big)$ regret in the no-delay setting.

[‡]In fact, we also show in Lemma A.7 that $\sigma_{\max} \lesssim \sqrt{d_{\text{tot}}}$, and in Lemma A.9 that there are delay sequences such that $\sigma_{\max} \ll \sqrt{d_{\text{tot}}}$ and $\sigma_{\max} \approx \sqrt{d_{\text{tot}}}$, respectively.

2.3.1 Regret Analysis

Here we provide a proof sketch of Theorem 2.1, whereas the full proof is deferred to Appendix A.2. Specifically, using the strong convexity property, we first decompose the regret:

$$\mathrm{Reg}_T(u) \leq \sum_{t=1}^{T} \left(\langle g_t, x_t - u \rangle - \frac{\lambda}{2}\|x_t - u\|_2^2 \right) = \underbrace{\sum_{t=1}^{T} \langle g_t, x_t^\star - u \rangle}_{\mathrm{Reg}_T^\star(u)} + \underbrace{\sum_{t=1}^{T} \langle g_t, x_t - x_t^\star \rangle}_{\texttt{Drift}_T} - \frac{\lambda}{2} \sum_{t=1}^{T} \|x_t - u\|_2^2, \tag{2.2}$$

where $x_t^\star = \arg\min_{x \in \mathcal{X}} \sum_{\tau=1}^{t-1} (\langle g_\tau, x \rangle + \frac{\lambda}{2}\|x - x_\tau\|_2^2)$ for $t \geq 2$ and $x_1^\star = x_1$ are the decisions assuming that all gradients before round t are observed.

Next, we analyze the term $\mathrm{Reg}_T^\star(u)$ and $\texttt{Drift}_T$ separately. For the term $\texttt{Drift}_T$, applying the Cauchy-Schwarz inequality and using the fact that $\|g_t\|_2 \leq L$ for all $t \in [T]$ by Assumption 2.1, we can obtain that

$$\texttt{Drift}_T \leq L \sum_{t=1}^{T} \|x_t^\star - x_t\|_2 \,. \tag{2.3}$$

For the term $\mathrm{Reg}_T^\star(u)$, following a standard FTRL analysis and using the optimality of $x_t^\star$, we are able to obtain that

$$\mathrm{Reg}_T^\star(u) \leq \frac{\lambda}{2} \sum_{t=1}^{T} \|x_t - u\|_2^2 + \sum_{t=1}^{T} \langle g_t, x_t^\star - x_{t+1}^\star \rangle \,.$$

Since the first term can be canceled by the last negative term shown in Equation (2.2), we only need to control the second term $\langle g_t, x_t^\star - x_{t+1}^\star \rangle$, which is further bounded by $L\|x_t^\star - x_{t+1}^\star\|_2$ via Cauchy-Schwarz and the fact that $\|g_t\|_2 \leq L$. Then, using a stability lemma for FTRL (Lemma D.1), we can show that

$$\|x_t^\star - x_{t+1}^\star\|_2 \leq \frac{2L}{\lambda(2t-1)} + \|x_t^\star - x_t\|_2 \,.$$

Interestingly, this inequality relates the Euclidean distance between adjacent "cheating" iterates $(x_t^\star)_{t \geq 1}$ in the stability term of FTRL to the distance between x_t and $x_t^\star$, which is also present in the $\texttt{Drift}_T$ term and intuitively quantifies the influence of delays on the regret.

Combining the inequalities involving $\mathrm{Reg}_T^\star(u)$ and $\texttt{Drift}_T$, we can finally bound the regret from above as follows:

$$\mathrm{Reg}_T \leq \sum_{t=1}^{T} \frac{2L^2}{\lambda(2t-1)} + 2L \sum_{t=1}^{T} \|x_t^\star - x_t\|_2 \leq \frac{L^2}{\lambda} \ln(2T+1) + 2L \sum_{t=1}^{T} \|x_t^\star - x_t\|_2 \,.$$

It remains to show how to bound $\|x_t^\star - x_t\|_2$ by $\mathcal{O}\big(\frac{L}{\lambda} \min\{\sigma_{\max} \ln T, \sqrt{d_{\mathrm{tot}}}\}\big)$, which is the key novelty in our analysis compared to previous works. Recalling the definitions of x_t and $x_t^\star$, we can apply the

stability lemma of FTRL (Lemma D.1) again and show for all $t \geq 2$ that

$$\frac{\lambda(t-1)}{2}\|x_t^\star - x_t\|_2^2 \leq \frac{\left\|\sum_{\tau \in m_t} g_\tau\right\|_2^2}{2\lambda(t-1)}\,, \tag{2.4}$$

meaning that $\sum_{t=1}^T \|x_t^\star - x_t\|_2 \leq \sum_{t=2}^T \frac{\left\|\sum_{\tau \in m_t} g_\tau\right\|_2}{\lambda(t-1)} \leq \sum_{t=2}^T \frac{L|m_t|}{\lambda(t-1)}$, where we also use the fact that $x_1^\star = x_1$. Here, we highlight the importance of including all previous decisions x_τ for $\tau \leq t$, instead of $\tau \in o_{t+1}$ only, in the regularization term of the update rule of x_{t+1} shown in Equation (2.1). Doing so particularly ensures that the updates of x_t and $x_t^\star$ share the same regularization terms, which is crucial in leading to a diameter-free upper bound for $\|x_t^\star - x_t\|_2$ using the stability lemma.

Finally, we study the term $\sum_{t=2}^T \frac{|m_t|}{t-1}$. Directly bounding $|m_t|$ from above by $\sigma_{\max}$ leads to the first $\sigma_{\max} \ln T$ bound. To further obtain the $\sqrt{d_{\text{tot}}}$ bound, it is crucial to observe that $\sum_{\tau \leq t} |m_\tau| \leq (t-1)^2$ since $m_\tau \subseteq [\tau - 1]$. Therefore, by also using Orabona (2025, Lemma 4.13) we are able to prove that $\sum_{t=2}^T \frac{|m_t|}{t-1} \leq \sum_{t=2}^T \frac{|m_t|}{\sqrt{\sum_{\tau \leq t} |m_\tau|}} \leq 2\sqrt{d_{\text{tot}}}$, which concludes the regret analysis.

2.4 Delayed OCO with Exp-concave Losses

In this section, we consider the delayed OCO problem with exp-concave losses. Exp-concave losses are a more general class of loss functions that require more sophisticated techniques to be tackled. To address this problem, we design Algorithm 2.2, a variant of Online Newton Step (ONS) which effectively handles delayed feedback. Specifically, after receiving the gradients g_τ for all $\tau \in o_{t+1} \backslash o_t$, we select x_{t+1} as the minimizer of the cumulative surrogate loss over all the already observed gradients and the past actions, with an additive squared ℓ_2-regularization term. For simplicity, in this section we omit dependencies on curvature parameters, Lipschitz constants, and domain diameter; they appear explicitly in the theorem statements. The following result provides a first regret bound for Algorithm 2.2.

Algorithm 2.2: Delayed ONS for exp-concave functions

input: $\beta > 0$, learning rate rule $\{\eta_t\}_{t \geq 1}$,
initialize: $x_1 \in \mathcal{X}$
for $t = 1, 2, \ldots$ **do**
    Play x_t; receive $g_\tau = \nabla \ell_\tau(x_\tau)$ for all $\tau \in o_{t+1} \setminus o_t$
    Update

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \sum_{\tau \in o_{t+1}} \Big(\langle g_\tau, x \rangle + \frac{\beta}{2} \langle g_\tau, x - x_\tau \rangle^2 \Big) + \frac{\eta_t}{2}\|x\|_2^2 \tag{2.5}$$

Theorem 2.2. *Assume that $\ell_1, \ldots, \ell_T$ are α-exp-concave and let $\beta = \frac{1}{2}\min\{\frac{1}{4LD}, \alpha\}$. Then, under Assumption 2.1 and Assumption 2.2, Algorithm 2.2 with $0 < \eta_0 \leq \eta_1 \leq \cdots \leq \eta_T$ guarantees that*

$$\mathrm{Reg}_T = \mathcal{O}\Big(\frac{n}{\beta} \ln\Big(1 + \frac{\beta L^2 T}{\eta_0 n}\Big) + \eta_T D^2 + \min\{B_1, B_2\}\Big),$$

where $B_1 = \left(\frac{L^2}{\eta_0} + \frac{1}{\beta}\right) n d_{\max} \ln\left(1 + \frac{\beta L^2 T}{\eta_0 n}\right)$ and $B_2 = L^2 \sum_{t=1}^T \frac{|m_t|}{\eta_{t-1}}$.

We can now introduce two careful tunings of the time-varying learning rates $(\eta_t)_{t\geq 1}$ to derive the regret bounds $\mathcal{O}(d_{\max} n \ln T)$ and $\mathcal{O}(\sqrt{d_{\text{tot}}})$ individually.

Simple tuning. With a constant learning rate constant $\eta_t = 1$ for all $t \in [T]$, Algorithm 2.2 directly obtains $\mathcal{O}(d_{\max} n \ln T)$ regret. Alternatively, setting $\eta_t = \frac{L}{D}\sqrt{\sum_{s\leq t} |m_s| + |m_t| + 1}$ for all $t \geq 1$, Algorithm 2.2 achieves $\mathcal{O}(\sqrt{d_{\text{tot}}})$ regret; here, the $|m_t|+1$ term is an essentially tight worst-case estimation of $|m_{t+1}|$, since $m_{t+1} \subseteq m_t \cup \{t\}$.

Note that either of these two bounds can be significantly better than the other under different delay sequences, e.g., as shown by our Lemma A.10 in the appendix. Therefore, we ideally want to achieve $\mathcal{O}(\min\{d_{\max} n \ln T, \sqrt{d_{\text{tot}}}\})$ regret via a single choice of the learning rates. In fact, we can show that it is indeed possible to obtain such a bound by a careful delay-adaptive learning rate tuning.

Adaptive tuning. The adaptive learning rate is given by $\eta_0 = 1$ and $\eta_t = \min\{a_t, b_t\} + 1$ for all $t \geq 1$, where

$$a_t = \frac{2}{LD}\left(L^2 + \frac{1}{\beta}\right) n d_{\max}^{\leq t} \ln\left(1 + \frac{\beta L^2 T}{n}\right) , \tag{2.6}$$

$$b_t = \frac{L}{D}\sqrt{\sum_{s\leq t} |m_s| + |m_t| + 1} . \tag{2.7}$$

The overall idea behind this learning rate tuning is to keep track of both the $d_{\max} n \ln T$ and $\sqrt{d_{\text{tot}}}$ regret guarantees over the rounds via a_t and b_t, respectively. Then, η_t is set depending on the best of the two, i.e., $\min\{a_t, b_t\}$, which then leads to achieve the best of both regret bounds. Note that this adaptive tuning requires the knowledge of the time-stamps of the received gradients since we need to compute $d_{\max}^{\leq t} = \max_{\tau\leq t} \min\{d_\tau, t - \tau\}$ which, we recall, is the maximum delay that has been perceived up to round t. The following corollary provides a regret bound for Algorithm 2.2 with this adaptive tuning. The full proof of Corollary 2.1 can be found in Appendix A.3.

Corollary 2.1. *Assume that $\ell_1, \ldots, \ell_T$ are α-exp-concave and let $\beta = \frac{1}{2}\min\{\frac{1}{4LD}, \alpha\}$. Then, under Assumption 2.1 and Assumption 2.2, Algorithm 2.2 with the adaptive learning rate $\eta_t = \min\{a_t, b_t\}+1$, where a_t and b_t are defined in Equations* (2.6) *and* (2.7)*, guarantees that*

$$\mathrm{Reg}_T = \mathcal{O}\bigg(\frac{n}{\beta}\ln\Big(1 + \frac{\beta L^2 T}{n}\Big) + D^2 + \min\{C_1, C_2\}\bigg),$$

where $C_1 = \left(\frac{D}{L} + 1\right)\left(L^2 + \frac{1}{\beta}\right) n d_{\max} \ln\left(1 + \frac{\beta L^2 T}{n}\right)$ and $C_2 = \left(L^2 + LD\right)\left(\sqrt{d_{\text{tot}}} + 1\right)$.

Corollary 2.1 shows Algorithm 2.2 with the adaptive learning rate obtains regret $\mathcal{O}\big(\min\big\{d_{\max} n \ln T, \sqrt{d_{\text{tot}}}\big\}\big)$. The main advantage of an adaptive learning rate is that it requires no prior knowledge of d_{tot} or $d_{\max}$, nor does it rely on a doubling trick that would throw away information via resets.

2.4.1 Regret Analysis

In this section, we provide a proof sketch of Theorem 2.2 and Corollary 2.1, while their full proofs are deferred to Appendix A.3. Specifically, using the exp-concavity property and Lemma A.3, we

decompose the overall regret as follows:

$$\mathrm{Reg}_T(u) \le \underbrace{\sum_{t=1}^{T} \langle g_t, x_t^\star - u\rangle}_{\mathrm{Reg}_T^\star(u)} + \underbrace{\sum_{t=1}^{T} \langle g_t, x_t - x_t^\star\rangle}_{\texttt{Drift}_T} - \frac{\beta}{2}\sum_{t=1}^{T}\big(\langle g_t, x_t - u\rangle\big)^2 , \tag{2.8}$$

where we define $x_1^\star = x_1$ and, for $t \ge 2$, $x_t^\star = \arg\min_{x\in\mathcal{X}} \sum_{\tau=1}^{t-1}(\langle g_\tau, x\rangle + \frac{\beta}{2}\left(\langle g_\tau, x - x_\tau\rangle\right)^2) + \frac{\eta_{t-1}}{2}\|x\|_2^2$ to be the decisions assuming that all gradients before round t are observed.

For the term $\mathrm{Reg}_T^\star(u)$, following a standard FTRL analysis, we are able to obtain that

$$\mathrm{Reg}_T^\star(u) \le \frac{\eta_T}{2}\|u\|_2^2 + \frac{\beta}{2}\sum_{t=1}^{T}\left(\langle g_t, u - x_t\rangle\right)^2 + \sum_{t=1}^{T}\min\left\{LD, \|g_t\|_{A_{t-1}^{-1}}^2\right\}. \tag{2.9}$$

where $A_{t-1} = \eta_{t-1}I + \beta\sum_{\tau=1}^{t-1} g_\tau g_\tau^\top$. Applying Lattimore and Szepesvári (2020, Lemma 19.4), the last sum on the right-hand side of the above inequality satisfies

$$\sum_{t=1}^{T}\min\left\{LD, \|g_t\|_{A_{t-1}^{-1}}^2\right\} = \mathcal{O}\left(\frac{n}{\beta}\ln\left(1 + \frac{\beta L^2 T}{n}\right)\right). \tag{2.10}$$

Now we consider the $\texttt{Drift}_T$ term. By applying the Cauchy-Schwarz inequality followed by the stability lemma (Lemma D.1) again, it follows that for all $t \ge 1$,

$$\texttt{Drift}_T \le \sum_{t=1}^{T}\|g_t\|_{A_{t-1}^{-1}}\|x_t - x_t^\star\|_{A_{t-1}} \le 4\sum_{t=1}^{T}\|g_t\|_{A_{t-1}^{-1}}\left(\sum_{\tau\in m_t}\|g_\tau\|_{A_{t-1}^{-1}}\right). \tag{2.11}$$

By applying Lemma A.11, it holds that

$$\texttt{Drift}_T = \mathcal{O}\left(\left(L^2 + \frac{1}{\beta}\right) n d_{\max}\ln\left(1 + \frac{\beta L^2 T}{n}\right)\right). \tag{2.12}$$

At the same time, we can also prove that

$$\texttt{Drift}_T = \mathcal{O}\left(L^2\sum_{t=1}^{T}\frac{|m_t|}{\eta_{t-1}}\right). \tag{2.13}$$

Combing Equations (2.8) to (2.13) concludes the proof of Theorem 2.2. To prove Corollary 2.1, we carefully consider the adaptive learning rate tuning and separate the analysis into two cases. In case $a_T \le b_T$ at the end of the T rounds, we utilize a delayed version of the elliptical potential lemma (Lemma A.11) to achieve the logarithmic regret. On the other hand, if $b_T < a_T$ we split the regret analysis at the last round $\tau^\star$ at which $a_{\tau^\star} \le b_{\tau^\star}$. Then, we use again the logarithmic bound up to round $\tau^\star$ and the $\sqrt{d_{\mathrm{tot}}}$ bound for the remaining rounds. It suffices to observe that the first bound is no worse than $\sqrt{d_{\mathrm{tot}}}$ since $a_{\tau^\star} \le b_{\tau^\star}$ to conclude the proof.

2.5 Online Linear Regression with Delayed Labels

Here we consider the problem of online linear regression (OLR) with delays. This setting essentially corresponds to a variant of OCO where the domain is $\mathcal{X} = \mathbb{R}^n$ and loss functions are $\ell_t(x) = \frac{1}{2}(\langle z_t, x\rangle - y_t)^2$ comparing any point $x \in \mathbb{R}^n$ to a label $y_t \in \mathbb{R}$ given some feature vector $z_t \in \mathbb{R}^n$; to be precise, the predicted label by a given point x corresponds to the inner product $\langle z_t, x\rangle$. At each round t, the learner first observes an n-dimensional feature vector z_t before performing its prediction x_t, but the true label y_t is only revealed at a later round $t + d_t$. A common assumption on feature vectors and labels in this setting, analogous to the ones we introduced in Section 2.2 for instance, is their boundedness.

Assumption 2.3. *The feature vectors $z_1, \dots, z_T$ and the labels $y_1, \dots, y_T$ are bounded, i.e., $\|z_t\|_2 \le Z$ and $|y_t| \le Y$ for any $t \in [T]$, given $Y, Z \ge 0$.*

Algorithm 2.3: Delayed VAW forecaster with clipping

input: learning rate rule $\{\eta_t\}_{t\ge 1}$
initialize: $\rho_1 = 0$
for $t = 1, 2, \dots$ **do**
  Observe z_t
  Update

$$x_t = \operatorname*{arg\,min}_{x\in\mathbb{R}^n} \sum_{\tau\in o_t} -y_\tau\langle z_\tau, x\rangle + \frac{\eta_t}{2}\|x\|_2^2 + \frac{1}{2}\sum_{\tau\le t}\big(\langle z_\tau, x\rangle\big)^2 \tag{2.14}$$

  Play $\widetilde{x}_t = x_t \cdot \min\big\{\frac{\rho_t}{|\langle z_t, x_t\rangle|}, 1\big\}$
  Receive y_τ for all $\tau \in o_{t+1} \setminus o_t$
  Set $\rho_{t+1} = \max_{\tau\in o_{t+1}} |y_\tau|$

Note that the loss ℓ_t becomes exp-concave when the domain is also *bounded*. If this were the case, we could solve this problem by designing a version of ONS that can handle delayed labels. In OLR, however, the domain is unconstrained as it corresponds to the whole n-dimensional Euclidean space, which makes it particularly challenging to simply adapt one of the techniques seen so far without further assumptions. We instead design an algorithm for this problem (see Algorithm 2.3) that corresponds to an adaptation of the Vovk-Azoury-Warmuth (VAW) forecaster (Azoury and Warmuth, 2001, Vovk, 2001) in order to handle delayed labels. We can then prove that the regret guarantee for this algorithm in the delayed OLR setting becomes as stated in Theorem 2.3 below (whose proof is in Appendix A.4).

Theorem 2.3. *In the OLR problem with delayed labels under Assumption 2.3, Algorithm 2.3 guarantees for any $0 < \eta_0 \le \eta_1 \le \cdots \le \eta_T$ that*

$$\mathrm{Reg}_T(u) \le \frac{\eta_T}{2}\|u\|_2^2 + nY^2 \ln\!\left(1 + \frac{Z^2 T}{\eta_0 n}\right) + \mathcal{O}\Big(Y^2\big(\sigma_{\max} + \min\{M_1, M_2\}\big)\Big)\,,$$

where $M_1 = nd_{\max}\ln\big(1 + \frac{Z^2T}{\eta_0 n}\big)$ and $M_2 = Z^2\sum_{t=1}^T \frac{|m_t|}{\eta_t}$.

The idea behind the regret analysis is once again to decompose the regret into a cheating regret

term and a drift term:

$$\mathrm{Reg}_T(u) = \underbrace{\sum_{t=1}^{T}\big(\ell_t(x_t^\star) - \ell_t(u)\big)}_{\mathrm{Reg}_T^\star(u)} + \underbrace{\sum_{t=1}^{T}\big(\ell_t(\widetilde{x}_t) - \ell_t(x_t^\star)\big)}_{\texttt{Drift}_T},$$

where $(\widetilde{x}_t)_{t\geq 1}$ are the actions played by Algorithm 2.3, while $(x_t^\star)_{t\geq 1}$ are the "cheating" iterates that assume to have knowledge about all labels from past rounds. To bound the cheating regret $\mathrm{Reg}_T^\star(u)$, it is important to leverage the curvature of the squared loss. Specifically, by definition,

$$\mathrm{Reg}_T^\star(u) = \sum_{t=1}^{T}\frac{\langle z_t, x_t^\star\rangle^2 - \langle z_t, u\rangle^2}{2} + \sum_{t=1}^{T}\langle -y_t z_t, x_t^\star - u\rangle\,.$$

Then, we can study the second sum via the standard tools for the regret analysis of FTRL with respect to the linear losses $x \mapsto -y_t\langle z_t, x\rangle$, which yields

$$\mathrm{Reg}_T^\star(u) \leq \frac{\eta_T}{2}\|u\|_2^2 + nY^2\ln\!\left(1 + \frac{Z^2T}{\eta_0 n}\right).$$

This is exactly the first line in the regret guarantee presented in Theorem 2.3, and it corresponds to the part that does not depend on delays.

On the other hand, the drift term $\texttt{Drift}_T$ requires much more care and novel techniques. By the convexity of ℓ_t, we have that $\texttt{Drift}_T \leq \sum_{t=1}^{T}\langle \nabla\ell_t(\widetilde{x}_t), \widetilde{x}_t - x_t^\star\rangle$. Here we immediately observe the importance of the additional clipping of x_t to define the selected point $\widetilde{x}_t$, which is inspired from the clipping ideas by Cutkosky (2019), Mayo et al. (2022). Its scope is to guarantee that the predicted label $\langle z_t, \widetilde{x}_t\rangle$ falls within the range of true labels; the reason for this is to avoid the gradient of ℓ_t evaluated at $\widetilde{x}_t$ to blow up, otherwise obstructing an attempt to nicely bound $\texttt{Drift}_T$. We also remark that, differently form Mayo et al. (2022), we do not require to clip the labels used in the iterates update too. If we had knowledge of Y, we could use it to clip to the interval $[-Y, Y]$, thus guaranteeing $\ell_t(\widetilde{x}_t) \leq Y$. However, since we want to assume *no prior knowledge of* Y, the best clipping we can do at any time t is via ρ_t. Doing so requires to handle possible rounds when the label falls outside the clipping interval, which in turn requires a careful analysis that accounts for the feedback to be revealed only after some delay (as ρ_t could possibly be updated much later in time). We are then able to prove that

$$\texttt{Drift}_T = \mathcal{O}\big(Y^2\sigma_{\max} + Y^2\min\{M_1, M_2\}\big)\,.$$

which is the delay-dependent part of the regret; the $Y^2\sigma_{\max}$ term, in particular, is the one due to clipping mistakes.

Given any $\gamma > 0$, we may now set $\eta_0 = \gamma$ and $\eta_t = \gamma(\min\{a_t, b_t\} + 1)$ for all $t \geq 1$, where

$$a_t = 2nd_{\max}^{\leq t}\ln\left(1 + \frac{Z^2T}{\gamma n}\right),\ b_t = Z\sqrt{\textstyle\sum_{s\leq t}|m_s|}\,. \tag{2.15}$$

By doing so, we obtain the following Corollary 2.2 which provides a regret bound for Algorithm 2.3 with this adaptive tuning, and whose proof is deferred to Appendix A.4.

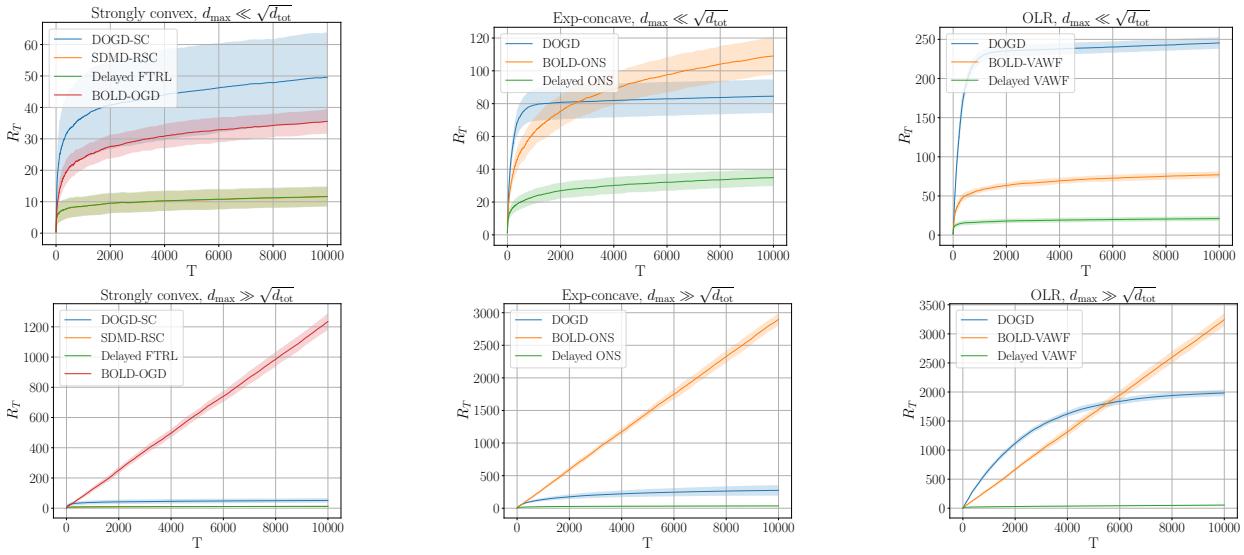


Figure 2.1: Comparison with relevant baselines. The shaded areas consider a range centered around the mean with half-width corresponding to the empirical standard deviation over 20 repetitions.

Corollary 2.2. *In the OLR problem with delayed labels under Assumption 2.3, Algorithm 2.3 with the adaptive learning rate* $\eta_t = \gamma(\min\{a_t, b_t\} + 1)$*, where* a_t *and* b_t *are defined in Equation* (2.15) *for any* $\gamma > 0$ *guarantees that*

$$\mathrm{Reg}_T \leq \frac{\gamma\|u\|_2^2}{2} + nY^2 \ln\left(1 + \frac{Z^2 T}{\gamma n}\right) + \mathcal{O}\big(\min\{Q_1, Q_2\}\big),$$

where $Q_1 = \big(\gamma\|u\|_2^2 + Y^2\big) n d_{\max} \ln\left(1 + \frac{Z^2 T}{\gamma n}\right)$ *and* $Q_2 = \big(\gamma Z\|u\|_2^2 + (Z+1)Y^2\big)\sqrt{d_{\text{tot}}}$.

To achieve this final result, we leverage similar ideas from the adaptive tuning for delayed ONS in Corollary 2.1, as mentioned above, together with a nontrivial relation between $\sigma_{\max}$ and $\sqrt{d_{\text{tot}}}$ to handle the additive $Y^2\sigma_{\max}$ term from the clipping errors (see Lemma A.7). We remark that here we used directly Z for the tuning, which requires its knowledge since the first round; we could easily do without this prior knowledge by using $Z_t = \max_{\tau \leq t}\|z_\tau\|_2$ instead because we always observe all the previous and the current feature vectors by the beginning of round t.

2.6 Experiments

In this section, we evaluate the performance of the proposed algorithms on three types of loss functions in the delayed OCO setting. All experiments are conducted over $T = 10000$ round and results are averaged over 20 independent trials. To showcase the advantage of our algorithms, we consider two delay regimes. For the first case, each delay d_t is independently and uniformly sampled from the set $\{0, 1, \ldots, 5\}$, thus leading to $\mathbb{E}[\sqrt{d_{\text{tot}}}] = \Theta(\sqrt{T})$ and $\mathbb{E}[\sigma_{\max}] \leq \mathbb{E}[d_{\max}] \leq 5$. In the second case, we define $p = T^{-1/3} = 0.1$. Then, for each t, d_t is sampled from the same distribution with probability $1 - p$, and it is set to be $d_t = T - t$ with probability p. In this case, $\mathbb{E}[\sqrt{d_{\text{tot}}}] = o(T)$, $\mathbb{E}[d_{\max}] \geq T(1 - (1-p)^T)$, and $\mathbb{E}[\sigma_{\max}] = \mathcal{O}(pT)$. We compare our algorithms against several baselines designed for delayed feedback settings. Below, we describe how we construct losses, together with the baseline algorithms we compare against. We provide additional experiments

in Appendix A.6.

Strongly convex loss. We consider the following strongly convex losses $\ell_t(x) = \frac{1}{2}(\langle z_t, x\rangle - y_t)^2 + \frac{1}{2}\|x\|_2^2$. The feasible set is the ball $\mathcal{X} = \{x \in \mathbb{R}^5, \|x\|_2 \leq 2\}$. Each coordinate of the feature vector $z_t \in \mathbb{R}^5$ at round t is uniformly chosen from $[-1, 1]$ while $y_t = \langle z_t, \mathbf{1}\rangle + \epsilon_t$, where ϵ_t is an i.i.d. standard Gaussian noise. We evaluate Algorithm 2.1 on this loss sequence and compare its performance with DOGD-SC (Wan et al., 2022a), SDMD-RSC (Wu et al., 2024, Algorithm 6), and BOLD-OGD which applies the reduction proposed by Joulani et al. (2013) to OGD.

Exp-concave loss. The loss functions we consider for exp-concave ones are $\ell_t(x) = \frac{1}{2}\big(\langle z_t, x\rangle - y_t\big)^2$. The other configurations are the same as the experiments in the strongly convex case. We evaluate our Algorithm 2.2 and compare its performance with that of DOGD (Quanrud and Khashabi, 2015) and BOLD-ONS, which applies the reduction proposed in Joulani et al. (2013) to ONS (Hazan et al., 2007).

Online linear regression. We still consider the loss function $\ell_t(x) = \frac{1}{2}\big(\langle z_t, x\rangle - y_t\big)^2$ for all $t \in [T]$, the same one as used in the exp-concave setting. The only difference is that the action space is now unconstrained ($\mathcal{X} = \mathbb{R}^5$). We empirically evaluate Algorithm 2.3 on this loss sequence and compare the performance with DOGD (Quanrud and Khashabi, 2015) and BOLD-VAW, which is again a combination of the reduction in Joulani et al. (2013) and the VAW forecaster (Azoury and Warmuth, 2001, Vovk, 2001).

Experimental results. Figure 2.1 shows the mean cumulative regret and its standard deviation over 20 rounds for the instances with strong convexity, exp-concavity, and OLR under the two previously mentioned delay regimes. For strongly convex losses, we find that our algorithm performs much better than DOGD-SC (Wan et al., 2022a) and have similar performances compared to SDMD-RSC, which is proven to only achieve $\mathcal{O}(d_{\max} \ln T)$ regret (Wu et al., 2024). However, we point out that this mismatch in the empirical performance and the theoretical guarantee of SDMD-RSC is due to a loose analysis of this algorithm. In fact, we show that SDMD-RSC can also achieve the same $\mathcal{O}(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\})$ regret via a refined analysis. The proof is deferred to Appendix A.5.

For both exp-concave and OLR settings, our algorithms consistently outperform DOGD, which does not leverage the curvature of the loss function, as well as the reduction-based algorithms proposed in Joulani et al. (2013), under both delay regimes, showing the effectiveness of our algorithms under different delay conditions.

Chapter 3

Exploiting Intermediate Feedback in Multi-Armed Bandits with Delayed Feedback

3.1 Introduction

The impact of delay on the performance of sequential decision makers, measured by regret, has been extensively studied under full information and bandit feedback, and in stochastic and adversarial environments (Joulani et al., 2013, Pike-Burke et al., 2018, Lancewicki et al., 2021, Zimmert and Seldin, 2020a). Yet, in many real-life situations, *intermediate observations* may be available to the learner. For example, a health check-up might give a preliminary indication on the effect of a treatment, an advertisement click might be a precursor for an upcoming purchase, and preliminary reviews might provide some information regarding an upcoming acceptance or rejection decision. In this chapter, we investigate when and how intermediate observations can be used to reduce the impact of delays in observing the final outcome of an action in a multi-armed bandit setting.

Online learning with delayed feedback and intermediate observations was studied by Mann et al. (2019) in a full-information setting, and subsequently by Vernade et al. (2020) in a non-stationary stochastic bandit setting. In the paper of Vernade et al. (2020), at each round the learner chooses an action and immediately observes a signal (also called state) belonging to a finite set. The actual loss (i.e., feedback) incurred by the learner in that time step is only received with delay, which can be fixed or random. More formally, the observed state is drawn from a distribution that only depends on the chosen action, and the incurred loss is drawn from a distribution that only depends on the observed state (and not on the chosen action), forming a Markov chain.

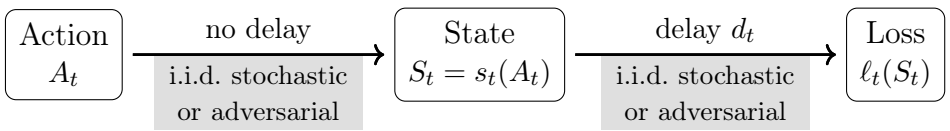


Figure 3.1: Scheme depicting the delayed feedback setting with intermediate observations.

The work of Vernade et al. (2020) studies a setting where mappings s_t from actions to states are non-stationary and losses ℓ_t over states are i.i.d. stochastic. In this chapter, we instead consider

two possible regimes for the action-state mappings s_t (stochastic and adversarial) and two possible regimes for the mappings ℓ_t from states to losses (also stochastic and adversarial). Altogether, we study four different regimes, defined by the combination of the first and the second mapping type (see Figure 3.1).

We characterize (within logarithmic factors) the minimax regret rates for all of them, by giving upper and lower bounds. Similar to Vernade et al. (2020), we assume that the states are observed instantaneously, and that the losses are observed with some delay $d \in \mathbb{N}$. We show that the minimax regret rate is fully determined by the regime of the state-loss mapping, regardless of the regime of the action-state mapping. The results are informally summarized in Table 3.1, where K denotes the number of actions, S denotes the number of states, and T denotes the time horizon. It is assumed that the losses belong to the $[0, 1]$ interval. All of our upper bounds hold with high probability (with respect to the learner's internal randomization) irrespective of the regime of the action-state mapping.

State-loss mapping	Regret bounds	References
Adversarial	$\sqrt{dT} + \sqrt{KT}$	Cesa-Bianchi et al. (2019) Theorem 3.7
Stochastic	$\min\left\{\sqrt{ST} + d\sqrt{S}, \sqrt{dT}\right\} + \sqrt{KT}$	Theorems 3.2 and 3.3 Corollary 3.2

Table 3.1: Summary of our results with fixed delay d, ignoring logarithmic factors.

We recall that, up to logarithmic factors, the minimax regret rate in multi-armed bandits with delays without intermediate observations is of order $\sqrt{(K+d)T}$ (Cesa-Bianchi et al., 2019, Zimmert and Seldin, 2020a). Therefore, given our findings we conclude that, if the mapping from states to losses is adversarial, then intermediate observations do not help (in the minimax sense) because the regret rates are the same irrespective of whether the intermediate observations are used or not, and irrespective of whether the mapping from actions to states is stochastic or adversarial. However, if the mapping from states to losses is stochastic, and the number S of states is smaller than the delay d, then intermediate observations are helpful, and we provide an algorithm, `AdaMetaBIO`, which is able to exploit them. Our result improves on the $\widetilde{\mathcal{O}}\big(\sqrt{KST}\big)$ regret bound obtained by Vernade et al. (2020) for the case of stochastic and stationary action-state mapping.

Our algorithm also applies to a more general setting of non-uniform delays $(d_t)_{t\in[T]}$ where we achieve a high-probability regret bound of order $\sqrt{KT + \min\{ST, d_{\text{tot}}\}}$, ignoring logarithmic factors once more and terms not depending on T. This improves upon the total delay term $d_{\text{tot}} = d_1 + \cdots + d_T$ similarly to the respective term in the fixed delay setting.

Roadmap. We provide a formal definition of the problem in Section 3.2. In Section 3.3, we introduce two algorithms, `MetaBIO` and `AdaMetaBIO`, for the model of bandits with intermediate observations. Section 3.4 contains the analysis of both algorithms, where we prove high-probability regret bounds for the setting of adversarial action-state mappings and stochastic losses. We provide regret lower bounds in Section 3.5, and experimental validation of our results in Section 3.6.

3.1.1 Related Works

Adaptive clinical trials have served an inspiration for the multi-armed bandit model (Thompson, 1933) and, interestingly, they have also pushed the field to study the effect of delayed feedback (Simon, 1977, Eick, 1988). In the bandit setting, Joulani et al. (2013) have studied a stochastic setting with random delays, whereas Neu et al. (2010, 2014) have studied an adversarial setting with constant delays. Cesa-Bianchi et al. (2019) have shown an $\Omega(\max\{\sqrt{KT}, \sqrt{dT \ln K}\})$ lower bound for adversarial bandits with uniformly delayed feedback, and an upper bound matching the lower bound within logarithmic factors by using an Exp3-style algorithm (Auer et al., 2002b), whereas Zimmert and Seldin (2020a) have reduced the gap to the lower bound down to constants by using a Tsallis-INF approach (Zimmert and Seldin, 2021). Follow up works have studied adversarial multi-armed bandits with non-uniform delays (Thune et al., 2019, Bistritz et al., 2019, 2022a, György and Joulani, 2021, van der Hoeven and Cesa-Bianchi, 2022b) with Zimmert and Seldin (2020a) providing a near-optimal algorithm, and Masoudian et al. (2022a) and Masoudian et al. (2024a) deriving best-of-both-worlds extensions and a matching lower bound for special sequences of delays. Two key techniques for handling non-uniform delays are the skipping technique, introduced by Thune et al. (2019), and algorithm parametrization by the number of outstanding observations (an observed quantity at action time related to delays), as opposed to the delays (an unobserved quantity at action time), introduced by Zimmert and Seldin (2020a). Finally, the presence of delays has been further considered in more complex extensions of multi-armed bandits (van der Hoeven et al., 2023b).

3.2 Problem Setting

We consider an online learning setting with a finite set $\mathcal{A} \coloneqq [K]$ of $K \geq 2$ actions and a finite set $\mathcal{S} \coloneqq [S]$ of $S \geq 2$ states. In each round $t \in [T]$, the learner picks an action $A_t \in \mathcal{A}$ and receives a state $S_t \coloneqq s_t(A_t) \in \mathcal{S}$ as an intermediate observation according to some unknown action-state mapping $s_t \in \mathcal{S}^{\mathcal{A}}$. The learner then incurs a loss $\ell_t(S_t) \in [0,1]$, which *exclusively* depends on the state associated to the selected action and is only observed at the end of round $t + d_t$, where the delay $d_t \geq 0$ is (fully) revealed to the learner only when the loss observation is received. The difficulty of this learning task depends on three elements, all initially unknown to the learner:

- the sequence of action-state mappings $s_1, \ldots, s_T \in \mathcal{S}^{\mathcal{A}}$;
- the sequence of loss vectors $\ell_1, \ldots, \ell_T \in [0,1]^{\mathcal{S}}$;
- the sequence of delays $d_1, \ldots, d_T \in \mathbb{N}$, where $d_t \leq T - t$ for all $t \in [T]$ without loss of generality.

Note that unlike standard bandits, as remarked above, here the losses are functions of the states instead of the actions. However, since actions are chosen without a-priori information on the action-state mappings, learners have no direct control on the losses they will incur and, because of the delays, they also have no immediate feedback on the loss associated with the observed states. Note also that, for all $t \geq 1$, the states $s_t(a)$ for $a \neq A_t$ and the losses $\ell_t(s)$ for $s \neq S_t$ are never revealed to the algorithm. For brevity, we refer to this setting as (delayed) Bandits with Intermediate Observations (BIO).

In the setting of stochastic losses, we assume the loss vectors $\ell_t \in [0,1]^{\mathcal{S}}$ are sampled i.i.d. from some fixed but unknown distribution Q, and let $\theta \in [0,1]^{\mathcal{S}}$ be the unknown vector of expected losses for the states. That is, $\ell_t(s) \sim Q(\cdot \mid s)$ has mean $\theta(s)$ for each $t \in [T]$ and $s \in \mathcal{S}$. Note

that we allow dependencies between the stochastic losses of distinct states in the same round, but require losses to be independent across rounds. In the setting of stochastic action-state mappings, we assume that each observed state S_t is independently drawn from a fixed but unknown distribution $P(\cdot \mid A_t)$. If both losses and action-state mappings are stochastic, then $\ell_t(S_t)$ is independent of A_t given S_t. When losses or action-state mappings are adversarial, we assume an oblivious adversary as in previous chapters.

Our main quantity of interest is the regret measured via the learner's cumulative loss $\sum_{t=1}^{T} \ell_t(S_t)$, where $S_t = s_t(A_t)$ and $(A_t)_{t\in[T]}$ is the sequence of actions chosen by the learner. In the case of stochastic losses, we define the performance of the learner by $\sum_{t=1}^{T} \theta(S_t)$. In the case of stochastic action-state mappings, we average each instantaneous loss over the random choice of the state: $\sum_s \ell_t(s)P(s \mid A_t)$ for adversarial losses and $\sum_s \theta(s)P(s \mid A_t)$ for stochastic losses. Regret is always computed according to the best fixed action in hindsight with respect to some appropriate notion of cumulative loss. In particular, for stochastic state-action mappings, the cumulative losses of the best action are

$$\min_{a\in\mathcal{A}} \sum_{t=1}^{T} \sum_{s\in\mathcal{S}} \ell_t(s)P(s \mid a) \qquad \text{and} \qquad \min_{a\in\mathcal{A}} \sum_{t=1}^{T} \sum_{s\in\mathcal{S}} \theta(s)P(s \mid a)\ ,$$

respectively, whereas for adversarial state-action mappings they are, intuitively,

$$\min_{a\in\mathcal{A}} \sum_{t=1}^{T} \ell_t(s_t(a)) \qquad \text{and} \qquad \min_{a\in\mathcal{A}} \sum_{t=1}^{T} \theta(s_t(a))\ .$$

3.3 A Reduction to Standard Delayed Feedback

In this section, we introduce `MetaBIO` (Algorithm 3.1), a meta-algorithm that transforms any algorithm $\mathcal{B}$ tailored for the delayed setting *without* intermediate observations into an algorithm for our setting. We then propose `AdaMetaBIO`, a modification of `MetaBIO` that delivers an improved regret bound for our setting. The idea of `MetaBIO` is to reduce the impact of delays using the information we get from intermediate observations. More precisely, if we have *enough* observations for the current state S_t at time t, we immediately feed to $\mathcal{B}$ an *estimate* of the mean loss of this state as if it were the actual loss at time t; otherwise, we wait for d_t time steps and refine our estimate using the additional loss observations.

The are two key steps in the design of our algorithm: *how* we construct the mean estimate and *when* we use it instead of waiting for the actual loss. They are the steps highlighted in green in Algorithm 3.1 (Lines 10 and 16). For all $t \in [T]$ and all $s \in \mathcal{S}$, we use $\widetilde{\theta}_t(s)$ to denote the estimate of $\theta(s)$ at round t and $n_t(s)$ to denote the number of observations for state s that we want to observe before using $\widetilde{\theta}_t(s)$. We add a subscript t to $\mathcal{L}(s)$ in Algorithm 3.1 to denote the set of loss observations $\mathcal{L}_t(s) \coloneqq \{(j, \ell_j(s)) : j + d_j \le t, S_j = s\}$ for state s that we have collected by the end of round t. Thus, $\widetilde{\theta}_t(s)$ is computed by using $N_t(s) \coloneqq |\mathcal{L}_t(s)|$ loss observations.

Fixed delay setting. When all rounds have delay d, we simply choose $n_t(s) \coloneqq d$ for all $s \in \mathcal{S}, t \in [T]$. In other words, if we have at least d observations for some state, then we can compensate for the effect of delays and construct a well-concentrated mean estimate around the actual mean. Let $\widehat{\theta}_t(s) \coloneqq \sum_{j\in\mathcal{L}_t(s)} \ell_j(s)/N_t(s)$. Then our mean loss estimate is a lower confidence bound for $\theta(s)$

Algorithm 3.1: `MetaBIO`

input: Algorithm $\mathcal{B}$ for standard delayed bandits, confidence parameter $\delta \in (0,1)$
initialize $\mathcal{L}(s) \leftarrow \emptyset$ for all $s \in \mathcal{S}$
for $t = 1, \dots, T$ **do**
  get A_t from $\mathcal{B}$ and play it
  observe $S_t = s_t(A_t)$
  for $j : j + d_j = t$ **do**
    receive $(j, \ell_j(S_j))$
    update $\mathcal{L}(S_j) \leftarrow \mathcal{L}(S_j) \cup \{(j, \ell_j(S_j))\}$
  initialize feedback set $\mathcal{M}_t \leftarrow \emptyset$
  compute $n_t(S_t)$
  if $|\mathcal{L}(S_t)| \geq n_t(S_t)$ **then**
    add t to $\mathcal{M}_t$
  for $j : j + d_j = t \wedge |\mathcal{L}(S_j)| < n_j(S_j)$ **do**
    add j to $\mathcal{M}_t$
  for $j \in \mathcal{M}_t$ **do**
    compute $\widetilde{\theta}_j(S_j)$ from $\mathcal{L}(S_j)$ ▷ using δ
    feed $\big(j, A_j, \widetilde{\theta}_j(S_j)\big)$ to $\mathcal{B}$

defined by

$$\widetilde{\theta}_t(s) \coloneqq \max\left\{0, \widehat{\theta}_t(s) - \frac{1}{2}\varepsilon_t(s)\right\} \tag{3.1}$$

for $\varepsilon_t(s) \coloneqq \sqrt{\frac{2}{N_t(s)} \ln \frac{4ST}{\delta}}$.

Arbitrary delay setting. In the arbitrary delay setting, where we do not have preliminary knowledge of delays, we cannot really use the delays to set $n_t(s)$. Instead, at the *end* of time t, we have access to the number of outstanding observations $\sigma_t \coloneqq \big|\{j \in [t] \,:\, j + d_j > t\}\big|$, which is the number of yet-to-arrive loss observations at the end of round t.* Then, for any $s \in \mathcal{S}$, we may set $n_t(s) \coloneqq \sigma_t$. With this choice, incurring zero delay at some round implies that we received at least half of all the loss observations we could have received in the no-delay setting (see Appendix B.2.4). In Section 3.4 we see that this ensures our mean estimate is well concentrated around its mean.

Since Algorithm 3.1 waits for the actual loss at time t only if $N_t(S_t) < \sigma_t$, then $\widetilde{d}_t \coloneqq d_t \mathbb{I}\{N_t(S_t) < \sigma_t\}$ is the actual delay incurred by the algorithm, and $\mathcal{L}_{t+\widetilde{d}_t}(s)$ is the set of loss observations used to compute the estimate of the mean loss at time t. Because some losses may arrive at the same time, the high-probability analysis of `MetaBIO` requires these observations to be ordered. More precisely, we construct our mean estimate at time $t + \widetilde{d}_t$ for the feedback of round t using the set

$$\mathcal{L}'_t(s) \coloneqq \left\{(j, \ell_j(s)) \in \mathcal{L}_{t+\widetilde{d}_t}(s) \,\middle|\, j + \widetilde{d}_j < t + \widetilde{d}_t \,\vee\, j < t\right\}. \tag{3.2}$$

Letting $N'_t(s) \coloneqq |\mathcal{L}'_t(s)|$, we define the empirical mean

$$\widehat{\theta}_t(s) \coloneqq \sum_{j \in \mathcal{L}'_t(s)} \frac{\ell_j(s)}{N'_t(s)}. \tag{3.3}$$

*This differs from prior work that considers outstanding observations at the *beginning* of the round.

Then, we set $\varepsilon_t(s) \coloneqq \sqrt{\frac{2}{N'_t(s)} \ln \frac{4ST}{\delta}}$ and define the mean loss estimator $\widetilde{\theta}_t(s)$ as a lower confidence bound similarly to Equation (3.1). We remark that, while $\widetilde{\theta}_t(s)$ is employed for the estimation of the mean loss $\theta_t(s)$ of the state s, the estimator is only ever adopted starting from time $t + \widetilde{d}_t$ with some (possibly nonzero) delay $\widetilde{d}_t$. We may thus use all the collected losses in $\mathcal{L}'_t(s) \subseteq \mathcal{L}_{t+\widetilde{d}_t}(s)$ for its definition. Therefore, once receiving the losses at the end of round t, Algorithm 3.1 constructs the estimator $\widetilde{\theta}_j(S_j)$ for the incurred loss at any (previous) round $j \in \mathcal{M}_t$ from the feedback set $\mathcal{M}_t$ using as much information as possible gathered thus far, i.e., losses in $\mathcal{L}'_j(S_j) \subseteq \mathcal{L}_t(S_j)$.

The `AdaMetaBIO` algorithm. As we already anticipated, the goal of intermediate observations is to reduce the impact of delays. However, if the number of states is too large compared to the average delay, then the information we get from intermediate observations could be misleading. We introduce `AdaMetaBIO` (Algorithm 3.2) to address this issue. Given a horizon T,[†] this algorithm runs $\mathcal{B}$ (which is tailored for the setting *without* intermediate observations) until the total incurred delay exceeds ST, and then switches to `MetaBIO`. We precise that `AdaMetaBIO` computes $\mathfrak{D}_t \coloneqq \sum_{j \le t} \sigma_j$ as the sum of outstanding observation counts up to round t, which is then used in the switching condition.

Algorithm 3.2: `AdaMetaBIO`

input: Algorithm $\mathcal{B}$ for standard delayed bandits, confidence parameter $\delta \in (0, 1)$
initialize $\mathfrak{D}_0 \leftarrow 0$
for $t = 1, \dots, T$ **do**
  get A_t from $\mathcal{B}$
  for $j : j + d_j = t$ **do**
    receive $(j, \ell_j(S_j))$
    feed $(j, A_j, \ell_j(S_j))$ to $\mathcal{B}$
  set $\sigma_t \leftarrow \sum_{j=1}^{t-1} \mathbb{I}\{j + d_j > t\}$
  update $\mathfrak{D}_t \leftarrow \mathfrak{D}_{t-1} + \sigma_t$
  if $\mathfrak{D}_t \left(3 \ln K + \ln(6/\delta)\right) > 49ST \ln \frac{8ST}{\delta}$ **then**
    break
if $t < T$ **then**
  run `MetaBIO`$(\mathcal{B}, \delta/2)$ for the remaining rounds

3.4 Regret Analysis

We analyze `MetaBIO` and `AdaMetaBIO` in the setting of adversarial action-state mappings and stochastic losses where the regret is defined by

$$\mathrm{Reg}_T \coloneqq \sum_{t=1}^{T} \theta(S_t) - \min_{a \in \mathcal{A}} \sum_{t=1}^{T} \theta(s_t(a)) \ .$$

Our analysis guarantees a bound on Reg_T that holds with high probability (and not just in expectation), hence the reason why Reg_T is not defined by taking the expectation over the internal randomization of the learner or the stochasticity of the environment (as done in all previous chapters).

[†]Note that we may remove the a-priori knowledge of T by using a doubling trick at the cost of a polylog factor in the regret. See Remark 3.1 for further details.

A related notion of regret is

$$\mathcal{R}eg_T \coloneqq \sum_{t=1}^{T} \ell_t(S_t) - \min_{a \in \mathcal{A}} \sum_{t=1}^{T} \ell_t(s_t(a)) \ ,$$

which considers the realized losses instead of their means. The two quantities are close with high probability: each inequality in

$$-\sqrt{2T \ln(2K/\delta)} \leq \mathrm{Reg}_T - \mathcal{R}eg_T \leq \sqrt{2T \ln(2/\delta)} \tag{3.4}$$

individually holds with probability at least $1 - \delta$ for any given $\delta \in (0, 1)$; see Lemma B.1 in Appendix B.

Let $d_{\mathrm{tot}} \coloneqq \sum_{t=1}^{T} d_t$ be the total delay. We start by showing an upper bound on the total actual (or effective) delay $\widetilde{d}_{\mathrm{tot}} = \sum_{t=1}^{T} d_t \mathbb{I}\{N_t(S_t) < \sigma_t\} \leq d_{\mathrm{tot}}$ incurred by `MetaBIO`. Then, we provide a high-probability regret analysis of both `MetaBIO` and `AdaMetaBIO`.

More precisely, we can show that `MetaBIO` incurs the delays of no more than $\min\{2S\sigma_{\max}, T\}$ rounds, where $\sigma_{\max} \coloneqq \max_{t \in [T]} \sigma_t$. In the worst case, these rounds correspond with those from the set

$$\Phi \in \operatorname*{arg\,max}_{\mathcal{J} \subseteq [T]} \Big\{ d_{\mathcal{J}} : |\mathcal{J}| = \min\{2S\sigma_{\max}, T\} \Big\} \ . \tag{3.5}$$

where we denote $d_{\mathcal{J}} \coloneqq \sum_{t \in \mathcal{J}} d_t$ for any $\mathcal{J} \subseteq [T]$.

Note that the set Φ is fully determined by the delay sequence $d_1, \ldots, d_T$. Moreover, the total delay incurred by `MetaBIO` cannot be worse than the sum of delays corresponding to the rounds in Φ, as stated in the lemma below.

Lemma 3.1 (Total effective delay)**.** *If* `MetaBIO` *is run with any algorithm $\mathcal{B}$ on delays $(d_t)_{t \in [T]}$, then its total effective delay is $\widetilde{d}_{\mathrm{tot}} \leq d_\Phi$.*

Lemma 3.1 (proof in Appendix B.2.1) implies that, if all delays are bounded by $d_{\max}$, then $\widetilde{d}_{\mathrm{tot}} \leq 2S\sigma_{\max} d_{\max}$, which does not depend on T. In the fixed-delay setting with delay d, for example, we get a total effective delay of at most $2Sd^2$, rather than the total delay dT we would incur without access to intermediate observations (when T is large enough).

We now turn `MetaBIO` into a concrete algorithm by instantiating $\mathcal{B}$. Specifically, we use `DAda-Exp3` (György and Joulani, 2021), a variant of `Exp3` which does not use intermediate observations and is robust to delays. `DAda-Exp3` guarantees the following regret bound.

Theorem 3.1 (György and Joulani (2021, Corollary 4.2))**.** *For any $\delta \in (0, 1)$, the regret of* `DAda-Exp3` *with respect to the realized losses in the adversarial bandits with arbitrary delays satisfies*

$$\mathcal{R}eg_T \leq 2\sqrt{3(2KT + d_{\mathrm{tot}}) \ln K} + \left(\sqrt{\frac{2KT + d_{\mathrm{tot}}}{3 \ln K}} + \frac{\sigma_{\max}}{2} + 1 \right) \ln \frac{2}{\delta}$$

with probability at least $1 - \delta$.

While Theorem 3.1 shows a high-probability bound on $\mathcal{R}eg_T$, Equation (3.4) shows that a high-probability bound for one notion of regret ensures a high-probability bound for the other. Although the original bound by György and Joulani (2021) was stated with $d_{\max}$ instead of $\sigma_{\max}$,

we can replace the former with the latter by observing that, in the analysis of György and Joulani (2021, Theorem 4.1), they only use $d_{\max}$ to upper bound the number of outstanding observations. Note that $\sigma_{\max}$ is never larger than $d_{\max}$, indicating it is a well-behaved term that is not vulnerable to a few large delays. See Masoudian et al. (2022a, Lemma 3) for a refined quantification of the relation between $\sigma_{\max}$ and $d_{\max}$.

If we consider a fixed confidence level $\delta \in (0,1)$, then we can make the learning rate η_t and the implicit-exploration term γ_t in `DAda-Exp3` depend on the specific value of δ so as to achieve an improved regret bound (see Appendix B.2.2). This allows us to show that in the BIO setting with adversarial action-state mappings and stochastic losses, the regret $\mathcal{R}eg_T$ of `DAda-Exp3` is bounded from above by

$$2\sqrt{2KTC_{K,6\delta}} + 2\sqrt{d_{\text{tot}}C_{K,6\delta}} + \frac{\sigma_{\max}+2}{2}\ln\frac{2}{\delta} \tag{3.6}$$

with probability at least $1-\delta$, where

$$C_{K,\delta} \coloneqq 3\ln K + \ln\frac{12}{\delta} \tag{3.7}$$

is a negligible logarithmic factor in K and $1/\delta$ only.

Next, we state the regret bound for `MetaBIO`. We remark that we initialize `DAda-Exp3` with confidence parameter $\delta/2$ so as to guarantee the high-probability bound as in Equation (3.6) with probability at least $1-\delta/2$ as required.

Theorem 3.2. *Let $\delta \in (0,1)$. If we run* `MetaBIO` *using* `DAda-Exp3`*, then the regret of* `MetaBIO` *in the BIO setting with adversarial action-state mappings and stochastic losses satisfies*

$$\mathrm{Reg}_T \le 2\sqrt{2KTC_{K,3\delta}} + 7\sqrt{ST\ln\frac{4ST}{\delta}} + 2\sqrt{d_\Phi C_{K,3\delta}} + \frac{\sigma_{\max}+2}{2}\ln\frac{4}{\delta} \tag{3.8}$$

with probability at least $1-\delta$.

We begin the analysis of Theorem 3.2 by decomposing the regret into two parts: (i) the regret $\mathcal{R}eg_T$ of `DAda-Exp3` with losses $\widetilde{\theta}_t(S_t)$, and (ii) the gap $\mathrm{Reg}_T - \mathcal{R}eg_T$, corresponding to the cumulative error of the estimates fed to `DAda-Exp3`. For the first part, we follow an approach similar to György and Joulani (2021) and apply Neu (2015, Lemma 1) to obtain a concentration bound for the loss estimates defined using importance weighting along with implicit exploration. When using the actual losses, the application of Neu (2015, Lemma 1) is straightforward. However, when the mean loss estimate $\widetilde{\theta}_t(S_t)$ is used rather than the actual loss, there is a potential dependency between the chosen action A_t and $\widetilde{\theta}_t(S_t)$. In Appendix B.2.3 we carefully design a filtration to show that we may indeed use the high-probability regret bound of `DAda-Exp3` in order to upper bound the first part (regret $\mathcal{R}eg_T$ defined in terms of the estimates $\widetilde{\theta}_t$).

The second part requires to bound the cumulative error of our estimator in Equation (3.3) for the observed states $(S_t)_{t\in[T]}$. To this end, we use the Azuma-Hoeffding inequality to control the error of these estimates. Doing so causes a $\widetilde{\mathcal{O}}(\sqrt{ST})$ term to appear in the regret bound. The detailed proof of this part is in Appendix B.2.4, together with the proof of Theorem 3.2.

The presence of the additive $\widetilde{\mathcal{O}}(\sqrt{ST})$ term in the regret bound implies that, when $S \gg \max\{d_{\text{tot}}/T, K\}$, using intermediate feedback leads to no advantage over ignoring it. So we ideally

want to recover the original bound in Equation (3.6) when this happens. `AdaMetaBIO` is an adaptive extension of `MetaBIO` that solves this issue and gives the following regret guarantee. The proof of this result is deferred to Appendix B.2.5. We remark that, to achieve this bound, before the eventual switch at some round t^* we use algorithm `DAda-Exp3` with confidence parameter set to $\delta/3$ so as to guarantee a high-probability bound on Reg_{t^*} with probability at least $1-\delta/2$ over the first t^* rounds (during which `DAda-Exp3` runs by itself).

Theorem 3.3. *Let $\delta \in (0,1)$. If we run* `AdaMetaBIO` *with* `DAda-Exp3`*, then the regret of* `AdaMetaBIO` *in the BIO setting with adversarial action-state mappings and stochastic losses satisfies*

$$\mathrm{Reg}_T \leq 3\min\left\{7\sqrt{ST\ln\frac{8ST}{\delta}}, \sqrt{d_{\mathrm{tot}}C_{K,2\delta}}\right\} + 6\sqrt{KTC_{K,2\delta}} + 2\sqrt{d_\Phi C_{K,2\delta}} + (\sigma_{\max}+2)\ln\frac{8}{\delta} \quad (3.9)$$

with probability at least $1-\delta$.

If we consider any upper bound $d_{\max}$ on the delays $(d_t)_{t\in[T]}$, we can further observe that the regret Reg_T of `AdaMetaBIO` (with `DAda-Exp3`) satisfies

$$\mathrm{Reg}_T = \widetilde{\mathcal{O}}\Big(\sqrt{KT} + \min\Big\{\sqrt{S}\big(\sqrt{T}+d_{\max}\big), \sqrt{d_{\max}T}\Big\}\Big)$$

with high probability. This also follows from the fact that, as previously mentioned, we can bound the total delay of `MetaBIO` by $d_\Phi \leq 2Sd_{\max}^2$.

Given the previous regret bounds, we observe that we may further improve the dependency on the delays by adopting the idea of skipping rounds with large delays when computing the learning rates. This "skipping" idea was introduced by Thune et al. (2019) and has been leveraged by György and Joulani (2021) to show that `DAda-Exp3` can achieve a refined high-probability regret bound—see György and Joulani (2021, Theorem 5.1). As a consequence, we can indeed provide an improved bound in our setting by following similar steps as in the proof of Theorem 3.2. The only main change is the adoption of the version of `DAda-Exp3` that uses the skipping procedure.

Corollary 3.1. *Let $\delta \in (0,1)$. If we run* `MetaBIO` *with* `DAda-Exp3` *with skipping (György and Joulani, 2021, Theorem 5.1), then the regret of* `MetaBIO` *in the BIO setting with adversarial action-state mappings and stochastic losses satisfies*

$$\mathrm{Reg}_T = \mathcal{O}\Bigg(\sqrt{KTC_{K,\delta}} + \sqrt{ST\ln\frac{ST}{\delta}} + \ln\frac{1}{\delta} + \sqrt{C_{K,\delta}\ln K}\min_{R\subseteq\Phi}\Big\{|R| + \sqrt{d_{\Phi\setminus R}\ln K}\Big\}\Bigg)$$

with probability at least $1-\delta$.

This result could also be extended in a similar way to `AdaMetaBIO`, so as to achieve the best result from the presence of intermediate feedback.

So far, we have provided some high-probability guarantees for the regret of both `MetaBIO` and `AdaMetaBIO`, by which we can derive some expectation bounds as well (e.g., by setting $\delta \approx 1/T$). However, using the empirical mean estimators $\widehat{\theta}_t$ as the mean loss estimators at time t and working directly with the expected regret allows us to improve the achievable bound by a polylogarithmic factor. Hence, for the expected regret we use `Tsallis-INF` (Zimmert and Seldin, 2020a), a learning algorithm for the standard delayed bandit problem that uses a hybrid regularizer to deal with delays

and gives a minimax-optimal expected regret bound in the standard delayed setting. The proof of this expected regret upper bound is in Appendix B.2.6.

Proposition 3.1. *If we execute* `AdaMetaBIO` *with* `Tsallis-INF` *(Zimmert and Seldin, 2020a), and use the switching condition* $\sqrt{8\mathfrak{D}_t \ln K} > 6\sqrt{ST\ln(2ST)}$ *at each round* $t \in [T]$*, where* $\mathfrak{D}_t = \sum_{j=1}^{t} \sigma_j$*, then the regret of* `AdaMetaBIO` *in the BIO setting with adversarial action-state mappings and stochastic losses satisfies*

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq 4\sqrt{2KT} + 2\sqrt{2d_\Phi \ln K} + 4\min\left\{3\sqrt{ST\ln(2ST)}, \sqrt{2d_{\mathrm{tot}} \ln K}\right\}.$$

Remark 3.1. *In* `MetaBIO`*, we can replace* T *by* t^2 *in the definition of the confidence intervals for Equation* (3.3) *and remove the need for prior knowledge of the time horizon* T*. In* `AdaMetaBIO`*, we could use a doubling trick to avoid the prior knowledge of* T *in the switching condition. On the other hand, it is not required to know the number of states* S *for expectation bounds on the regret of* `MetaBIO`*. However, removing the prior knowledge of* S *in the high-probability regret bounds is challenging. Indeed, to the best of our knowledge, there is no result in the BIO setting that avoids prior knowledge on the number of states. Lifting this requirement in the high-probability analysis is thus an interesting question for future work.*

3.5 Lower Bounds

The lower bounds in this section are for the expected regret $\mathbb{E}\left[\mathrm{Reg}_T\right]$. Since our algorithms provide high-probability guarantees, the upper bounds also apply to the expected regret. Throughout this section we will make use of constant delay, i.e., $d_t = d$ for all $t \in [T]$. We will first prove a general $\sqrt{KT}$ lower bound for all algorithms in BIO, after which we specialize to particular cases.

We start by proving a $\Omega\big(\sqrt{KT}\big)$ lower bound for any algorithm in our setting and for any combination of stochastic or adversarial action-state mappings and loss vectors. The construction is a reduction to the standard bandits lower bound construction.

Theorem 3.4. *Irrespective to whether the action-state mappings and loss vectors are stochastic or adversarial, there exists a sequence of losses such that any (possibly randomized) algorithm in BIO suffers regret* $\mathbb{E}\left[\mathrm{Reg}_T\right] = \Omega\big(\sqrt{KT}\big)$.

Proof. Our construction only uses two states h_1 and h_2. The loss vectors, which are deterministic and do not change over time, are defined as follows: $\ell_t(h_1) \coloneqq 1$ and $\ell_t(h_2) \coloneqq 0$ for all $t \geq 0$. The stochastic action-state mapping, which is also constant over time, is given by

$$s_t(a) \coloneqq \begin{cases} h_1 & \text{with probability } p_a \\ h_2 & \text{with probability } 1 - p_a \end{cases}$$

for all $a \in \mathcal{A}$ and $t \geq 0$, where the probabilities p_a are to be determined. Thus, the loss of an arm a is $\ell_t(s_t(a)) \coloneqq \ell_t(h_1) = 1$ with probability p_a and $\ell_t(s_t(a)) \coloneqq \ell_t(h_2) = 0$ with probability $1 - p_a$. Since the loss is determined by the state, the learner receives bandit feedback without delay. We can then choose p_a for $a \in \mathcal{A}$ to mimic the standard $\Omega\big(\sqrt{KT}\big)$ distribution-free bandit lower bound—e.g., see Slivkins (2019, Chapter 2). By Yao's minimax principle, the same lower bound also applies to

the case with adversarial action-state mappings. Since the loss vectors are deterministic, this covers all possible cases in BIO. □

Adversarial action-state mapping and stochastic losses. We first prove a lower bound of order $\sqrt{ST}$ for any number $K \geq 2$ of actions. However, we do need a minor generalization of our setting to allow correlation between unseen losses. Specifically, we allow all pairs of losses $\ell_j(s), \ell_{j'}(s')$ of distinct states $s \neq s'$ to be correlated if $j > j'$ and $j - j' \leq d$, while we guarantee the i.i.d. nature of losses for any fixed state. Since $\mathbb{E}[\ell_t(S_t)] = \mathbb{E}[\theta(S_t)]$, this does not affect the analysis for the upper bound on the regret of our algorithms since $\mathbb{E}[\mathrm{Reg}_T] \leq \mathbb{E}[\mathcal{R}eg_T]$ (see Lemma B.3). However, for a high-probability upper bound, we need to relate Reg_T and $\mathcal{R}eg_T$, which now leads to an additive $\widetilde{\mathcal{O}}(\sqrt{ST})$ term rather than an additive $\widetilde{\mathcal{O}}(\sqrt{T})$ term as in Equation (3.4).

In the proof of the $\sqrt{ST}$ lower bound, we leverage the fact that losses are independent only across time steps for a fixed state, while they may depend on the losses of the other states. Note that our lower bound holds even when the learner knows the action-state assignments beforehand. We provide a sketch of the proof of Theorem 3.5 below; see Appendix B.3 for the full proof.

Theorem 3.5. *Suppose that the action-state mapping is adversarial and the losses are stochastic and that $d_t = d$ for all $t \in [T]$. If $T \geq \min\{S, d\}$ then there exists a distribution of losses and a sequence of action-state mappings such that any (possibly randomized) algorithm suffers regret $\mathbb{E}[\mathrm{Reg}_T] = \Omega\big(\sqrt{\min\{S, d\}T}\big)$.*

Proof sketch. First, suppose that $S \leq 2d$. For the construction of the lower bound we only consider two actions and equally split the states over these two actions. Then, we divide the T time steps in blocks of length $S/2 \leq d$. In each block, each state has the same loss. Since the block length is smaller then the delay, we have effectively created a two-armed bandit problem with $T' = T/(S/2)$ rounds and loss range $[0, S/2]$, for which we can prove a $\Omega\big(S\sqrt{T'}\big) = \Omega\big(\sqrt{ST}\big)$ lower bound by showing an equivalent lower bound for the full information setting. If $S > 2d$, we use the same construction with only $2d$ states, and obtain a $\Omega\big(\sqrt{dT}\big)$ lower bound. □

Finally, we can show the following lower bound, whose proof can be found in Appendix B.3.

Theorem 3.6. *Suppose that the action-state mapping is adversarial, the losses are stochastic, and that $d_t = d$ for all $t \in [T]$. If $T \geq d + 1$ then there exists a distribution of losses and a sequence of action-state mappings such that any (possibly randomized) algorithm suffers regret*

$$\mathbb{E}[\mathrm{Reg}_T] = \Omega\Big(\min\Big\{(d+1)\sqrt{S}, \sqrt{(d+1)T}\Big\}\Big) .$$

This term is also present in the dynamic regret bound of `NSD-UCRL2`, but it is necessarily incurred from their analysis even in the stationary case (Vernade et al., 2020, Theorem 1).

This last lower bound implies that the regret of our algorithm is near-optimal. Since the lower bound of Theorem 3.4 applies to the case where the action-state mapping is adversarial and the losses are stochastic, we find the following result as a corollary of Theorem 3.4, Theorem 3.5, and Theorem 3.6.

Corollary 3.2. *Suppose that the action-state mapping is adversarial, the losses are stochastic, and that $d_t = d$ for all $t \in [T]$. If $T \geq 1 + \min\{S, d\}$, then there exists a distribution of losses and a*

sequence of action-state mappings such that any (possibly randomized) algorithm suffers regret

$$\mathbb{E}\left[\mathrm{Reg}_T\right] = \Omega\Big(\max\big\{\sqrt{KT}, \sqrt{\min\{S,d\}T}, (d+1)\sqrt{S}\big\}\Big)\,.$$

Stochastic action-state mappings and adversarial losses. In this case, we recover the standard lower bound for adversarial bandits with bounded delay.

Theorem 3.7. *Suppose that the action-state mapping is stochastic, the losses are adversarial, and that $d_t = d$ for all $t \in [T]$. Then there exists a stochastic action-state mapping and a sequence of losses such that any (possibly randomized) algorithm suffers regret $\mathbb{E}\left[\mathrm{Reg}_T\right] = \Omega\big(\max\big\{\sqrt{KT}, \sqrt{dT}\big\}\big)$.*

Proof. Since by Theorem 3.4 we already know that any algorithm must suffer $\Omega\big(\sqrt{KT}\big)$ regret, we only need to show a $\Omega(\sqrt{dT})$ lower bound. We use two states, h_1 and h_2. Our action-state mapping is deterministic and, for all $t \geq 0$, assigns $s_t(a) \coloneqq h_1$ to all but one action $a^\star$, to which the mapping assigns $s_t(a^\star) \coloneqq h_2$. We now have constructed a two-armed bandit problem with delayed feedback and T rounds, for which a $\Omega(\sqrt{dT})$ lower bound is known (Cesa-Bianchi et al., 2019). □

Adversarial action-state mappings, adversarial losses. Since we can recover the construction of the lower bound in Theorem 3.7, we immediately have the following result.

Corollary 3.3. *Suppose that the action-state mapping is adversarial, the losses are adversarial, and that $d_t = d$ for all $t \in [T]$. Then there exists an action-state mapping and a sequence of losses such that any (possibly randomized) algorithm suffers regret $\mathbb{E}\left[\mathrm{Reg}_T\right] = \Omega\big(\max\big\{\sqrt{KT}, \sqrt{dT}\big\}\big)$.*

3.6 Experiments

We empirically compare our algorithm `MetaBIO` with the following baselines: `DAda-Exp3` (György and Joulani, 2021) for adversarial delayed bandits without intermediate observations (which we used to instantiate the algorithm $\mathcal{B}$), the standard `UCB1` algorithm (Auer et al., 2002a) for stochastic bandits without delays and intermediate observations, and `NSD-UCRL2` (Vernade et al., 2020) for non-stationary stochastic action-state mappings and stochastic losses. We run all experiments with a time horizon of $T = 10^4$. All our plots show the cumulative regret of the algorithms considered as a function of time. The performance of each algorithm is averaged over 20 independent runs in every experiment, and the shaded areas consider a range centered around the mean with half-width corresponding to the empirical standard deviation of these 20 repetitions. In the first two experiments, we consider both fixed delays $d \in \{50, 100, 200\}$ and random delays $d_t \sim \mathrm{Laplace}(50, 25)$ sampled i.i.d. from the Laplace distribution with $\mathbb{E}\left[d_t\right] = 50$.

Experiment 1: stochastic action-state mappings. Here we use a stationary version of the experiments in Vernade et al. (2020)—see Table B.1 in Appendix B.4 for details. We set $K = 4$ and $S = 3$, while we repeat this experiment for the previously mentioned values of delays. Figure 3.2 shows that, across all delay regimes, `MetaBIO` largely improves on the performance of `DAda-Exp3` by exploiting intermediate observations.

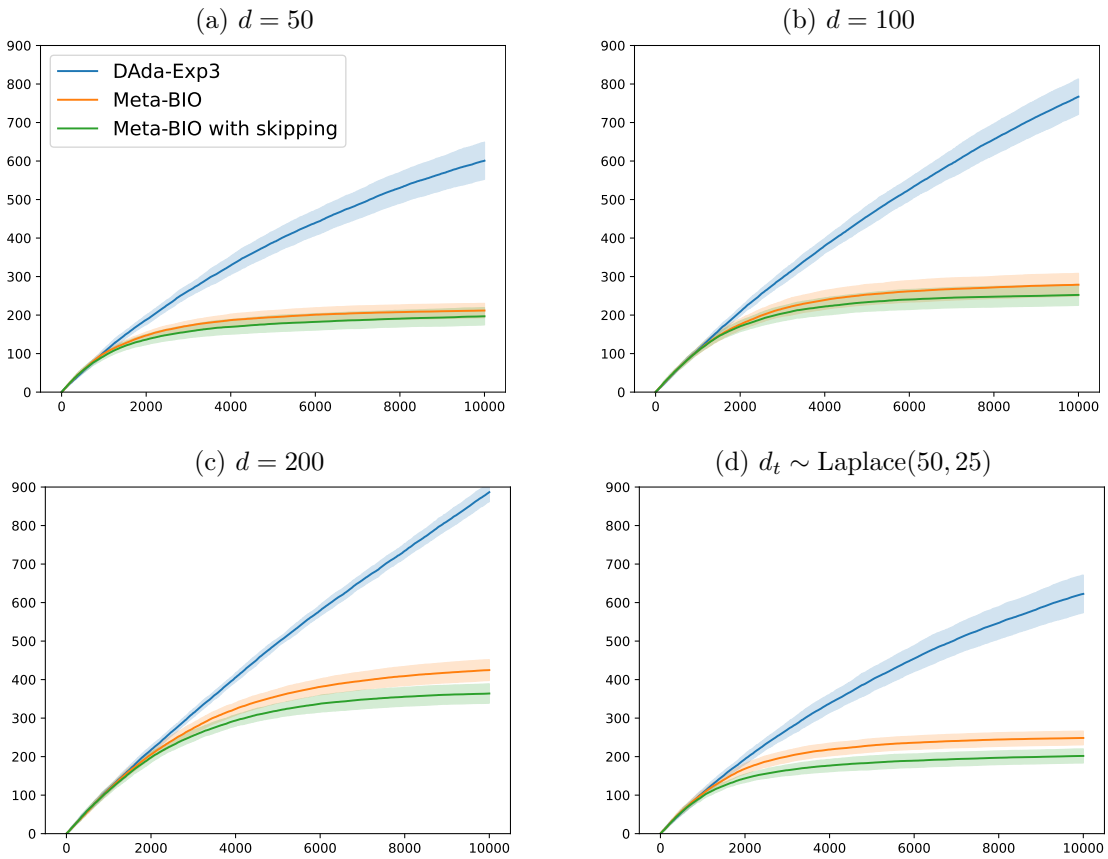


Figure 3.2: Cumulative regret over time for the stochastic action-state mapping when delays are fixed or random.

Experiment 2: adversarial action-state mappings. In this construction, we simulate the adversarial mapping using a construction adapted from Zimmert and Seldin (2021): we alternate between two stochastic mappings while keeping the loss means fixed. We set $K = 4$, $S = 3$, and we consider multiple instances for the different values of delays as in the previous experiment. The interval between two consecutive changes in the distribution of action-state mappings grows exponentially. See Table B.2 in Appendix B.4 for details. Figure 3.3 shows that `MetaBIO` and `MetaBIO` with "skipping" outperform both `UCB1` and `NSD-UCRL2`.

Experiment 3: utility of intermediate observations. Here we set $K = 8$, $d = 100$, and investigate how the performance of `MetaBIO` changes when the number S of states varies in $\{4, 6, 8, 10, 12\}$. The mean loss is always 0.2 for the optimal state and 1 for the others. The optimal action always maps to the optimal state. The suboptimal actions map to the optimal state with probability 0.6 and map to a random suboptimal state with probability 0.4. This implies that the expected loss of each arm remains constant when the number of states changes. Figure 3.4 shows that the regret gap between `MetaBIO` and `DAda-Exp3` shrinks as the number of states increases. This observation confirms our theoretical findings about the dependency of the regret on the number of states, which leads to a larger improvement the fewer they are.

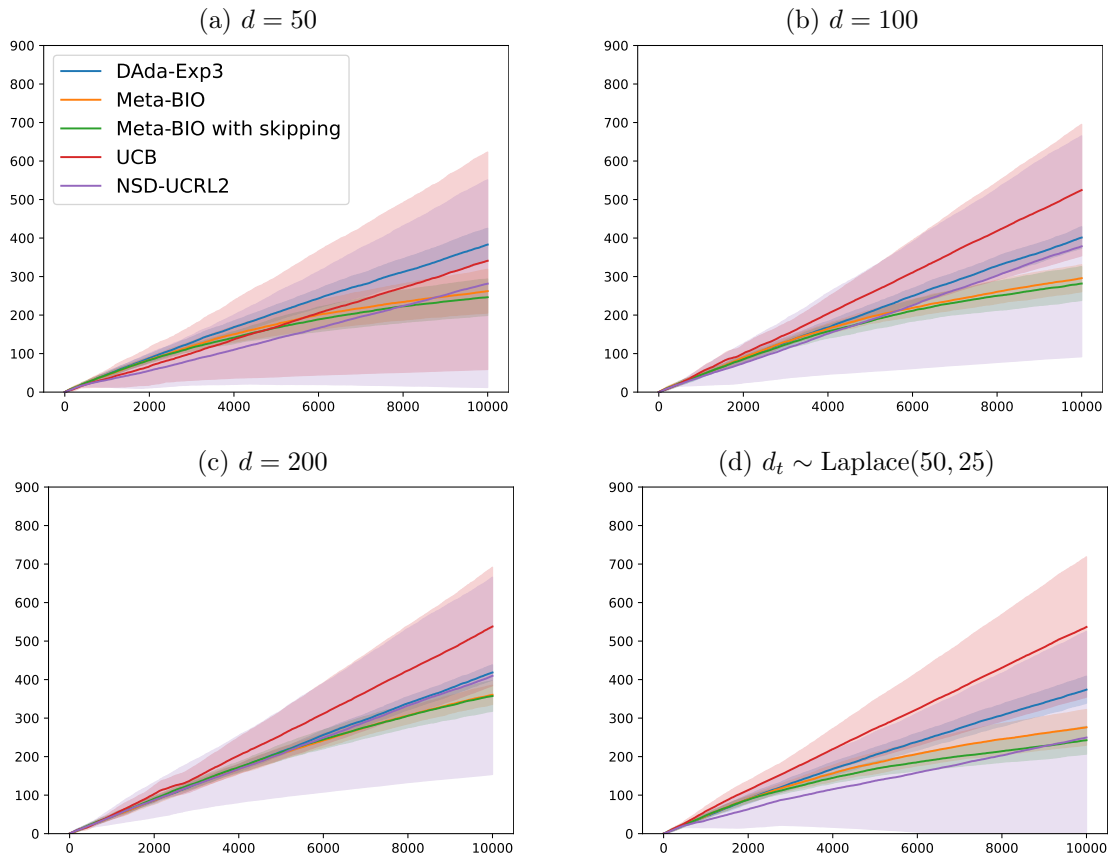


Figure 3.3: Cumulative regret over time for the adversarial action-state mapping when delays are fixed or random. All algorithms have small variance except for `UCB1` and `NSD-UCRL2`.

Experiment 4: performance of `AdaMetaBIO` when $S < d$. We use the same setting as in Experiment 1 with delay $d = 20$.[‡] Figure 3.6 shows the performance of `AdaMetaBIO` compared with both `DAda-Exp3` and `MetaBIO`. Before the switching point, `AdaMetaBIO` runs `DAda-Exp3` (up to independent internal randomization). Afterwards, `AdaMetaBIO` switches to `MetaBIO` (which in turn runs `DAda-Exp3` as a subroutine) and quickly aligns with its performance. Note that, at the switching time, `AdaMetaBIO` uses (via `MetaBIO`) the same instance of `DAda-Exp3` that was already running, rather than starting a new instance. It can be shown that our analysis of `AdaMetaBIO` applies to this variant as well without changes in the order of the bound.

Experiment 5: performance of `AdaMetaBIO` when $S > d$. We use a setting that is almost identical to that of Experiment 3, except we set $d = 4$ and $S = 14$. The performance of the three algorithms is shown in Figure 3.5. We can observe that `AdaMetaBIO` does not switch to `MetaBIO` and its performance is thus the same as that of `DAda-Exp3`, whereas `MetaBIO` incurs a larger regret.

[‡]Compared to the switching condition used for the analysis of `AdaMetaBIO`, we replace $49ST \ln \frac{8ST}{\delta}$ with ST. This change allows the switching condition to be triggered more easily to provide a better visualization of the behaviour of `AdaMetaBIO`, while it only introduces a polylog factor in its regret bound.

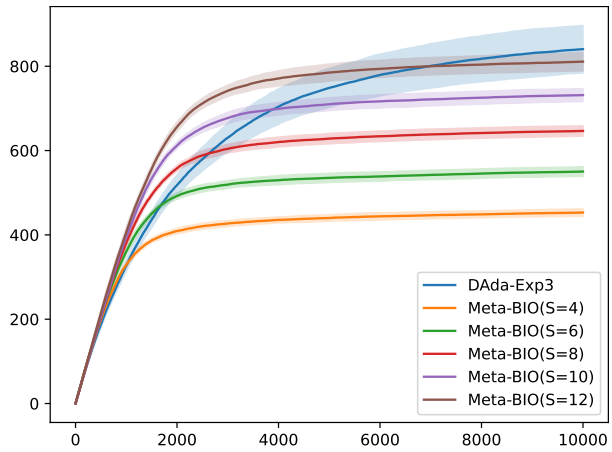


Figure 3.4: Cumulative regret over time of both `DAda-Exp3` and `MetaBIO` with different numbers of states $S \in \{4, 6, 8, 10, 12\}$.

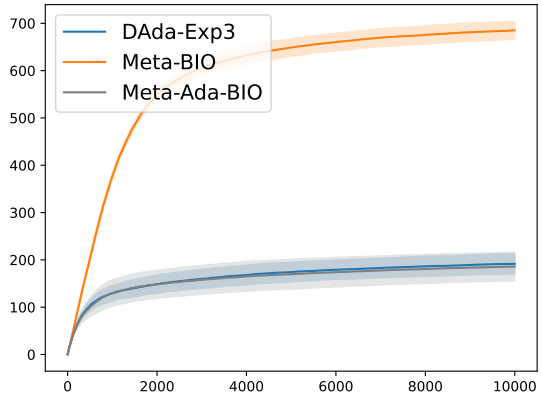


Figure 3.5: Cumulative regret over time of `DAda-Exp3`, `MetaBIO` and `AdaMetaBIO` when $S > d$.

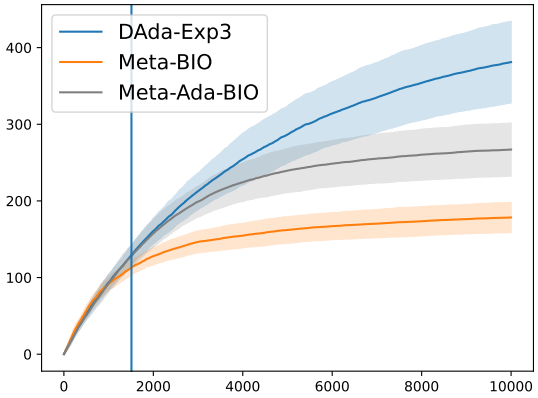


Figure 3.6: Cumulative regret over time of `DAda-Exp3`, `MetaBIO` and `AdaMetaBIO`. The vertical blue line marks the switching point of `AdaMetaBIO`.

Chapter 4

Distributed Online Convex Optimization under Stochastic Agent Availability and Random Networks

4.1 Introduction

In this chapter, we study a variant of distributed online convex optimization (DOCO) in which agents may not be available at every time step. When inactive, an agent neither contributes to the regret nor can it communicate with its neighbors. In practice, this scenario may arise due to machine failures, disconnections, or devices (e.g., mobile phones) being turned off. While the problem of intermittent agent availability has been investigated before in distributed convex optimization (Gu et al., 2021, Wang and Ji, 2022, Yan et al., 2024), we are not aware of any such study in the online framework. More specifically, we consider random agent activations where, in each round, each agent v becomes independently active with some unknown probability p_v. As a consequence, the active communication network at time t becomes stochastic, as it is induced by the random subset of active agents at time t. Because the number of active agents is a random variable, to ensure a uniform scaling of losses across time steps we define the global loss function as an average (as opposed to a sum) over the active agents. The individual regret of an agent u is then defined as the regret they accumulate with respect to this global loss function, but only during time steps in which they are active. On average, this corresponds to $p_u T$ time steps, implying that expected regret bounds should be compared to $\sqrt{p_u T}$, the typical regret rate in the absence of communication constraints. This notion of regret subsumes the standard notion of regret used in DOCO. We also introduce an alternative notion of regret, called network regret, which rather than relying on one single agent, is a cumulative average over active agents. The network regret accounts for $\left(1 - \prod_{v \in \mathcal{V}} (1 - p_v)\right) T$ time steps in expectation, which under some mild assumptions is close to T.

To propagate information about local losses, a possible approach is to use message passing. While it may seem adequate, it comes with significant drawbacks: in dense networks or when activation probabilities are high, the number of messages becomes prohibitively large; in sparse networks or with low activation probabilities, agents take too long to gather the necessary gradient information within a round. Instead, we use standard gossiping techniques (Xiao and Boyd, 2004b, Boyd et al., 2006) — widely used in DOCO — to aggregate gradient information from neighboring agents.

4.1.1 Related Works

To contextualize our contributions, we briefly review prior work on DOCO in time-varying communication networks. For space considerations, a more comprehensive literature review, including additional references, is deferred to Appendix C. DOCO in time-varying networks was first considered by Mateos-Núnez and Cortés (2014), who proved regret rates under the assumption that the union of communication networks over any m time steps is strongly connected. Akbari et al. (2015) considered unbalanced time-varying digraphs. Nedić and Olshevsky (2014) studied a time-varying sequence of directed graphs that is uniformly strongly connected. Hosseini et al. (2016) proposed a distributed algorithm that changes the weights on the communication links to adapt to the varying reliability of neighboring agents. They established the convergence rate of the algorithm as a function of the underlying network topology. It is worth mentioning that all aforementioned works are restricted to graphs that evolve deterministically. The work closest to ours is Lei et al. (2020), who studied the special case where communication networks are Erdös-Rényi graphs, in which each edge has a probability q of existing at each round. They proposed a gradient descent algorithm and proved regret upper bounds for the convex and strongly convex case, also extending their result to the bandit feedback framework. However, Lei et al. (2020) only consider stochastic edge availability. Our analysis of networks with random node availability provides a set of results that can be applied to both edge and node availability. In particular, we recover the bounds of Lei et al. (2020) for the full information setting as a special case of ours.

Main contributions. The main contributions of this chapter can be summarized as follows.

- We introduce two notions of regret relevant in the presence of random activations. The first, *individual regret*, is a natural extension of the standard regret used in DOCO, enabling direct comparison with state-of-the-art results. The second, *network regret*, is defined as a cumulative average over the active agents at each time step.
- We analyze `Gossip-FTRL`, a distributed variant of the FTRL algorithm for online convex optimization, and establish general upper bounds on both individual and network regret for arbitrary connected communication graphs $\mathcal{G}$ and arbitrary activation probabilities.
- In the p-uniform case (activation probabilities equal to some known $p \geq 1/N$), the expected individual regret of our algorithm is bounded by $\frac{1}{1-\rho} N^{1/4} p^{1/4} \sqrt{T}$, where T is the known time horizon, N is the known number of agents, and ρ is the unknown spectral gap of the expected gossip matrix supported on the active agents.
- For a standard choice of the gossip matrix, we show that $\frac{1}{1-\rho}$ is of order $\kappa(\mathcal{G})/p^2$, where $\kappa(\mathcal{G})$ is the condition number of $\mathcal{G}$ *. The maximal individual regret is bounded by $\mathcal{O}\left(\frac{\kappa(\mathcal{G})}{p^{3/4}} N^{1/4} \sqrt{T}\right)$. We also extend this result to non p-uniform cases and obtain $\mathcal{O}\left(I_p \frac{\kappa(\mathcal{G})}{p_{\min} \bar{p}^{1/4}} N^{1/4} \sqrt{\frac{T}{\mu}}\right)$ where $I_p = \bar{p}/p_{\min}$ is an imbalance factor.
- We provide a lower bound showing that any distributed online algorithm must suffer, on some $\mathcal{G}$, and for some activation probabilities, a network regret at least of order $\big(\kappa(\mathcal{G})\big)^{\delta/4} N^{1/2-\delta} \frac{1}{p_{\min}} \sqrt{T}$ for any $0 \leq \delta \leq 1/2$. This is the first lower bound for DOCO with random networks.

*defined more formally in Equation (4.7)

- We extend our upper bounds to the case where edges are randomly deleted after the selection of active agents. We thus establish a strict generalization of the results of Lei et al. (2020).

- To provide empirical support to our results, we run experiments on synthetic data comparing `Gossip-FTRL` with `DOGD` by Lei et al. (2020) for different choices of the relevant parameters.

Our most general bounds (Theorem 4.1) hold when agents only know $\{p_v\}_{v\in\mathcal{V}}$, N and T. In particular, agents need not know the structure of $\mathcal{G}$ and `Gossip-FTRL` is run with the same initialization for all agents. The more refined bounds in Corollary 4.2 and Corollary 4.3 also need preliminary knowledge of the spectral gap of $\mathcal{G}$ (or a suitable bound on it).

Technical challenges. The main technical challenge lies in controlling regret when the set of active agents varies randomly over time, unlike in standard DOCO. Partial participation impacts the rate of consensus and requires adapting the standard regret decomposition—into the regret of a virtual omniscient agent and the deviations from it—to handle random activation. Key hurdles in tightening the upper bounds are: (i) bounding deviations from the virtual agent so that the bound scales with the number of active agents $|S_t|$ rather than N; and (ii) bounding the virtual agent's regret in terms of the imbalance in activation probabilities, instead of N.

The lower bound also introduces novel technical difficulties: while previous arguments rely on deterministic feedback delays induced by the graph structure (e.g., bottlenecks), our setting requires accounting for stochastic feedback delays due to random agent activations, necessitating nontrivial probabilistic arguments in the analysis.

4.2 Problem Setting

In multi-agent online convex optimization, agents are nodes of a communication network represented by a connected and undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, N\} = [N]$ indexes the agents and the edge set $\mathcal{E}$ defines the communication structure among agents. We use $\mathcal{N}_v = \{u \in \mathcal{V} : (u, v) \in \mathcal{E}\}$ to denote the neighborhood of $v \in \mathcal{V}$. Let $\mathcal{X} \subset \mathbb{R}^n$ be the agents' common decision space, which we assume to be convex and closed. At every round $t = 1, 2, \dots$, each agent $v \in \mathcal{V}$ becomes independently active with fixed probability p_v. Without loss of generality, we assume $p_{\min} = \min_v p_v > 0$ and, for simplicity, $\sum_v p_v \geq 1$ (otherwise less than one active agent would be active per round on average). Note that this implies $p_{\max} = \max_v p_v \geq \frac{1}{N}$. We call *p-uniform* the special case when $p_v = p$ for all $v \in \mathcal{V}$. We assume that active agents v know which of their neighbors in $\mathcal{N}_v$ are active. Let S_t be the set of active agents at time t and $\mathcal{E}_t = \mathcal{E} \cap \{(u, v) : u, v \in S_t\}$ be the set of active edges at time t, i.e., edges in $\mathcal{E}$ whose both endpoints are active in that round. An adversary sets an unknown sequence $\ell_1(v, \cdot), \ell_2(v, \cdot), \dots$ of *local losses* $\ell_t(v, \cdot) : \mathcal{X} \to \mathbb{R}$ for each agent $v \in V$. The adversary is allowed to observe $(S_k)_{k<t}$ before setting $\ell_t(v, \cdot)$. For all $t \geq 1$ we assume $\ell_t(v, \cdot)$ is convex and L-Lipschitz with respect to an arbitrary norm $\|\cdot\|$.

We say that a doubly stochastic matrix W_t is a gossip matrix for the set S_t of active agents at time t if $W_t(v, v') = 0$ for all distinct $v, v' \in \mathcal{V}$ such that $(v, v') \notin \mathcal{E}_t$. Let W_t be a random gossip matrix (with respect to a graph $\mathcal{G}$ and activation probabilities p_v for $v \in \mathcal{V}$) and define $\rho = \sqrt{\lambda_2(\mathbb{E}[W_t W_t^T])}$, for i.i.d gossip matrices W_t, where we denote by $\lambda_i(\cdot)$ the i-th highest eigenvalue of a matrix (keeping

track of multiplicity). Clearly, $0 < \rho < 1$. In what follows, we often write ρ leaving $\mathcal{G}$ and $\{p_v\}_{v\in\mathcal{V}}$ implicitly understood from the context.

Next, we define the distributed online optimization protocol used in this chapter.
At each round $t = 1, 2, \ldots, T$,

1. Each active agent $v \in S_t$ chooses an action $x = x_t(v) \in \mathcal{X}$ and observes the gradient $\nabla\ell_t(v, x)$ of the local loss $\ell_t(v, \cdot)$.
2. Each active agent $v \in S_t$ sends a message $z_t(v)$ to their active neighbors and uses the messages received from the active neighbors to compute a new message $z_{t+1}(v)$.

Note that this protocol implicitly defines an active communication graph $\mathcal{G}_t = (S_t, \mathcal{E}_t)$ for round t. As the number of agents that incur loss in a step is a random variable, we define the *network loss* at step t as the average over the active agents of the local losses in that round,

$$\ell_t^{\text{net}}(S_t, \cdot) = \frac{1}{|S_t|}\sum_{v\in S_t} \ell_t(v, \cdot)$$

and let $\ell_t^{\text{net}}(\emptyset, \cdot) = 0$. Hence, unlike the standard DOCO model where ℓ_t^{net} scales linearly with N, in our model ℓ_t^{net} is independent of N.

The performance of each agent is evaluated using the *individual regret* $\text{Reg}_T(u)$ defined by

$$\text{Reg}_T(u) = \sum_{t\le T:\ u\in S_t} \ell_t^{\text{net}}\big(S_t, x_t(u)\big) - \min_{x\in\mathcal{X}} \sum_{t\le T:\ u\in S_t} \ell_t^{\text{net}}(S_t, x),$$

for all $u \in \mathcal{V}$. In this chapter, we provide uniform bounds in high probability and in expectation for the individual regret of any $u \in \mathcal{V}$.

Remark 4.1. *When $p_{\min} = 1$, we recover the standard DOCO setting Yan et al. (2013), Hosseini et al. (2013) and our regret $\text{Reg}_T(u)$ becomes*

$$\sum_{t=1}^{T} \ell_t^{\text{net}}\big(\mathcal{V}, x_t(u)\big) - \min_{x\in\mathcal{X}} \sum_{t=1}^{T} \ell_t^{\text{net}}(\mathcal{V}, x)\ . \tag{4.1}$$

In standard DOCO, ℓ_t^{net} is a sum over the N local losses. The resulting regret Reg'_T is then defined by

$$\max_{v\in\mathcal{V}} \sum_{t=1}^{T}\sum_{v'\in\mathcal{V}} \ell_t\big(v', x_t(v)\big) - \min_{x\in\mathcal{X}} \sum_{t=1}^{T}\sum_{v'\in\mathcal{V}} \ell_t(v', x)\ . \tag{4.2}$$

Hence, when $p_{\min} = 1$, $\max_{u\in\mathcal{V}} \text{Reg}_T(u) = \text{Reg}'_T/N$. Recently, Wan et al. (2024a) proved that $\text{Reg}'_T(u) = \widetilde{\Theta}\big(N(1-\rho)^{-1/4}\sqrt{T}\big)$ where the upper bound relies on accelerated gossiping and $\widetilde{\Theta}$ hides factors logarithmic in N. They also proved a lower bound of $\Omega\big(N(1-\rho)^{-1/4}\sqrt{T}\big)$ that consequently also holds up to a factor N for the individual regret when $p_{\min} = 1$.

Remark 4.2. *We sometimes refer to the network regret $\text{Reg}_T^{\text{net}}$, as opposed to the individual regret. It is defined as*

$$\sum_{t\le T} \frac{1}{|S_t|}\sum_{u\in\mathcal{V}} \ell_t^{\text{net}}\big(S_t, x_t(u)\big) - \min_{x\in\mathcal{X}} \sum_{t\le T} \frac{1}{|S_t|}\ell_t^{\text{net}}(S_t, x)$$

which should not be confused with the average of the individual regret across agents. In general, there is no clear ordering between the network regret and the maximal individual regret. For the p-uniform case, for example, when p is large, we expect the maximal individual regret to be larger, since it is

Algorithm 4.1: Gossip-FTRL An instance of this algorithm is run by each agent $v \in \mathcal{V}$.

Input: Learning rate $\eta > 0$
Initialize: $z_1(v) = 0$
for $t = 1, 2, \ldots$ **do**
  if $v \in S_t$ **then**
    Predict

$$x_t(v) = \arg\min_{x \in \mathcal{X}} \left\{ \langle z_t(v), x \rangle + \frac{1}{\eta} \psi(x) \right\}$$

    Observe $g_t(v) = \nabla \ell_t\big(v, x_t(v)\big)$
    Send $z_t(v)$ to $\mathcal{N}_v \cap S_t$
    Receive and store $z_t(j)$ from $j \in \mathcal{N}_v \cap S_t$
    Compute $W_t(v, j) > 0$ for $j \in \mathcal{N}_v \cap S_t$
    Update $z_{t+1}(v) = \sum_{j \in \mathcal{N}_v \cap S_t} W_t(v, j) z_t(j) + g_t(v)$
  else
    $z_{t+1}(v) = z_t(v)$

the case when $p = 1$, while for small values of p, we expect it to be smaller, since it only accounts for a fraction p of the time steps. In presenting our main results, we focus on the notion of individual regret to remain consistent with prior work. Nevertheless, we obtain counterparts of our upper bounds for the network regret.

Note that the network regret defined here bears similarities with that of Cesa-Bianchi et al. (2020), and even reduces to theirs if all the local losses are identical. However, their setting is not comparable with DOCO, precisely because losses in DOCO are agent-varying. In their setting, agents can achieve an expected network regret of order $\mathcal{O}(\sqrt{T})$ even without communicating. In DOCO, instead, ignoring communication leads to a linear expected network

4.3 The Gossip-FTRL Algorithm

We assume each agent runs an instance of Gossip-FTRL (Algorithm 4.1), a gossiping variant of FTRL with a regularizer $\psi : \mathcal{X} \mapsto \mathbb{R}$ that is μ-strongly convex with respect to the same norm $\|\cdot\|$ with respect to which the Lipschitzness of the losses is defined. Our analysis depends on the choice of ψ only through μ and the diameter $D^2 = \max_{x \in \mathcal{X}} \psi(x) - \min_{x' \in \mathcal{X}} \psi(x')$. At any time step t, the instance of Gossip-FTRL run by an active agent v computes a weight vector $W_t(v, \cdot)$ over the set $\mathcal{N}_v \cap S_t$ of active neighbors. In the section about the gossip matrix, we introduce a simple way of choosing these weights so that the *gossip matrix* $W_t(\cdot, \cdot)$ is an i.i.d. doubly stochastic matrix, which is a requirement for our analysis.

The algorithm considered here is a natural extension to arbitrary regularizers and random activations of those traditionally used for DOCO, e.g., (Hosseini et al., 2013), but the optimal choice for the learning rate is different, as it depends on the activation probabilities.

4.4 Upper Bounds

Recall that at each round, each agent $v \in \mathcal{V}$ is independently active with probability p_v. Let $\overline{p}$ be the average of these probabilities. The next result establishes an upper bound on the expected

individual regret of Algorithm 4.1 (all missing proofs are in Appendix C).

Theorem 4.1. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta > 0$ and i.i.d gossip matrices W_t. Then, the expected individual regret for each $u \in \mathcal{V}$ can be bounded by*

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq \frac{D^2 I_p}{\eta\bar{p}} + \frac{L^2\eta}{\mu}(6 + 2I_p + 3p_u\sqrt{\bar{p}N}\frac{\rho}{1-\rho})T\,, \tag{4.3}$$

where $\rho = \sqrt{\lambda_2(\mathbb{E}[W_1 W_1^\top])}$, and $I_p = \bar{p}/p_{\min}$ is an imbalance factor. In the p-uniform case, we have, for all $u \in \mathcal{V}$,

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq \frac{D^2}{p\eta} + \frac{L^2}{\mu}\eta(8 + 3p\sqrt{pN}\frac{\rho}{1-\rho})T\ . \tag{4.4}$$

If, in addition, $\eta = \frac{(D/L)\sqrt{\mu}}{2\sqrt{2}N^{1/4}\sqrt{T}p^{5/4}}$, then for any $u \in \mathcal{V}$,

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq 2\sqrt{2}\frac{DL}{\sqrt{\mu}}N^{1/4}p^{1/4}\frac{1}{1-\rho}\sqrt{T}\,. \tag{4.5}$$

Note that, in the p-uniform case, even lacking any knowledge on p, except for $pN \geq 1$, one can set $\eta = (D/L)N^{1/4}\sqrt{\mu/T}$ and get the suboptimal bound $\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq 8DLN^{\frac{3}{4}}\frac{1}{1-\rho}\sqrt{T/\mu}$.

In the first bound of Theorem 4.1, the presence of the imbalance factor shows that the regret grows with the heterogeneity of the activation probabilities p_v.

The bounds of Theorem 4.1 capture the structure of $\mathcal{G}$ through the reciprocal of the spectral gap $\frac{1}{1-\rho}$. In the section on the gossip matrix, we give upper bounds on $\frac{1}{1-\rho}$ for an appropriately chosen gossip matrix W_t. Specifically, combining Equation (4.3) with Theorem 4.3 and choosing an appropriate η which only requires knowing $p_{\min}$, $\bar{p}$, N, and the spectral radius of $\mathcal{G}$ we get,

$$\max_{u\in\mathcal{V}}\mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(DLI_p\frac{\kappa(\mathcal{G})}{p_{\min}\bar{p}^{1/4}}N^{1/4}\sqrt{\frac{T}{\mu}}\right)\ , \tag{4.6}$$

as proved in Corollary 4.2, where $\kappa(\mathcal{G})$ is the condition number of $\mathrm{Lap}(\mathcal{G})$, i.e.

$$\kappa(\mathcal{G}) = \lambda_1(\mathrm{Lap}(\mathcal{G}))/\lambda_{N-1}(\mathrm{Lap}(\mathcal{G})) \tag{4.7}$$

(note that $\lambda_N(\mathrm{Lap}(\mathcal{G})) = 0$) and represents a notion of connectivity of $\mathcal{G}$ (a smaller value of $\kappa(\mathcal{G})$ corresponds to better-connected graphs). Together with the assumption that $\sum_v p_v \geq 1$, this also yields $\max_{u\in\mathcal{V}}\mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(DLI_p\frac{\kappa(\mathcal{G})}{p_{\min}}N^{1/2}\sqrt{\frac{T}{\mu}}\right)$. This bound holds in particular when $p_{\max} = 1$. When $p_v = p$, Bound Equation (4.6) is replaced by $\mathcal{O}\left(DL\frac{\kappa(\mathcal{G})}{p^{3/4}}N^{1/4}\sqrt{T/\mu}\right)$, as proved in Corollary 4.2. This makes us lose a factor $p^{5/4}$ with respect with the bound $\mathcal{O}\left(\sqrt{pT}\right)$ for the case when all losses are equal across agents, which cancels the need for communication.

Comparison with previous bounds.

To compare with previous results, we restrict our analysis to the special case when $p_v = 1$ for all $v \in \mathcal{V}$. In this case, our bound Equation (4.5) is of order of $N^{1/4}\frac{1}{1-\rho}\sqrt{T}$. This matches the upper bounds of Hosseini et al. (2013), Yan et al. (2013)—recall that our global loss is divided by the number of active agents, so the upper bounds for the standard DOCO setting must be divided by N.

If the active graph $\mathcal{G}_t$ is an Erdős-Rényi random graph with parameter q, our setting reduces to

that of Lei et al. (2020) for convex losses and full feedback. As shown later in the section on random edges, our analysis recovers the upper bound of order $\frac{N^{1/4}}{q}\frac{\rho}{1-\rho}\sqrt{T}$ in (Lei et al., 2020, Theorem 1) when η is tuned based on N. In Corollary 4.3, we also prove a general bound that holds for all p and q and where ρ is expressed in terms of simple graph-theoretic quantities. When $p = 1$ and $q = 1$, Wan et al. (2024a) recently showed that using accelerated gossip one can achieve a bound of order $(1-\rho)^{-1/4}\ln N\sqrt{T}$ when η is tuned based on both N and ρ. Under the same tuning assumptions, our bound Equation (4.5) is instead of order $N^{1/4}\sqrt{\frac{\rho}{1-\rho}T}$.

Network regret. The network regret may be bounded as $\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{D^2 I_p}{\eta\bar{p}} + \frac{L^2\eta}{\mu}(6 + 2I_p + 3\sqrt{\bar{p}N}\frac{\rho}{1-\rho})T$ (Theorem C.1 in Appendix C). When $p_{\max} = 1$, this matches the bound of Theorem 4.1, indicating that the difference between the bounds primarily arises from the individual regret being accumulated over fewer time steps. In the p-uniform case, the network regret becomes $\mathcal{O}(DLN^{1/4}p^{-5/4}\sqrt{T/\sqrt{\mu}})$ with the right η.

Trade-off between regret and communication. While the activation probabilities are exogenous variables that cannot be tuned, it is interesting to compare their effect on regret and communication costs (measured as the expected number of simultaneously active pairs of agents in each round). In particular, while the maximal individual regret scales with $\kappa(\mathcal{G})/p^{3/4}$ in the p-uniform case, the communication cost scales with $p^2|\mathcal{E}|$.

High probability upper bounds. In the Appendix, we complement these results with high probability guarantees showing that the individual regret is upper bounded by $\mathcal{O}\left(DL\sqrt{NT}/\sqrt{1-\rho^2}\right)\ln(NT/\delta)$ with probability larger than $1-\delta$. The degradation of the bound with respect to the dependence on N and p is natural, since the analysis must involve high-probability upper and lower bounds on $|S_t|$, which yields a dependence on N instead of p.

4.5 Lower Bound

We provide a generalization of the lower bound from Wan et al. (2024a) to the case of random agent availability.

Theorem 4.2. *Let A be any algorithm for D-OCO on the decision set $\mathcal{X}$. Let $N = 2(M+1)$, where $M \geq 4$ is an even integer, and suppose $T \geq N^3$. Then, there exists a graph $\mathcal{G}$ with N nodes and a set of activation probabilities $\{p_v \mid v \in [N]\}$ with $p_{\min}N \geq 1$, and sequences of linear functions $\{\ell_t(v,\cdot)\}_{t=1}^T$, with each $\ell_t(v,\cdot)$ chosen adaptively based on $(S_k)_{k\leq t}$, and satisfying $\|\nabla\ell_t(v,\cdot)\|_2 \leq L$, such that the expected individual regret of A satisfies:*

$$\max_{u\in[N]} \mathbb{E}[\mathrm{Reg}_T(u)] \geq \frac{1}{2^7 p_{\min}} DL\,\kappa(\mathcal{G})^{\delta/2} N^{1/2-\delta}\sqrt{T}$$

for all $0 \leq \delta \leq \frac{1}{2}$, while the imbalance factor satisfies $I_p = \frac{2+p_{\min}(N-2)}{Np_{\min}} \leq 3$.

This result justifies the $1/p_{\min}$ scaling in the upper bound for the non-uniform case Equation (4.6). Notably, this dependence is not due to the imbalance factor, which remains below 3 in this setting. It shows that the dependence on $\{p_v\}_{v\in\mathcal{V}}$ obtained in Equation (4.6) is only suboptimal by a factor of $1/\bar{p}^{1/4}$. The dependence on $\kappa(\mathcal{G})$ and N is inherited from the lower bound established for standard DOCO in Wan et al. (2024a). The full proofs are deferred to Appendix C.5.

4.6 The Gossip Matrix

Following the literature on gossip algorithms, we set

$$W_t = I_N - b\,\mathrm{Lap}(\mathcal{G}_t)\,, \tag{4.8}$$

where $\mathcal{G}_t = (V, \mathcal{E}_t)$ and $b > 0$ is a parameter set so that $b \leq 1/\lambda_1(\mathcal{G}_t)$. One can easily verify that this a gossip matrix for S_t. Indeed, it is a symmetric and doubly-stochastic matrix, whose off-diagonal elements $W_t(i,j)$ are zero whenever either i or j are not in S_t. In particular, W_t is nonnegative because $\lambda_1(\mathcal{G}_t) \leq \lambda_1(\mathcal{G})$, since $\mathcal{G}_t$ is obtained by removing edges from $\mathcal{G}$. Note also that by knowing b and its active neighborhood, an agent v can compute $W_t(v,\cdot)$, as required by `Gossip-FTRL`. Finally, since the sets $S_1, S_2, \ldots$ of active agents are drawn i.i.d., the matrices $W_1, W_2, \ldots$ are also i.i.d.

A general bound on ρ. The following result provides a general upper bound that, when applied to the bounds of Theorem 4.1, characterizes the dependence of the regret both on the probabilities p_v and on the graph structure (through the Fiedler value $\lambda_{N-1}(\mathcal{G})$ or the condition number $\kappa(\mathcal{G})$). The full proofs in this section are deferred to Appendix C.6.

Theorem 4.3. *If W_t is set according to Equation (4.8), then*

$$\rho^2 \leq 1 - b p_{\min}^2 \lambda_{N-1}(\mathcal{G})\ . \tag{4.9}$$

Moreover, for $b = 1/\lambda_1(\mathcal{G})$ we have

$$\rho^2 \leq 1 - \frac{p_{\min}^2}{\kappa(\mathcal{G})}\ . \tag{4.10}$$

This choice of b, which is the best possible under the constraint that the gossip matrix is nonnegative, reveals that the regret is naturally controlled by the condition number of $\mathcal{G}$.

In the p-uniform case, we can refine this and obtain a closed-form expression for ρ.

Theorem 4.4. *If W_t is set according to Equation (4.8), and $b = 1/\lambda_1(\mathcal{G})$, then in the p-uniform case we have*

$$\rho^2 = 1 - \frac{2p^2}{\kappa(\mathcal{G})}\left(1 - \frac{1-p}{\lambda_1(\mathcal{G})} - \frac{p}{2\kappa(\mathcal{G})}\right)\ .$$

Wether in the p-uniform case or not, we can always derive the following bound on the factor $1/(1-\rho)$, which appears in Theorem 4.1.

Corollary 4.1. *If W_t is set according to Equation (4.8) and $b = 1/\lambda_1(\mathcal{G})$, then*

$$\frac{1}{1-\rho} \leq 2\frac{\kappa(\mathcal{G})}{p_{\min}^2}\,.$$

Combining this with Equation (4.4), we immediately get the following result.

Corollary 4.2. *Assume each agent runs an instance of `Gossip-FTRL` with learning rate $\eta > 0$. If the gossip matrix W_t is chosen as in Equation (4.8) with $b = 1/\lambda_1(\mathcal{G})$ and η tuned with respect to $\{p_v\}_{v\in\mathcal{V}}$ and N, the expected individual regret can be bounded by*

$$\max_{u\in\mathcal{V}} \mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(DLI_p \frac{\kappa(\mathcal{G})}{p_{\min}\bar{p}^{1/4}} N^{1/4}\sqrt{T/\mu}\right)\ , \tag{4.11}$$

in the general case and

$$\max_{u\in\mathcal{V}} \mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(DL\frac{\kappa(\mathcal{G})}{p^{3/4}}N^{1/4}\sqrt{T/\mu}\right) \tag{4.12}$$

in the p-uniform case, for all $p \leq 1$.

The bounds in Equation (4.11) and Equation (4.12) capture the intuition that bottlenecks in $\mathcal{G}$ (causing a small Fiedler value or a high condition number) negatively impact the regret due to a slower propagation of the information in the network.

Note that the appropriate choice of η only requires knowing $p_{\min}$, $\bar{p}$ N, and the spectral radius of $\mathcal{G}$

Spectral properties of the gossip matrix in the p-uniform case. To better visualize the dependence on the graph structure, we study specific graphs of practical importance. Specifically, we give results in the p-uniform case for cliques, strongly regular graphs, and grids (see Figure 4.1).

Clique. We have $\lambda_1(\mathcal{G}) = \lambda_{N-1}(\mathcal{G}) = N$ and

$$\rho^2 = 1 - 2p^2 + p^2\left(\frac{2(1-p)}{N} + p\right) .$$

Strongly regular. Let $\mathcal{G}$ be strongly regular with parameters k (the degree of any node), m (the number of common neighbors for any two adjacent nodes), and n (the number of common neighbors for any two nonadjacent nodes). Then $\lambda_1(\mathcal{G}) = k - s$, $\lambda_{N-1}(\mathcal{G}) = k - r$, and

$$\rho^2 \leq 1 - \frac{k-r}{k-s}p^2 ,$$

where $\begin{cases} r = \frac{m-n+\sqrt{(m-n)^2+4(k-n)}}{2} , \\ s = \frac{m-n-\sqrt{(m-n)^2+4(k-n)}}{2} . \end{cases}$

In particular, when $\mathcal{G}$ is the lattice graph, i.e. the graph with vertices $[M]^2$ and an edge between any two vertices in the same rows or columns (yielding $k = 2M-2$, $m = M-2$ and $n = 2$), $\rho \leq 1 - \frac{1}{2}p^2$.

2-dim grid. We have $\lambda_1(\mathcal{G}) = 4 - 4\cos(\pi(M-1)/M)$ and $\lambda_{N-1}(\mathcal{G}) = 2 - 2\cos(\pi/M)$, where $M = \sqrt{N}$ is the grid side length. Then

$$\rho^2 \leq 1 - \frac{1-\cos(\pi/M)}{2-2\cos(\pi(M-1)/M)}p^2 .$$

Note that $\frac{1-\cos(\pi/M)}{2-2\cos(\pi(M-1)/M)} \sim \frac{\pi^2}{4M^2}$, which goes to zero when $M \to \infty$.

Figure 4.1 shows the empirical behavior of ρ^2 for $b = 1/\lambda_1(\mathcal{G})$. The quantity $\widehat{\rho}^2$ is the second eigenvalue of W_1^2 averaged over 1000 different draws of active agents, where each agent is activated with probability p ranging from 0 to 1. We also plot the exact value ρ^2 (Theorem 4.4) and its upper bound ρ^2_{up} Equation (4.10).

Figure 4.1 reveals that for dense graphs, ρ^2 decreases quickly as $p \to 1$, implying a better regret rate. For the clique we have $\rho = 0$, implying an expected regret rate of order $\sqrt{T}$, which is independent of N—see Equation (4.4). On the other hand, in sparse graphs ρ may remain high. For

example, in the grid $\rho > 0.9$ for all p. Note also that ρ_{up}^2 approximates ρ^2 well, especially when p is small.

4.7 Random Edges

We now study a setting where, after agents are activated, edges between pairs of active agents are independently deleted with probability $1 - q$. More specifically, given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the active graph $\mathcal{G}_t = (S_t, \mathcal{E}_t)$ at time t is defined by $\Pr\big((i,j) \in \mathcal{E}_t\big) = q \Pr(i,j \in S_t)\mathbb{I}\big\{(i,j) \in \mathcal{E}\big\}$. When $\Pr(i,j \in S_t) = 1$ for all distinct $i, j \in \mathcal{V}$ we recover the model of Lei et al. (2020). To simplify, we only consider the p-uniform case. In that case, we write $\mathcal{G}_t \sim \mathcal{G}(\mathcal{G}, p, q)$. Note that $\mathcal{G}_1, \mathcal{G}_2, \ldots$ is i.i.d. because $S_1, S_2, \ldots$ is i.i.d.; moreover, if W_t is chosen as in Equation (4.8), then $W_1, W_2, \ldots$ is also an i.i.d. sequence. Using Equation (4.4), we can prove the following result. The full proofs are deferred to Appendix C.7.

Corollary 4.3. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta > 0$. If the gossip matrix W_t is chosen as in Equation* (4.8) *with $b = 1/\lambda_1(\mathcal{G})$, then*

$$\rho^2 = 1 - \frac{2p^2 q}{\kappa(\mathcal{G})}\left(1 - \frac{1-pq}{\lambda_1(\mathcal{G})} - \frac{pq}{2\kappa(\mathcal{G})}\right) .$$

By tuning η with respect to p and N,the expected individual regret of each $u \in \mathcal{V}$ on $\mathcal{G}_1, \mathcal{G}_2, \ldots$ drawn i.i.d. from $\mathcal{G}(\mathcal{G}, p, q)$ can be bounded by

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(\frac{\kappa(\mathcal{G})}{q}\frac{N^{1/4}}{p^{3/4}}\sqrt{T}\right) . \tag{4.13}$$

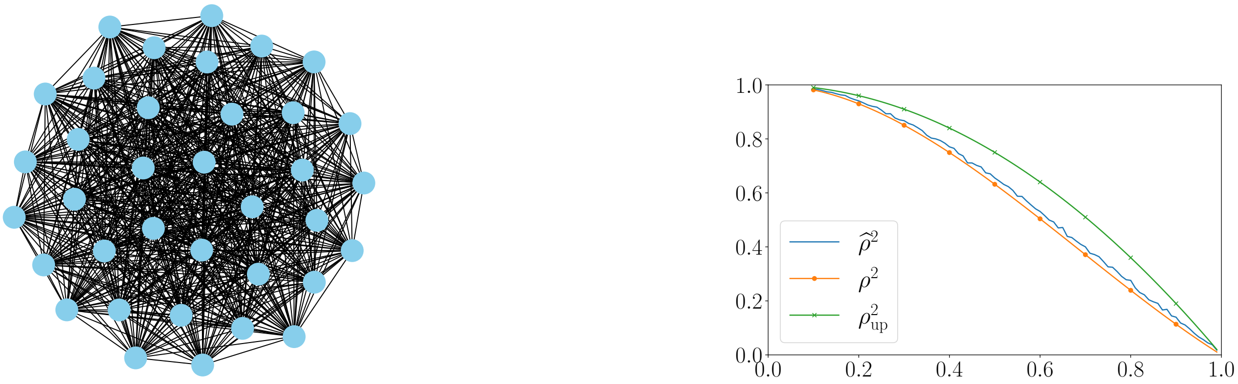
4.8 Experiments

We empirically evaluate our theoretical results on synthetic data. As the closest setting to ours is the one considered by Lei et al. (2020), we compare `Gossip-FTRL` with `DOGD`. Recall that while `Gossip-FTRL` can deal with arbitrary values of p and q in $(0, 1]$, `DOGD` is designed for settings with $p = 1$ (agents are always active) and $0 < q \leq 1$ (edges of $\mathcal{G}$ are active with probability q). To run `DOGD` when $p < 1$, we feed a zero gradient vector to instances run by agents that are inactive in that round. Our synthetic data are generated based on the distributed linear regression setting of Yuan et al. (2020). In particular, the agents' decision space $\mathcal{X}$ is the 10-dimensional Euclidean ball of radius 2 centered at the origin. The local loss functions are $\ell_t(v, \boldsymbol{x}) = \frac{1}{2}\big(\langle \boldsymbol{w}_t(v), \boldsymbol{x}\rangle - y_t(v)\big)^2$ for all $v \in \mathcal{V}$ and $\boldsymbol{x} \in \mathcal{X}$. The feature vectors $\boldsymbol{w}_t(v)$ are generated independently, by picking each coordinate independently and uniformly at random in $[-1, 1]$. The labels $y_t(v)$ are generated according to $y_t(v) = \varepsilon_t(v)$ for $1 \leq v < \lceil N/2 \rceil$ and $y_t(v) = \langle \boldsymbol{w}_t(v), \mathbf{1}\rangle + \varepsilon_t(v)$ for the remaining agents, where $\varepsilon_t(v)$ is independent Gaussian noise (zero mean and unit variance). As it corresponds to roughly T rounds (as opposed to pT) and offers greater practical relevance, we opt to report the network regret rather than the individual regret.

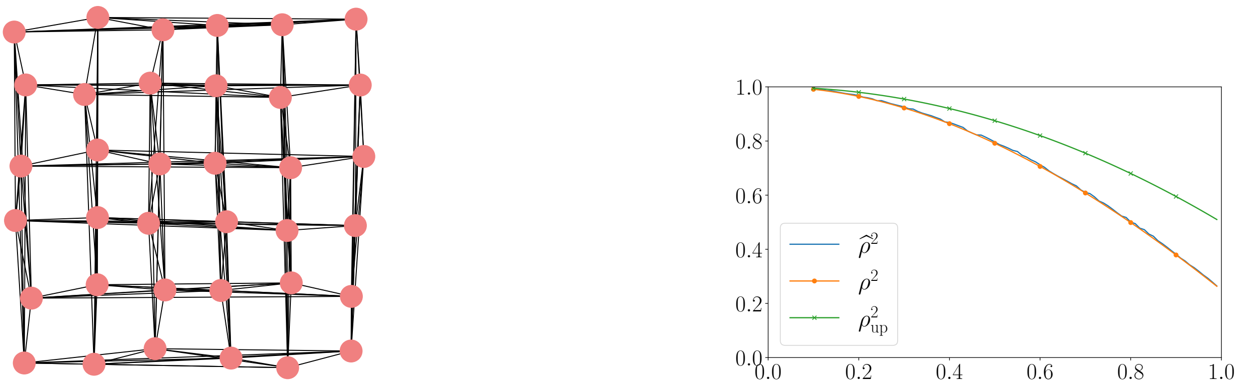
Each of the following experiments is run with $|\mathcal{V}| = N = 36$ and $T = 1000$. Plots are averages over 20 repetitions, where repetitions use the same labels and feature vectors and only agent (and possibly edge) activations are drawn afresh in each repetition. Both algorithms are tuned according to the

theoretical specifications (ignoring constant factors): we set $\eta = p^{-3/4}N^{-1/4}T^{-1/2}$ for `Gossip-FTRL` (see the optimal setting of η for the network regret in Theorem C.1) and $\eta = N^{-1/4}T^{-1/2}$ for `DOGD`.

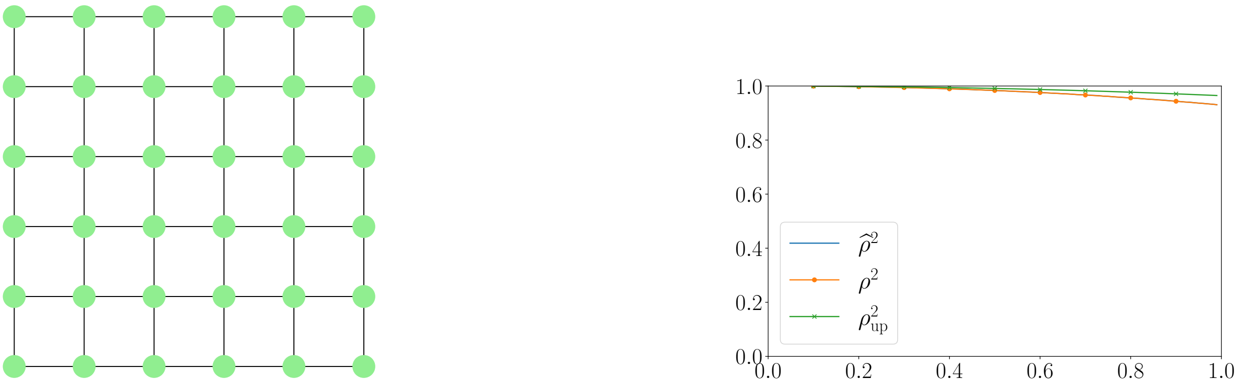
Our experiments show that `Gossip-FTRL` performs consistently better than `DOGD`, although the difference is not huge. Both algorithms are surprisingly robust to sparsity induced by low values of q when $\mathcal{G}$ is dense (Figure 4.2a and Figure 4.2c). When $\mathcal{G}$ is sparse, the regret goes up much more quickly as q becomes smaller (Figure 4.2b and Figure 4.2d). Figure 4.3 shows the behavior of `Gossip-FTRL` and `DOGD` on a grid for pairs (p, q) in the set $\{0.4, 0.6, 0.8\}^2$. In Figure 4.4 the plot of the network regret for `Gossip-FTRL` is consistent with $\mathrm{Reg}_T^{\mathrm{net}}$ scaling with $\frac{1}{p^{9/4}}$ as predicted by the third bound of Theorem C.1 Equation (C.7)—at least for sufficiently small values of p—and `DOGD` exhibits a similar behavior. Finally, Figure 4.5 shows the impact of $\lambda_{N-1}(\mathcal{G})$ on the network regret of `Gossip-FTRL`. The regret decreases as $\lambda_{N-1}(\mathcal{G})$ is increased by adding more edges to the bottleneck between the two cliques.



(a) 36-node clique

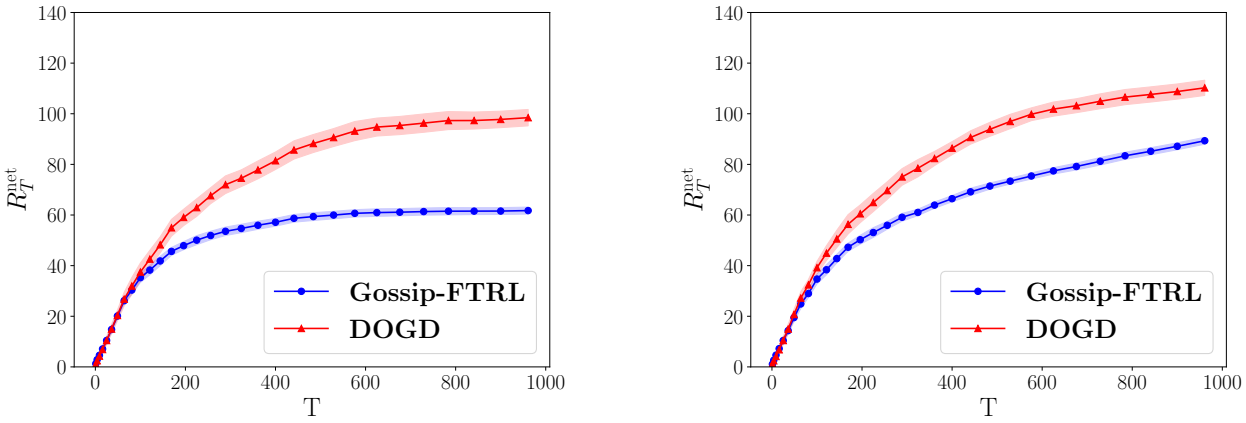


(b) 6×6 lattice



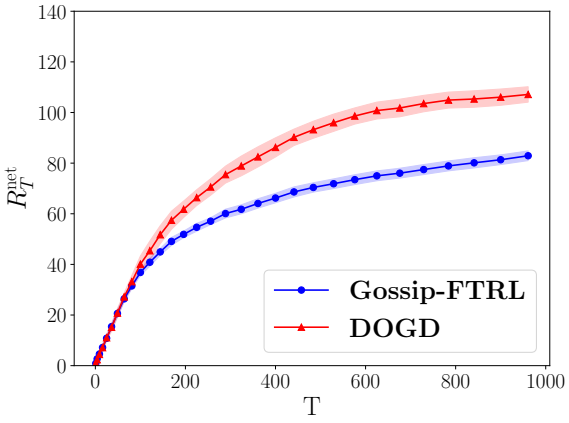
(c) 6×6 grid

Figure 4.1: Empirical estimate $\widehat{\rho}$ compared to ρ and the upper bound $\widehat{\rho}_{\text{up}}$ Equation (4.10) for $b = 1/\lambda_1(\mathcal{G})$ plotted as a function of $p \in [0, 1]$.

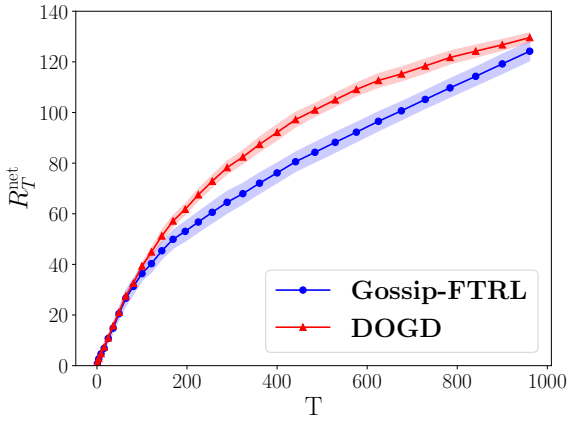


(a) Clique $p = 0.5, q = 1$.

(b) Grid $p = 0.5, q = 1$.



(c) Clique $p = 0.5, q = 0.05$.



(d) Grid $p = 0.5, q = 0.5$.

Figure 4.2: Growth over $T = 1000$ steps of the network regret of `Gossip-FTRL` and `DOGD` on a clique and on a grid for $N = 36$ and for different choices of p, q.

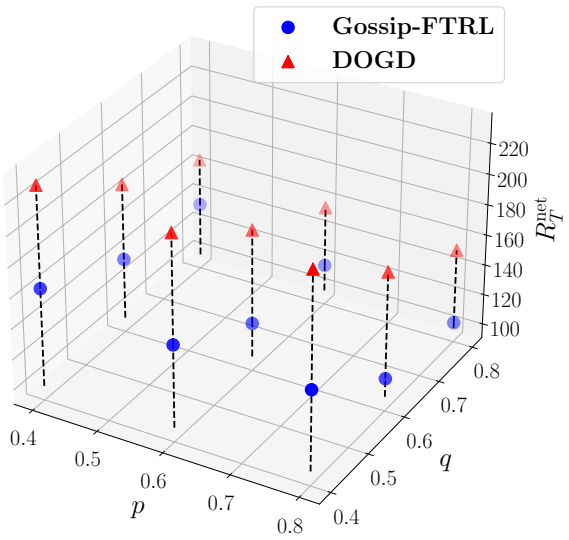


Figure 4.3: Network regret of `Gossip-FTRL` and `DOGD` after $T = 1000$ steps on a grid with $N = 64$.

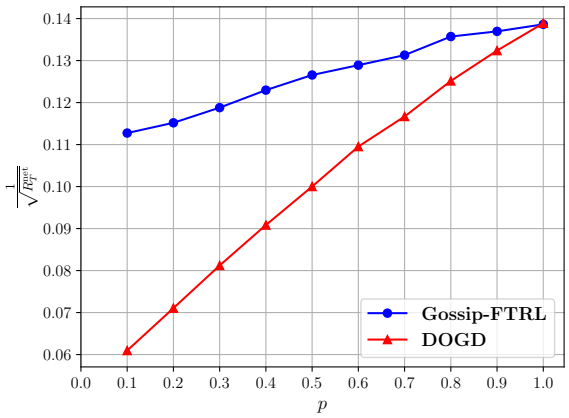


Figure 4.4: Plot of $\left(\text{Reg}_T^{\text{net}}\right)^{-1/2}$ for `Gossip-FTRL` on a clique for $p \in [0, 1]$ and $T = 1000$.

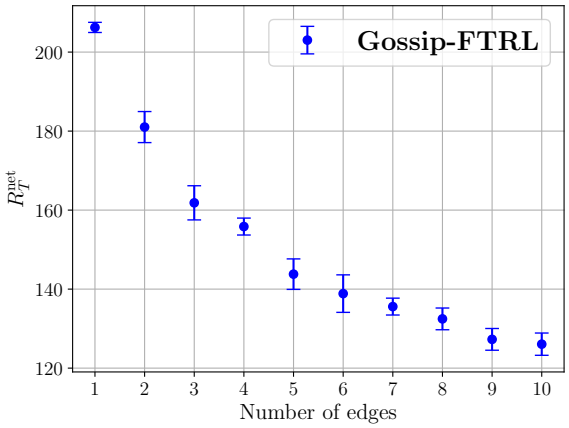


Figure 4.5: Network regret of `Gossip-FTRL` after $T = 1000$ steps when $p = 0.5$, $q = 1$, and $\mathcal{G}$ is made up by two cliques joined by a varying number of random edges.

Chapter 5

Distributed Online Convex Optimization with Delayed Feedback

5.1 Introduction

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 (see in Chapter 2), the interplay between decentralization and delayed feedback introduces unique complexities that remain less understood. Several works have considered delays in distributed settings, but most assume either bounded time-invariant (Cao and Basar, 2022) or known delays (Nguyen et al., 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 et al. (2024) made progress by proposing a distributed algorithm that handles arbitrary delays in DOCO. 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 distributed online learning algorithms that adapt to unknown, time- and agent-varying delays while maintaining near-optimal regret guarantees?

In this chapter, we answer this question affirmatively by developing novel distributed 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 $\widetilde{\mathcal{O}}\big(\sqrt{N^3 d_{\text{tot}}} + \frac{N\sqrt{N}\sqrt{T}}{(1-\sigma_2)^{1/4}}\big)$, 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.*

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

Our algorithm is inspired by the recent advance in DOCO (Wan et al., 2024b) but with an important adaptive learning rate mechanism combined with a distributed 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 et al. (2024) whose regret bound is no better than $\mathcal{O}\left(\frac{N^2}{(1-\sigma_2)^2}\sqrt{d_{\text{tot}}} + \frac{N\sqrt{N}}{1-\sigma_2}\sqrt{T}\right)$, 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.[†] 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 T and $1-\sigma_2$ is tight.

- We then consider the case where the loss functions are all strongly convex, and extend our framework to derive regret bounds of $\mathcal{O}\left(\frac{N^{3/2}}{\alpha}\delta_{\max}\ln(T) + \frac{N\ln N\ln T}{\alpha\sqrt{1-\sigma_2}}\right)$, where $\delta_{\max}$ is the maximum number of missing observations averaged over agents, showing that strong convexity enables improved regret guarantee under DOCO with delayed feedbacks. 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.

5.1.1 Related Works

Distributed online convex optimization Distributed online convex optimization (DOCO) 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 distributed optimization focused on offline settings, leveraging techniques from gossip algorithms — originally used to achieve consensus (i.e. local convergence to an average $\bar{x}$ when each agent u in a graph holds static local information $x(u)$) — to enable distributed optimization Boyd et al. (2011), Nedic and Ozdaglar (2009a). The first formal treatment of the online counterpart was given by Hosseini et al. (2013), who analyzed a dual averaging algorithm and established sublinear regret guarantees. Specifically, they showed that a regret bound of order $(1-\sigma_2)^{-1/2}\sqrt{NT}$ 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 distributed mirror descent Shahrampour and Jadbabaie (2018), for which a similar regret rate is provable and accelerated gossiping for DOCO Wan et al. (2024b). The method from (Wan et al., 2024b) notably improves the previous regret bound by a factor scaling loosely with $\sqrt{N}/\log(N)\cdot(1-\sigma_2(W))^{-1/4}$. The DOCO framework has seen various extensions, including work on settings with random communication graphs Hosseini et al. (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 This chapter 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

[†]We also remark that Nguyen et al. (2024) requires β-smoothness for the loss functions for all agents, which is not assumed in this chapter.

delays (Joulani et al., 2013). Various aspects of delayed feedback have been explored, including adaptive regret guarantees (Joulani et al., 2016a), diverse delay structures (Gatmiry and Schneider, 2024, Bar-On and Mansour, 2025, Ryabchenko et al., 2025), and limited-feedback scenarios (Cesa-Bianchi et al., 2016b, Cella and Cesa-Bianchi, 2020, Zimmert and Seldin, 2020b, Lancewicki et al., 2022, van der Hoeven et al., 2023a).

DOCO with delayed feedbacks In DOCO 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 distributed 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 DOCO and its bandit counterpart with event-triggered communications and delayed feedback. For the more challenging setting with time- and agent-varying delays, Nguyen et al. (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.

5.2 Preliminary

Throughout this chapter, 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 DOCO, agents are organized in a communication network defined by a connected and undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The node set $\mathcal{V} = [N]$ corresponds to the N agents, and $\mathcal{E}$ denotes the set of edges indicating permissible communication among agents. We use $\mathcal{V}$ and $[N]$ interchangeably throughout the chapter. Each agent $u \in \mathcal{V}$ is associated with an arbitrary and unknown sequence of *local loss functions* $\ell_1(u, \cdot), \ell_2(u, \cdot), \ldots \ell_T(u, \cdot)$ decided by an adversary, where $\ell_t(u, \cdot) : \mathcal{X} \subseteq \mathbb{R}^n \to [0, 1]$ for $t \in [T]$ has a bounded feasible domain and is L-Lipschitz with respect to ℓ_2 norm.

Assumption 5.1 (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 5.2 (Lipschitzness)**.** *For every $t \in [T]$, we assume that $\ell_t(u, \cdot)$ is convex and L-Lipschitz with respect to $\|\cdot\|_2$ for all $u \in \mathcal{V}$.*

The learning protocol of DOCO 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 $u \in \mathcal{V}$ selects an action $x_t(u) \in \mathcal{X}$ simultaneously and suffers a loss $\ell_t(u, x_t(u))$. For each agent u, instead of observing the gradient $\nabla\ell_t(u, x_t(u))$ 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 $t + d_t(u) \leq T$, for all $u \in \mathcal{V},\ t \in [T]$ 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 $\mathcal{G}$. The goal for all agents is to minimize each agent's regret defined as follows, which is in terms of the global loss function $\sum_{v\in\mathcal{V}} \ell_t(v, x)$:

$$\mathrm{Reg}_T(u) \triangleq \max_{x\in\mathcal{X}} \Big(\sum_{t=1}^{T} \sum_{v\in\mathcal{V}} (\ell_t(v, x_t(u)) - \ell_t(v, x)) \Big). \tag{5.1}$$

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

It remains to introduce how agents communicate their information with each other in this network. Specifically, following previous works of DOCO (Yan et al., 2013, Hosseini et al., 2013, Wan et al., 2024b), 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 $\mathcal{G}$.

Definition 5.1. *A matrix $W \in [0, 1]^{N\times N}$ is a valid communication matrix with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if W satisfies that (i) $W(u, v) = 0$ if $u \neq v$ and $(u, v) \notin \mathcal{E}$; W is symmetric and doubly-stochastic meaning that (ii) $W(u, v) \geq 0,\ \forall u, v \in \mathcal{V}$; (iii) $W(u, v) = W(v, u),\ \forall u, v \in \mathcal{V}$; (iv) $\sum_{v\in\mathcal{V}} W(u, v) = 1,\ \forall u \in \mathcal{V}$. Consequently, a valid communication W is positive semi-definite with $0 \leq \sigma_2(W) < 1$ where $\sigma_2(W)$ is the second-largest eigenvalue.*

A typical construction of this matrix is as follows:

$$W = I_N - c \cdot \mathrm{Lap}(\mathcal{G}), \tag{5.2}$$

where $I_N \in \mathbb{R}^{N\times N}$ denotes the identity matrix and $\mathrm{Lap}(\mathcal{G})$ denotes the Laplacian of the graph $\mathcal{G}$ with $\mathrm{Lap}(\mathcal{G})(i, i) = \deg(i)$ for all $i \in \mathcal{V}$, $\mathrm{Lap}(\mathcal{G})(i, j) = -1$ if $i \neq j, (i, j) \in \mathcal{E}$, and $\mathrm{Lap}(\mathcal{G})(i, j) = 0$ if $i \neq j, (i, j) \notin \mathcal{E}$. c is a certain constant such that $0 < c \leq 1/\sigma_1(\mathrm{Lap}(\mathcal{G}))$, with $\sigma_1(\mathrm{Lap}(\mathcal{G}))$ being the largest eigenvalue of the Laplacian $\mathrm{Lap}(\mathcal{G})$. In particular, building row $W(u, \cdot)$ defined in Equation (5.2) 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 $\{x(u)\}_{u\in\mathcal{V}}$ for each agent where $x(u) \in \mathbb{R}^n$ represents the information agent u wants to communicate. In the context of DOCO, this information can correspond to various quantities such as predictions (Shahrampour and Jadbabaie, 2018) or loss gradients (Hosseini et al., 2013). In order to approximate the averaged vector $\bar{x} = \frac{1}{N}\sum_{u\in\mathcal{V}} x(u)$, Liu and Morse (2011) considers the following accelerated gossip process:

$$x^{k+1}(u) = (1 + \theta) \sum_{v\in\mathcal{N}_u} W(u, v)x^k(v) - \theta x^{k-1}(u), \tag{5.3}$$

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

Proposition 5.1 (Proposition 1 in Ye et al. (2023a)). *The iterations of Equation* (5.3) *with* $\theta = \left(1 + \sqrt{1 - \sigma_2^2(W)}\right)^{-1}$ *ensure that*

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

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 $u \in \mathcal{V}$, 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 $\delta_{\max} = \max_{t\in[T]} \frac{1}{N} \sum_{u\in\mathcal{V}} |m_t(u)|$ to be the maximum number of per-round missing observations averaged over all agents and $d_{\text{tot}} = \frac{1}{N} \sum_{t\in[T]} \sum_{u\in\mathcal{V}} d_t(u)$ to be the total delay averaged over all agents.

5.3 DOCO 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 $\widetilde{\mathcal{O}}\left(N\sqrt{d_{\text{tot}}} + \frac{N^{5/4}\sqrt{T}}{(1-\sigma_2(W))^{1/4}}\right)$ 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 $1 - \sigma_2(W)$.

5.3.1 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. (2024b), 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. (2024b), 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 (5.4). 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:

$$\theta = \frac{1}{1 + \sqrt{1 - \sigma_2^2(W)}}, \quad B = \left\lceil \frac{\sqrt{2}\ln(\sqrt{14N})}{(\sqrt{2}-1)\sqrt{1-\sigma_2(W)}} \right\rceil. \tag{5.6}$$

After completing all iterations within block s, each agent updates her decision for the next block by solving a Follow the Regularized Leader (FTRL) Equation (5.5) with learning rate $\eta_s(u)$. Then, different from Wan et al. (2024b) 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.

Algorithm 5.1: Accelerated Distributed 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:

$$z_s^{k+1}(u) = (1+\theta)\sum_{v\in\mathcal{V}} W(u,v) z_s^k(v) - \theta z_s^{k-1}(u). \tag{5.4}$$

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

Compute $x_{s+1}(u)$ for next block as follows:

$$x_{s+1}(u) = \arg\min_{x\in\mathcal{X}} \langle z_s^B(u), x\rangle + \frac{1}{\eta_s(u)}\|x\|_2^2. \tag{5.5}$$

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 $g_\tau(u) \triangleq \nabla \ell_\tau(x_{s(\tau)}(u))$, $s(\tau)$ is the block τ lies in.

Compute $z_{s+1}^{-1}(u)$ and $z_{s+1}^0(u)$ for next block:

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

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. (2024b).

The pseudo code of our algorithm is formally shown in Algorithm 5.1 and the following theorem shows that our algorithm achieves $\mathcal{O}(N\sqrt{d_{\text{tot}}} + N^{5/4}\sqrt{T}/(1-\sigma_2(W))^{1/4})$ when $\eta_s(u)$ is fixed over all blocks and is dependent on d_{tot}.

Theorem 5.1. *Assume each agent $u \in \mathcal{V}$ runs an instance of Algorithm 5.1 with a valid communication matrix W, parameters θ and B defined in Equation* (5.6)*, and a fixed learning rate*

$$\eta_s(u) = \eta = \frac{D}{L\sqrt{d_{\text{tot}} + \sqrt{N}BT}}. \tag{5.7}$$

Then, under Assumption 5.1 and Assumption 5.2, the regret is bounded as

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

Furthermore, when $d_t(u) = d(u)$ for all $t \in [T]$, we have

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

with $\eta = \frac{D}{L\sqrt{d_{\mathrm{tot}}+BT}}$.

Two remarks are as follows. First, note that Nguyen et al. (2024) considers the exact same case where d_{tot} is known and obtain a regret bound no better than $\mathcal{O}\left(\frac{N^2}{(1-\sigma_2)^2}\sqrt{d_{\mathrm{tot}}} + \frac{N\sqrt{N}}{1-\sigma_2}\sqrt{T}\right)$. Comparing to their results, our result not only achieves a better dependency on the spectral gap $\sigma_2(W)$ 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 $N^{5/4}/(1-\sigma_2(W))^{1/4}\sqrt{T}$ in Theorem 5.1 instead of $N\sqrt{N}/(1-\sigma_2(W))\sqrt{T}$ in their bound. For the delay related terms, 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. Second, we save an additional $N^{1/4}$ factor when delays are time-invariant and agent-specific. i.e when $d_t(u) = d(u)$ for all $t \in [T]$. Specifically, our results improves upon the $\mathcal{O}(N\sqrt{d_{\mathrm{tot}}} + \frac{N^{1.5}\sqrt{T}}{1-\sigma_2(W)})$ achieved by Cao and Basar (2022) and also matches the lower bound up to logarithmic factors in this setting as will be shown later. In addition, our bound also recovers the regret bound proven in Wan et al. (2024b) when $d(u) = 0$ for all $u \in \mathcal{V}$.

5.3.2 Proof Sketch

The full proof of Theorem 5.1 is deferred to Appendix D 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 $\bar{x}_s = \arg\min_{x\in\mathcal{X}}\{\langle\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{(s-1)B+1}(v)} g_\tau(v), x\rangle + \frac{N}{\eta}\|x\|_2^2\}$ 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 the analysis on online learning with delayed feedback, we further split it 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 $\mathcal{O}(\eta NL^2(d_{\mathrm{tot}} + BT))$.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 5.1. 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 $\mathcal{O}(\eta\sqrt{N}TL)$, which is the main technical part of the proof and require an involved analysis. When the delay is fixed over all rounds, we further show a tighter $\mathcal{O}(\eta TL)$ bound for $\sum_{s=1}^{T/B}\|x_s(v)-\bar{x}_s\|_2$, which removes an extra $\sqrt{N}$ factor. Finally, picking η optimally leads to our

final bound.

5.3.3 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 DOCO, 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, 2014b, Gyorgy and Joulani, 2021). In contrast, in the distributed 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 5.1 to perform the decision update and track the average gradients under delay. In addition, each agent also runs an instance of Algorithm 5.2 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 5.2. From a high level, Algorithm 5.2 closely mirrors the accelerated gossip routine of Algorithm 5.1, but instead focuses on gossiping the cumulative number of missing observations. Concretely, Algorithm 5.2 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 (5.7) by this local estimate till block $s-1$ as shown in Equation (5.8). At the end of the epoch s, similar to Algorithm 5.1, 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 5.1 alongside Algorithm 5.2 (with the same θ and B described in Equation (5.6)) to use $\eta_s(u)$ computed in Equation (5.8) to update $x_{s+1}(u)$. The following theorem shows that with this adaptive learning rate tuning, we achieve $\widetilde{\mathcal{O}}(N^{1.5}\sqrt{d_{\text{tot}}} + N^{1.5}\sqrt{T}/(1-\sigma_2(W))^{1/4})$ without knowing d_{tot}.

Theorem 5.2. *Assuming each agent $u \in [N]$ runs an instance of Algorithm 5.2 with a valid communication matrix W and parameters θ and B defined in Equation* (5.6) *together with an instance of Algorithm 5.1 parametrized by the same W, θ and B and using $\eta_s(u)$ computed by Algorithm 5.2. Then, under Assumption 5.1 and Assumption 5.2, the regret is bounded as*

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

Compared to Theorem 5.1 using a fixed learning rate with prior knowledge of d_{tot}, the degradation is a factor of $\sqrt{N}$ in the delay-dependent term and $N^{1/4}$ in the T-dependent term, coming from some technical issues when using adaptive learning rates. The proof of Theorem 5.2 is provided in Appendix D. 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

Algorithm 5.2: Accelerated Gossip Routine for the Adaptive Learning Rate for Agent u

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

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:

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

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

Count missing observations at the end of the block

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

· Update

$$\eta_{s+1}(u) = \frac{D}{L\sqrt{\sqrt{N}BT + B\cdot\zeta_s^B(u) + 3sB^2}}. \tag{5.8}$$

Compute first iterates for next block:

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

possibly non-decreasing learning rates $\eta_s(u)$.

5.3.4 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 5.3. *Let d be the constant feedback delay suffered by all agents $u \in [N]$ in the network. Then, there exists a graph $\mathcal{G} = (\{0,1,\ldots,N\}, \mathcal{E})$, with $N = 2M+1$ where M is an even integer, and a sequence of L-Lipschitz loss functions $\{\ell_1(0,\cdot),\ldots,\ell_1(N,\cdot)\},\ldots,\{\ell_T(0,\cdot),\ldots,\ell_T(N,\cdot)\}$ such that any algorithm has to suffer regret at least:*

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

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

Compared to this lower bound, our obtained upper bounds are optimal in the dependence on T, $1-\sigma_2(W)$, and d_{tot}, though they still incur 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 D. Our proof is adapted from the construction in Wan et al. (2024b) which considers 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)

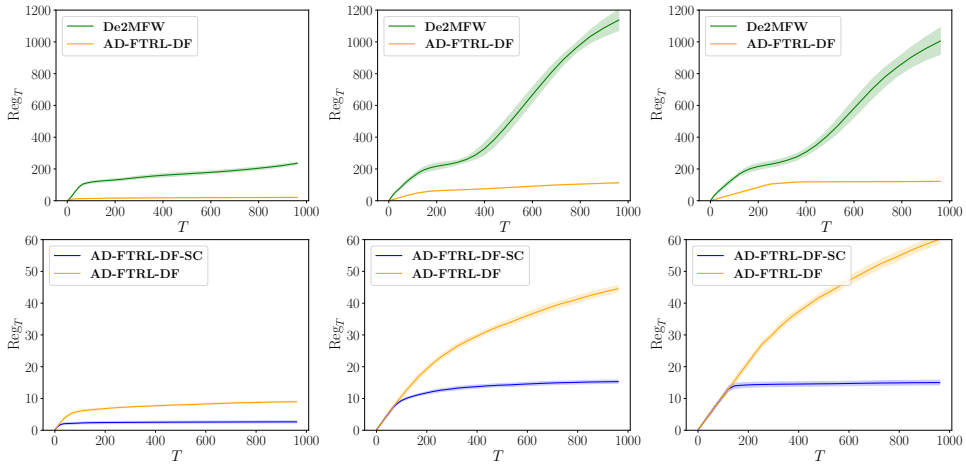


Figure 5.1: Comparison with relevant baselines across three network topologies—complete (left), grid (middle), and cycle (right)—under convex losses (top row) and strongly convex losses (bottom row).

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.

5.4 DOCO with Strongly-Convex Loss Functions

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

Assumption 5.3 (strong convexity)**.** *For every $t \leq T$ and $v \in \mathcal{V}$, we assume that $\ell_t(v, \cdot)$ is α-strongly convex: $\forall x, y \in \mathcal{X}$*

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

In order to show an improved regret bound when losses are strongly convex in DOCO with feedback delay, following the algorithm proposed in Wan et al. (2024b) for strongly convex functions, we propose our algorithm AD-FTRL-DF-SC outlined in Algorithm 5.3. Compared to AD-FTRL-DF shown in Algorithm 5.1, 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 (5.9)); second, we do not need to apply a gossip-based communication among agents to tune the learning rate adaptively but only need $\eta_s(u) = \frac{2}{\alpha s B}$ for all $u \in [N]$. The following theorem shows that Algorithm 5.3 achieves $\mathcal{O}(\frac{N\sqrt{N}\delta_{\max}}{\alpha} + N \ln N \ln T / (\alpha\sqrt{1 - \sigma_2(W)}))$ regret.

Theorem 5.4. *Assume each agent $u \in \mathcal{V}$ runs an instance of AD-FTRL-DF-SC with a valid communication matrix W and parameters θ and B defined in Equation* (5.6)*. Then, under Assumption 5.1, 5.2 and Assumption 5.3, the global regret is bounded as*

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

Algorithm 5.3: Accelerated Distributed 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**

$\eta_s = \frac{2}{\alpha s B}$

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:

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

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

Compute action $x_{s+1}(u)$:

$$x_{s+1}(u) = \arg\min_{x\in\mathcal{X}} \langle z_s^B(u), x\rangle + \frac{1}{\eta_s}\|x\|_2^2.$$

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

$$y_s^+(u) = \sum_{\tau \in o_{sB+1}(u)\setminus o_{(s-1)B+1}(u)} g_\tau(u) - \alpha B x_s(u). \tag{5.9}$$

Compute first iterates for next block:

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

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* $\bar{d} = \frac{1}{N}\sum_{v\in\mathcal{V}} d(v)$ *and the global regret is bounded as*

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

The full proof is deferred to Appendix D. To our knowledge, there are no previous results for DOCO 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 et al. (2025a), 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 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. (2024b) when $d(u) = 0$ for all $u \in [N]$. Finally, in Appendix D, 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), and $1 - \sigma_2(W)$.

5.5 Experiments

In this section, we evaluate the performance of our proposed algorithms in the delayed DOCO 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 et al. (2020), we define the local losses for all agents $v \in \mathcal{V}$ as

$$\ell_t(v, x) = \tfrac{1}{2}\big(\langle w_t(v), x\rangle - y_t(v)\big)^2, \tag{5.10}$$

where each feature vector $w_t(v)$ has independent coordinates drawn uniformly from $[-1, 1]$. Labels are generated as follows: for $1 \le v < N/2$, $y_t(v) = \varepsilon_t(v)$, 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:

$$\ell_t(v, x) = \tfrac{1}{2}\big(\langle w_t(v), x\rangle - y_t(v)\big)^2 + \tfrac{1}{2}\|x\|_2^2. \tag{5.11}$$

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 (5.2) 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 5.1) with adaptive learning rate tuning (Algorithm 5.2) against De2MFW (Nguyen et al., 2024). In the strongly convex loss setting, we compare our algorithm AD-FTRL-DF-SC (Algorithm 5.3) against AD-FTRL-DF (Algorithm 5.1) with adaptive learning rate tuning.

Results. Figure 5.1 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 (5.10) and Equation (5.11) for all three topologies. From the plots, we observe that for the losses defined in Equation (5.10), 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.

Chapter 6

Distributed Stochastic Multi-Armed Bandits Over Random Networks

6.1 Introduction

The emergence of large-scale cooperative systems holds true in various applications ranging from sensor networks (Ganesan et al., 2004, Zhu et al., 2016) to federated learning (Ye et al., 2023b, McMahan et al., 2017) and edge computing (Wang et al., 2022a, Ghoorchian and Maghsudi, 2020). It has naturally motivated interest in distributed multi-agent multi-armed bandit (MA-MAB) problem, where multiple agents collaboratively learn to optimize rewards. MA-MAB settings are typically categorized as homogeneous (Landgren et al., 2016a, Martínez-Rubio et al., 2019) or heterogeneous (Zhu et al., 2021, Xu and Klabjan, 2023), depending on whether the reward distributions for the same arm are identical across agents. The heterogeneous setting, in which reward distributions vary across agents, is significantly more general but more challenging, and thus has attracted growing attention. It introduces substantial difficulties, as agents must make sequential decisions under uncertainty while relying on limited information about both their own rewards and the actions or observations of other agents.

Another challenging aspect of MA-MAB lies in the underlying communication protocol, which constrains how agents share information with one another. Decentralized MA-MAB settings are more realistic than centralized ones—where all agents can communicate with any other agent—as they restrict communication to immediate neighbors defined by a graph structure. Much of the existing work has focused on time-invariant graphs (Zhu et al., 2021), where the communication graph remains fixed throughout. However, the complexity of many real-world decentralized systems, such as wireless ad-hoc networks, necessitates the use of time-varying graphs (Zhu and Liu, 2023), particularly random graphs (Xu and Klabjan, 2023). This added complexity significantly complicates both the communication and learning dynamics. Notably, (Xu and Klabjan, 2023) is the first to consider classical Erdős–Rényi (E-R) random graphs in the MA-MAB setting, where any two agents can communicate with probability p at each time step. However, it is possible that some pairs of agents can never communicate directly due to inherent topological constraints. This scenario has been formulated as a more general version of the E-R graph, where two agents can communicate with probability p only if there is an edge between them in a base graph. Note that when the base graph is a complete graph, it is equivalent to classical (E-R) random graphs, implying consistency.

To date, this setting remains unexplored, which motivates our work.

To date, a line of work has studied regret bounds under various graph structures, where connectivity or sequential connectivity is typically required. For example, in the context of time-invariant graphs, Martínez-Rubio et al. (2019) and Zhu et al. (2021) analyze distributed bandits over connected graphs and derive $\log T$ regret bounds. For time-varying graphs, Zhu et al. (2025) obtains $\log T$ regret bounds under the assumption of B-connectivity, where the union of any l consecutive graphs must be connected. In the classical Erdős–Rényi model (Erdos et al., 1960) over fully connected agent communication, Xu and Klabjan (2023) derive a regret of order $\log T$, but only under the assumption that $p \geq 1/2 + 1/2\sqrt{1 - (\epsilon/NT)^{2/N-1}}$ which is larger than $1/2$ and can even approach 1 when the number of agents N or the time horizon T is large. This is a strong assumption, as it may not hold in many real-world settings, but is required in their analysis to ensure that the graph is connected with high probability. Relaxing this connectivity requirement to allow arbitrary p presents a significant challenge. Moreover, their regret bound does not reflect how the link probability p impacts the regret. Incorporating p into the regret expression would significantly improve our understanding of how to choose p in practice—a gap that remains open. In this chapter, we address both of these gaps. To this end, we address the following key research question: *Can we solve MA-MAB under new Erdős–Rényi random networks and heterogeneous rewards, and derive regret bounds that captures graph complexity under much milder assumptions?*

Contribution. We provide an affirmative answer to the above question through the following contributions. Methodologically, we solve the MA-MAB problem over general Erdős–Rényi communication networks using a gossip algorithm, which is widely adopted in distributed settings. Moreover, we adopt an algorithm based on arm elimination with a minimal number of arm pulls required for each arm to guarantee sufficient information collecting for each agent, which addresses the E-R communication networks.

Analytically, we study the regret of MA-MAB under our proposed algorithm over general Erdős–Rényi communication networks for any $p \in (0, 1]$, leading to novel contributions. We are the first to 1) explore general Erdos-Rényi graphs induced by any fixed connected base graphs beyond classical setting that assumes a complete base graphs; 2) obtain the a tighter and more interpretable regret bound, which generalizes the bound from the homogeneous fixed-graph setting studied in Martínez-Rubio et al. (2019) to our more challenging heterogeneous setting, as shown in Section 6.4; 3) relax the assumption of a sufficiently large p used in Xu and Klabjan (2023) and reduce the order of N in the upper bound in their analysis of classical Erdős–Rényi models with complete base graphs. Precisely, we obtain a regret upper bound of order of $\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log T}{\Delta_k} + \frac{N^2 \log T}{p\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))} + \frac{KN^2 \log T}{p}\right)$ where the first term accounts for an optimal centralized regret, aligning with the lower bound established in Section 6.4.2, and the last two terms capture the effects of both the link probability p and the algebraic connectivity $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ of the base graph. Moreover, we uniquely characterize how the regret upper bounds are influenced by p and $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$, which highlights a tradeoff between the communication cost (i.e., the number of communication rounds) and regret performance.

Numerically, we implement our proposed algorithm and conduct experiments to validate our theoretical results with multiple random graph settings and edge generation probability p. We also compare our methods with existing approaches to demonstrate their effectiveness.

6.1.1 Related Works

Distributed online algorithm. Our framework builds on the classical line of work on gossip algorithms (Xiao and Boyd, 2004a, Boyd et al., 2006). Specifically, when an agent has access only to local information and communicates solely with its immediate neighbors, it adopts a gossip algorithm to aggregate information from agents beyond its local neighborhood, based on a weight matrix. Notably, gossip algorithms are widely used in distributed optimization. For example, Duchi et al. (2011), Nedic and Ozdaglar (2009b) address distributed convex optimization problems using gossip algorithms. Subsequently, Hosseini et al. (2013), Yan et al. (2013) extend the gossip approach to the online setting and achieve a regret of order $\sqrt{T}$, assuming convex loss functions. Later, Mateos-Núnez and Cortés (2014) consider distributed online optimization over B-connected and design a distributed online primal-dual algorithm coupled with a gossip protocol, also achieving a $\sqrt{T}$ regret bound. More recently, Lei et al. (2020) study the same problem over random communication networks (Erdős–Rényi networks) and obtain the same regret order. They further characterize how regret is affected by the link probability p in the Erdős–Rényi model and the algebraic connectivity of the base graph $\mathcal{G}$. We consider the same graph topology but focus on multi-agent multi-armed bandits, which differ significantly from online convex optimization and introduce the additional challenge of learning the dynamics of bandits. For a more comprehensive survey on distributed online optimization, we recommend the reader to Li et al. (2023) and Yuan et al. (2024).

Distributed multi-agent multi-armed bandit. Along the line of work on MA-MAB, several studies (Landgren et al., 2016a,b, Zhu et al., 2020, Chawla et al., 2020, Wang et al., 2022b, 2020, Zhu et al., 2025, Martínez-Rubio et al., 2019, Agarwal et al., 2022, Sankararaman et al., 2019, Zhu et al., 2021, Zhu and Liu, 2023, Xu and Klabjan, 2023, Yi and Vojnovic, 2023) have investigated both homogeneous and heterogeneous settings. In homogeneous settings, numerous work incorporate gossip algorithms to reduce regret in terms of the number of agents; information sharing among agents accelerates the concentration of reward observations. For example, Landgren et al. (2016a,b) first formulate this problem and solve it using gossip algorithms. Martínez-Rubio et al. (2019) achieve the optimal centralized regret—independent of the number of agents and matching that of the single-agent bandit—plus an additional term that depends on the spectral gap of the communication matrix. (Chawla et al., 2020) characterizes the regret-communication trade-off, considers circular ring graphs, and establishes regret improvement. Wang et al. (2022b, 2020) further focus on optimizing communication efficiency to minimize the number of communication rounds while guaranteeing regret performance. In contrast, we consider more challenging heterogeneous settings and also characterize the regret-communication trade-off. Here gossiping enables regret reduction in terms of the order of T; without information from other agents, the regret can easily be linear in T Xu and Klabjan (2025). In this direction, Zhu et al. (2021) is the first to study heterogeneous rewards over a time-invariant connected graph and establishes regret bounds of order $\log T$. (Zhu and Liu, 2023) extends this to B-connected graphs, also achieving regret bounds of order $\log T$. Recently, Xu and Klabjan (2023) propose a gossip-based algorithm for the classical E-R model and obtain regret bounds of order $\log T$ when p is sufficiently larger than $1/2$ to ensure the graph is connected with high probability. More generally, Yi and Vojnovic (2023) consider MA-MAB with heterogeneous rewards in the adversarial environment and establishes a regret bound of order $T^{2/3}$. In contrast, we consider the stochastic setting with general Erdős–Rényi random networks, without any assumption

on p.

6.2 Setting and Notations

In this section, we formally define the problem of interest, starting by the presentation of general notations.

General Notations. For a matrix $M \in \mathbb{R}^{p\times q}$, let $[M]_{i,j}$ denote the entry in the i-th row and j-th column. Given a doubly stochastic matrix $P \in \mathbb{R}^{d\times d}$, we denote by $\lambda_2(P)$ its second largest eigenvalue. Let $e_i \in \mathbb{R}^d$ be the i-th standard basis vector, $\mathbf{1} \in \mathbb{R}^d$ the all-ones vector, and I_d the $d \times d$ identity matrix. We use $\mathbb{I}\cdot$ to denote the indicator function, which equals 1 if the condition inside holds, and 0 otherwise. Moreover, we use $[n] = \{1, 2, \cdots, n\}$ to denote a set of indices.

Multi-agent Multi-armed Bandit. We consider a multi-agent multi-armed bandit (MA-MAB) setting involving N agents. The bandit problem is run over a time horizon of T. In each round $t \in [T]$, every agent $i \in [N]$ selects an arm $A_i(t) \in [K]$. The local reward of arm k for agent i follows an unknown, time-invariant probability distribution $\mathbb{P}_{i,k}$ supported on $[0, 1]$, with mean $\mu_{i,k} = \mathbb{E}[X \sim \mathbb{P}_{i,k}] \in [0, 1]$. When agent i selects arm $A_i(t)$ at round t, it only observes a local stochastic reward $X_{i,A_i(t)}(t)$ drawn independently from the distribution $\mathbb{P}_{i,A_i(t)}$ each round. However, the agent's true objective depends on the global reward, defined as $X_{A_i(t)}(t) \coloneqq \frac{1}{N}\sum_{j\in[N]} X_{j,A_i(t)}(t)$. Accordingly, we define the global mean reward for arm k as $\mu_k \coloneqq \frac{1}{N}\sum_{j\in[N]} \mu_{j,k}$, where $\mu_{j,k}$ is the mean of $\mathbb{P}_{j,k}$. We denote $T_{i,k}(t)$ to be the total number of times agent i has selected arm k up to time t by $T_{i,k}(t) = \sum_{s=1}^{t} \mathbb{I}(A_i(s) = k)$.

Throughout, we focus on a decentralized setting where N agents are distributed on undirected time-varying graphs and communicate via the graphs. Specifically, we consider a communication protocol based on **Erdős–Rényi random graphs**. More precisely, agents communicate via a time-varying graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$, where the vertex set $\mathcal{V} = [N]$ is fixed, but the edge set $\mathcal{E}_t$ may vary at each round t. Uniquely, each **communication graph** $\mathcal{G}_t$ is generated based on an underlying undirected, fixed, and connected (but not necessarily complete) **base graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that defines all feasible communication edges. The connectedness of the base graph is essential to avoid linear regret as proved in Theorem 4 in Xu and Klabjan (2025). In every round t, each edge in graph $\mathcal{G}$ is independently in $\mathcal{G}_t$ with probability $p \in (0, 1]$. Two agents can communicate if and only if there is an edge between them in $\mathcal{G}_t$—namely, an active edge. Formally, the random graph generation reads as follows.

Assumption 6.1 (Erdős–Rényi Random Graph)**.** *In each round t, the random communication graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$, generated from the base graph $\mathcal{G}$ meets:*

$$\mathbb{P}\left((i, j) \in \mathcal{E}_t\right) = \begin{cases} p, & \text{if } (i, j) \in \mathcal{E}, \\ 0, & \text{otherwise,} \end{cases}$$

for all vertices $i, j \in \mathcal{V}$.

Thus, $\mathcal{E}_t \subseteq \mathcal{E}$ for all t. Let $\mathcal{N}_i = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$ denote the neighbors of agent i in the base graph $\mathcal{G}$, and $\mathcal{N}_i(t) = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}_t\}$ denote the active neighbors of agent i in round t, clearly

satisfying $\mathcal{N}_i(t) \subseteq \mathcal{N}_i$.

The objective is to design a distributed algorithm π that minimizes the total global regret over T rounds. Let $\mu^* = \max_{k\in[K]} \mu_k$ be the optimal global average reward. The total regret for algorithm π is defined as

$$\mathrm{Reg}_T(\pi) = NT\mu^* - \sum_{i\in[N]} \sum_{t=1}^{T} \mu_{A_i(t)} = \sum_{i\in[N]} \sum_{t=1}^{T} \Delta_{A_i(t)} \tag{6.1}$$

where $\Delta_k = \mu^* - \mu_k$ is the global suboptimality gap for arm k.

6.3 Algorithm: Gossip Successive Elimination

In this section, we present the proposed methodology. Unlike prior work on gossip bandits, which assumes fixed graphs or graphs that become connected within a few rounds (with high probability), our setting faces the core challenges of randomness and disconnection in the communication graph. To address this, our algorithm guarantees an upper bound on the number of rounds needed for any agent's information to reach its neighbors via gossip, enabling accurate estimate construction. It also uses a round-robin strategy combined with an arm elimination strategy (Even-Dar et al., 2006) to ensure that agents maintain consensus on arm pull counts. More specifically, we present the precise steps in Algorithm 6.1, namely, Gossip Successive Elimination (GSE), which integrates the arm elimination strategy with a gossip-based communication protocol.

The agent needs several parameters, including the link probability p and algebraic connectivity $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ as inputs. Initially, agent i's active arm set $\mathcal{S}_i$ is set to the full set of arms $[K]$, while both the local reward estimate $\widehat{\mu}_{i,k}(t)$ and global reward estimate $z_{i,k}(t+1)$ are initialized to zero ($t = 0$). Notably, Algorithm 6.1 requires knowledge of the link probability p and the algebraic connectivity $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$. Following the standard practice in existing work Martínez-Rubio et al. (2019), Zhu et al. (2021), Xu and Klabjan (2023), we assume that $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ is known, while allowing the link probability p to remain unknown. In this case, the algorithm can still be implemented, and the corresponding regret upper bound preserved, by adding a short burn-in phase to estimate a value $\widehat{p} \in (p/2, p]$. This estimation requires only $\mathcal{O}(\log T)$ time steps. We refer the reader to Appendix E.3 for the details and theoretical analysis of this procedure.

At each round t, agent i observes its active neighbors $\mathcal{N}_i(t)$ in the communication graph $\mathcal{G}_t$, which is randomly generated from the base graph $\mathcal{G}$ according to Assumption 6.1. As part of our key contributions, we consider a weighting matrix W_t for gossip as

$$W_t = I_N - \frac{1}{N} \mathrm{Lap}(\mathcal{G}_t), \tag{6.2}$$

where $\mathrm{Lap}(\mathcal{G}_t)$ denotes the Laplacian matrix of the communication graph $\mathcal{G}_t$.

The execution steps of agents run as follow. Agent i selects arm $A_i(t)$ from the active set $\mathcal{S}_i$ with the least number of pulls $T_{i,k}(t)$, observes the local reward of $A_i(t)$, and updates the reward estimates as follows:

$$\widehat{\mu}_{i,k}(t) = \frac{1}{T_{i,k}(t) \vee 1} \sum_{\tau=1}^{t} \mathbb{I}\{A_i(\tau) = k\} X_{i,k}(\tau),$$

Algorithm 6.1: `Gossip Successive Elimination` for Agent $i \in [N]$

1: **Input:** Algebraic connectivity $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$, total time horizon T, set of arms $[K]$, link probability p
2: **Initialization:** Active set $\mathcal{S}_i \leftarrow [K]$, local reward estimate $\widehat{\mu}_{i,k}(0) \leftarrow 0$, global reward estimate $z_{i,k}(1) \leftarrow 0$ for all $k \in [K]$.
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Select arm $A_i(t) \in \mathcal{S}_i$ with the minimum pull count $T_{i,k}(t)$ and update $T_{i,k}(t)$
5: Receive feedback and update statistics for each arm using Equation (6.3)
6: Remove arm $k \in \mathcal{S}_i$ if there exists an arm $k' \in \mathcal{S}_i$, $k' \neq k$, satisfying the elimination condition in Equation (6.6)
7: Update the active set $\mathcal{S}_i$ according to Equation (6.7)

$$z_{i,k}(t+1) = \sum_{j \in \mathcal{N}_i(t) \cup \{i\}} [W_t]_{i,j} z_{j,k}(t) + \widehat{\mu}_{i,k}(t) - \widehat{\mu}_{i,k}(t-1), \tag{6.3}$$

where $X_{i,k}(\tau)$ is the feedback observed by agent i from pulling arm k at round τ. The global estimate $z_{i,k}(t)$ is updated via a gossip protocol: at each round t, agent i aggregates its own and its active neighbors' estimates, weighting each neighbor $j \in \mathcal{N}_i(t) \cup \{i\}$ by the corresponding entry $[W_t]_{i,j}$ of the matrix W_t. We also define the gossip-based upper and lower confidence bounds for arm k, which remain key criteria for arm elimination and updating $\mathcal{S}_i$, as

$$\mathrm{GUCB}_{i,k}(t) = z_{i,k}(t) + c_{i,k}(t), \ \mathrm{GLCB}_{i,k}(t) = z_{i,k}(t) - c_{i,k}(t).$$

Here the confidence bound $c_{i,k}(t)$ reads as

$$c_{i,k}(t) = \sqrt{\frac{4\log(T)}{N \max\{T_{i,k}(t) - KL^*, 1\}}} + \frac{4(\sqrt{N} + \tau^*)}{\max\{T_{i,k}(t) - KL^*, 1\}}, \tag{6.4}$$

with

$$\tau^* = \left\lceil \frac{2N\log(T)}{p\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))} \right\rceil, \ L^* = N \left\lceil -\frac{2\log(NT)}{\log(1-p)} \right\rceil. \tag{6.5}$$

The confidence radius in Equation (6.4) decomposes into two parts. The first term captures the **estimation error**, arising from the statistical variance of independently sampled rewards for each arm. The second term captures the **consensus error**, which stems from the cumulative approximation error error incurred as agents reach consensus via a gossip-based communication protocol.

Arm elimination occurs if and only if there exists an arm $k' \neq k$ in $\mathcal{S}_i$ that meets the condition

$$\mathrm{GLCB}_{i,k'}(t) \geq \mathrm{GUCB}_{i,k}(t), \tag{6.6}$$

which means that arm k' has a higher global reward estimate than arm k with high probability. Subsequently, the active set $\mathcal{S}_i$ is updated as

$$\mathcal{S}_i \leftarrow \bigcap_{j \in \mathcal{N}_i(t) \cup \{i\}} \mathcal{S}_j. \tag{6.7}$$

This procedure coupled with the arm-elimination strategy ensures that, for every agent, each arm in the active set is pulled a roughly equal number of times, ensuring consensus.

6.4 Regret Analyses

In this section, we analyze the regret of the proposed algorithm and establish an upper bound on the corresponding regret, demonstrating its theoretical effectiveness. Additionally, we derive a lower bound for our new problem setting, highlighting the problem's inherent complexity and showing that the algorithm is nearly optimal up to some interpretable factors.

6.4.1 Upper Bound

We start by presenting the regret upper bound for Algorithm 6.1. To that end, we first introduce several technical lemmas that play a key role in the regret analysis. The proofs can be found in Appendix E.2.

Lemma 6.1. *Let us assume that the communication protocol follows Assumption 6.1. Then we have that for Algorithm 6.1, for any agent $i \in [N]$ and any arm $k \in [K]$, and any $t \in [T]$, the following holds with probability at least $1 - \frac{3NK}{T}$,*

$$|z_{i,k}(t) - \mu_k| \leq c_{i,k}(t),$$

where $c_{i,k}(t)$ is the confidence bound defined in Equation (6.4), and τ^* and L^* are the parameters introduced in Equation (6.5).

Importantly, we show that the estimation error compared to the global mean value is upper bounded by the confidence bound $c_{i,k}(t)$. Notably, $c_{i,k}(t)$ is monotonically decreasing in the number of pulls of arm k by agent i, i.e., $T_{i,k}(t)$, when $T_{i,k}(t) > KL^*$. Here, KL^* is the minimal number of pulls required to ensure that agent i has collected sufficient information from all other agents. This also implies that the global estimate $z_{i,k}(t)$ becomes increasingly accurate and approaches the true mean reward μ_k as the number of pulls increases. As a result, based on Lemma 6.1, we derive the following regret bound for individual agents.

Lemma 6.2. *Let $\mathrm{Reg}_{i,T}(\pi) = T\mu^* - \sum_{t=1}^{T} \mu_{a_i(t)}$ denote the regret incurred by agent i using policy π. Under the communication protocol assumed in Assumption 6.1, the regret of agent i under Algorithm 6.1 is guaranteed to meet*

$$\mathrm{Reg}_{i,T}(\mathrm{GSE}) \leq \sum_{k:\Delta_k>0} \left(\frac{64 \log(T)}{N\Delta_k} + 16 \left(\sqrt{N} + \tau^* \right) \right) + KL^* + 3KN\Delta_{\max}$$

where $\Delta_{\max} = \max_{k\in[K]} \Delta_k$ denotes the largest reward gap across all arms.

In other words, Lemma 6.2 shows that the regret incurred by any individual agent can be effectively bounded. This result naturally extends to the global regret, as stated in the theorem below.

Theorem 6.1. *The global regret defined in Equation* (6.1) *for Algorithm 6.1 is bounded as:*

$$\mathrm{Reg}_T(\mathrm{GSE}) = \sum_{i\in[N]} \mathrm{Reg}_{i,T}(\mathrm{GSE}) \leq \mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log(T)}{\Delta_k} + \frac{N^2\log(T)}{p\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))} + \frac{KN^2\log(NT)}{p}\right),$$

where $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ *is the second smallest eigenvalue of* $\mathrm{Lap}(\mathcal{G})$.

We continue our discussion on how the base graph $\mathcal{G}$ topology affects the regret bound. In addition to the parameter p, which determines the difference between $\mathcal{G}$ and $\mathcal{G}_t$, another key factor is $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$, which reflects the topology of the base graph $\mathcal{G}$. This value is the algebraic connectivity or Fiedler value of $\mathcal{G}$, which reflects how well connected the overall graph is. To illustrate this, we next provide more explicit regret upper bounds by specifying $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ for different base graph topologies commonly used in distributed optimization Duchi et al. (2011). The following corollary summarises how the choice of the random gossip matrix in Equation (6.2) interacts with different topologies of the base graph. The proof of $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ for various base graphs $\mathcal{G}$ can be found in Corollary 1 of Duchi et al. (2011).

Corollary 6.1. *For specific choices of the base graph* $\mathcal{G}$*, the regret upper bound in Theorem 6.1 simplifies as follows:*

1) When $\mathcal{G}$ *is a complete graph with* $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G})) = N$*, and refining* $L^* = \left\lceil -\frac{2\log(NT)}{\log(1-p)} \right\rceil$ *in Equation* (6.5)*, the regret upper bound simplifies to:*

$$\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log T}{\Delta_k} + \frac{KN\log T}{p}\right).$$

2) When $\mathcal{G}$ *is a* $\sqrt{N}\times\sqrt{N}$ *2D grid, we have* $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G})) = 2\left(1-\cos\left(\frac{\pi}{\sqrt{N}}\right)\right) = \Theta(1/N)$*. The corresponding regret bound becomes:*

$$\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log T}{\Delta_k} + \frac{N^2\,(K+N)\log T}{p}\right).$$

3) When $\mathcal{G}$ *is an expander graph with a bounded ratio between minimum and maximum node degrees, we have* $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G})) = \Theta(1)$*. In this case, the regret bound simplifies to:*

$$\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log T}{\Delta_k} + \frac{KN^2\log T}{p}\right).$$

Remark 6.1 (Comparison of Regret Bounds)**.** *In Theorem 6.1, for any fixed connected base graph* $\mathcal{G}$ *and* $0<p\leq 1$*, we obtain the optimal centralized regret* $\mathcal{O}\left(\sum_{k:\Delta_k>0}\frac{\log T}{\Delta_k}\right)$ *(see the lower bound in Section 6.4.2) plus an additional term* $\mathcal{O}\left(\frac{N^2\log T}{p\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))} + \frac{KN^2\log T}{p}\right)$*. We emphasize that our regret bound outperforms existing work—many of which are special (degenerated) cases of our more general framework—and is easier to interpret. Notably, Xu and Klabjan (2023) study MA-MAB under the classical E-R model (where* $\mathcal{G}$ *is a complete graph) with p largely over* $1/2$*, and derive a regret bound of* $\mathcal{O}\left(\sum_{k:\Delta_k>0}\frac{N\log T}{\Delta_k}\right)$*. In the same setting, based on Corollary 6.1, we obtain a regret*

bound of $\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log(T)}{\Delta_k} + KN\log(T)\right)$*, which is significantly smaller. Also, Zhu and Liu (2023) consider B-connected graphs, which do not capture E-R random graphs; the distinction between the two models is discussed in detail in Yuan et al. (2024). Finally, when* $p = 1$*, the communication graph becomes time-invariant, which corresponds to most existing work where connected graphs are assumed. For example, Zhu et al. (2021) and Zhu and Liu (2023) study such settings. The former derive a regret bound of* $\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{N^2\log T}{\Delta_k}\right)$*, which is worse than our results with a dependency on* N*. The latter obtain* $\mathcal{O}\left(\max\left(\sum_{k:\Delta_k>0} \frac{N\log T}{\mathcal{N}_k\Delta_k}, K_1, K_2\right)\right)$*, where* K_1 *and* K_2 *depend on* T *but lack explicit formulas, may grow arbitrarily large, and are difficult to interpret—at least to the best of our knowledge.*

Remark 6.2 (Regret and Communication Trade-off)**.** *We emphasize that our regret upper bound exhibits a novel trade-off between regret performance and communication efficiency in the presence of random graphs. It is straightforward to observe that increasing* p *reduces the regret bound but increases communication overhead, thereby lowering communication efficiency. For classical E-R graphs, the expected number of agent-to-agent communications per agent over the given time horizon is* pNT*, while the expected regrets decrease as* p *increases. Therefore, for a fixed time horizon* T *and a given base graph* $\mathcal{G}$*, the parameter* p *can be tuned to balance communication overhead and reward maximization, informing practical decision making.*

6.4.2 Lower Bound

In this section, we establish a lower bound on the global regret, as defined in Equation (6.1). Unlike the upper bound, this lower bound applies to any reasonable algorithm under a specific problem instance, highlighting the problem complexity. Detailed proofs are deferred to Appendix E.2. We begin by introducing several definitions related to the problem instance and the notion of a reasonable algorithm.

Definition 6.1 (Gaussian Instance)**.** *An instance* ν *is called a* Gaussian instance *if, for every agent* $i \in [N]$ *and arm* $k \in [K]$*, the reward distribution* $\mathbb{P}_{i,k}$ *is Gaussian with unit variance.*

Definition 6.2 (Consistent Policy)**.** *Let* $\mathcal{I}$ *be a class of problem instances. Let* $\mathrm{Reg}_T^\nu(\pi)$ *denote the regret incurred by policy* π *on instance* ν*. A policy (algorithm)* π *is said to be* consistent *on* $\mathcal{I}$ *if there exist constants* $C > 0$ *and* $s \in (0,1)$ *such that* $\mathrm{Reg}_T^\nu(\pi)$ *meets* $\mathrm{Reg}_T^\nu(\pi) \le CT^s$ *for all instances* $\nu \in \mathcal{I}$*.*

We next present the problem instance constructed to establish the regret lower bound. Note that an alternative, equivalent expression of the global regret reads as

$$\mathrm{Reg}_T^\nu(\pi) = \sum_{k\in[K]} \Delta_k \sum_{i\in[N]} \mathbb{E}[T_{i,k}(T)].$$

Thus, to derive a lower bound on regret, it suffices to lower bound the total expected number of pulls $\sum_{i\in[N]} \mathbb{E}[T_{i,k}(T)]$ for suboptimal arms $k \in [K]$. To this end, we consider a Gaussian instance ν where each $\mathbb{P}_{i,k} = \mathcal{N}(\mu_{i,k}, 1)$ with the random graph communication protocal based any connected

base graph $\mathcal{G}$ and connection probability p, and construct a perturbed instance ν' such that:

$$\mathbb{P}'_{i,a} = \begin{cases} \mathcal{N}(\mu_{i,a}, 1), & \text{if } a \neq k, \\ \mathcal{N}(\mu_{i,a} + (1+\varepsilon)\Delta_a, 1), & \text{if } a = k, \end{cases}$$

for a small constant $\varepsilon \in (0,1)$ representing the level of perturbation. The communication protocal for ν' is the same as that for ν. Under this perturbation, which defines a new problem instance, we derive the following information-theoretic inequality:

$$\sum_{j\in[N]} \mathbb{E}[T_{j,k}(T)] \cdot \frac{(1+\varepsilon)^2\Delta_k^2}{2} \geq \log\left(\frac{NT\varepsilon\Delta_k}{4\left(\mathrm{Reg}_T^{\nu}(\pi) + \mathrm{Reg}_T^{\nu'}(\pi)\right)}\right), \tag{6.8}$$

which imposes a lower bound on the total number of pulls of arm k across all agents.

By applying this inequality to a consistent policy π and rearranging the terms, we obtain the following lower bound on regret. The formal statement reads as follows.

Theorem 6.2. *Let π be a consistent policy on the class $\mathcal{I}$ of Gaussian instances for some $s \in (0,1)$. Then, for all instances $\nu \in \mathcal{I}$ and any $\varepsilon \in (0,1]$, the following holds:*

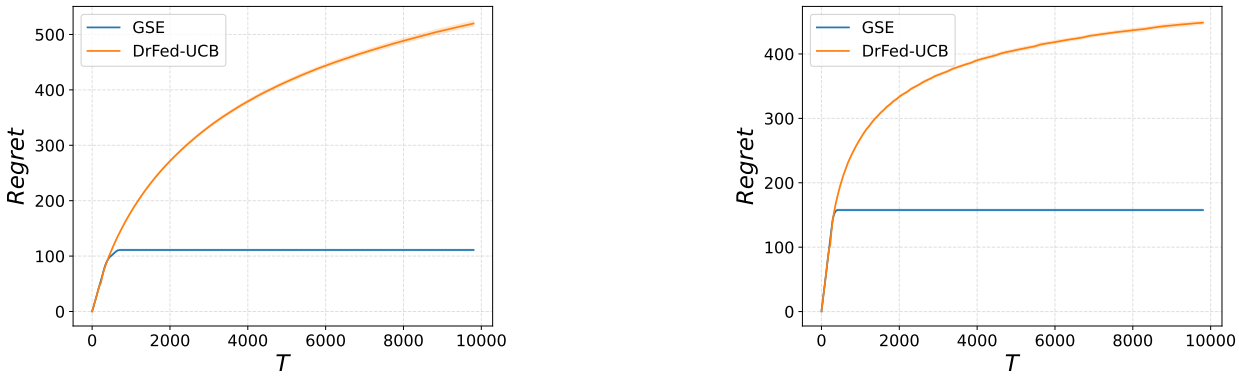
$$\lim_{T\to\infty} \frac{\mathrm{Reg}_T^{\nu}(\pi)}{\log T} \geq \sum_{k:\Delta_k>0} \frac{2(1-s)}{(1+\varepsilon)^2\Delta_k}.$$

Remark 6.3 (Comparison with Upper Bounds)**.** *Recall that the regret upper bound in Theorem 6.1 consists of two components: a centralized term and additional terms that capture the influence of the communication graph. The lower bound in Theorem 6.2 shows that the centralized component, $\mathcal{O}\left(\sum_{k:\Delta_k>0} \frac{\log T}{\Delta_k}\right)$, is tight, thereby establishing the optimality of the centralized regret achieved by Algorithm 6.1. This result is intuitive: the problem effectively involves N agents collaboratively solving a global multi-armed bandit task. With appropriate information sharing, the collective performance can match that of a single-agent bandit problem with full access to all rewards. Hence, achieving a global regret of the same order as the classical centralized bandit setting is both natural and optimal.*

6.5 Experiments

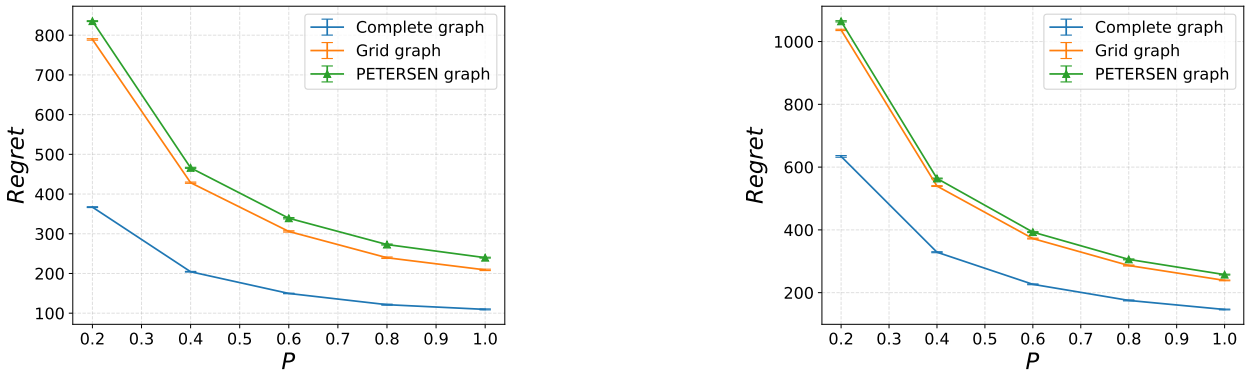
In this section, we demonstrate the effectiveness of our algorithm through numerical experiments on both synthetic and real-world datasets. The objective is twofold. First, we show that the cumulative regret of our algorithm grows logarithmically with respect to T and is significantly smaller than that of existing benchmarks, thereby validating our theoretical findings. We use DrFed-UCB, proposed by (Xu and Klabjan, 2023), as the baseline. Second, we conduct a simulation study to examine how the regret depends on the link probability p and the algebraic connectivity of the base graph $\mathcal{G}$, as reflected in the regret bound. We evaluate the impact of different values of p across various base graphs, including the complete graph, grid, and Petersen graph (Holton and Sheehan, 1993). Each edge in the base graph $\mathcal{G}$ appears in the communication graph $\mathcal{G}_t$ with probability p.

Experimental Settings. For synthetic experiment setting, We set $T = 10000$, $N = 16$, and $K = 5$; for the Petersen graph, we use $N = 10$ by definition. For the comparison with DrFed-UCB,



(a) Synthetic data (b) Real-world data

Figure 6.1: Comparison of the empirical results of our algorithm and DrFed-UCB. The base graph is a complete graph and the link probability $p = 0.9$.



(a) Synthetic data (b) Real-world data

Figure 6.2: Regret of our algorithm on different base graphs with different p value.

we consider a complete graph and a high link probability ($p = 0.9$), as required therein. Before the game starts, we sample each q_i independently and uniformly from the interval $[0, 1]$ for each agent i. The local mean reward of arm k on agent i is given by $\mu_{i,k} = q_i \cdot \frac{k-1}{K-1}$, and the global mean reward of arm k is $\mu_k = \frac{k-1}{K-1} \cdot \frac{\sum_{i\in[N]} q_i}{N}$. At each time step t, each agent $i \in [N]$ selects an arm and observes the local reward. For real-world experiments, we use the MovieLens dataset and refer to Yi and Vojnovic (2023) for details. We set the horizon $T = 10{,}000$, and select 20 users as agents ($N = 20$) and 5 genres as arms ($K = 5$). At each time step t, each agent randomly selects a movie from the genres. All ratings (rewards) of movies are normalised to $[0, 1]$.

Experimental results. All experiments are performed with 20 independent replications. The shaded areas consider a range centered around the mean with half-width corresponding to the empirical standard deviation over 20 repetitions. In Figure 6.1, we observe that our algorithm consistently outperforms DrFed-UCB on both synthetic and real-world datasets. In all runs, after an initial exploration period, our algorithm eliminates a significant number of suboptimal actions, resulting in near-constant regret thereafter. In Figure 6.2, we observe that increasing the link probability p improves the algorithm's performance, clearly validating the regret–communication trade-off. Additionally, different base graphs significantly impact the regret under the same p value—with the complete graph yielding the lowest regret.

Chapter 7

Summary and Discussion

In Chapter 2, we study how to leverage the curvature of the loss functions in online convex optimization with delayed feedback so as to improve regret guarantees. For strongly convex functions, we derive an algorithm achieving $\mathcal{O}(\min\{\sigma_{\max} \ln T, \sqrt{d_{\text{tot}}}\})$ regret, improving upon previous work (Wan et al., 2022a, Wu et al., 2024), which only obtain $\mathcal{O}(d_{\max} \ln T)$ regret. We also derive $\mathcal{O}(\min\{d_{\max} n \ln T, \sqrt{d_{\text{tot}}}\})$ for exp-concave losses and online linear regression, answering an open question proposed in Wan et al. (2022a). It is still left open whether $\mathcal{O}(\min\{\sigma_{\max} n \ln T, \sqrt{d_{\text{tot}}}\})$ is achievable for exp-concave losses.

In Chapter 3, we study a K-armed bandit with delayed feedback and intermediate finite-state observations, where the state is observed immediately but the loss arrives after an adversarial delay. Our results show that the central factor governing regret is the mapping from states to losses: if this mapping is adversarial, the regret matches the classical delayed bandit rate $\widetilde{\mathcal{O}}\big(\sqrt{(K+d)T}\big)$, so intermediate observations bring no benefit. However, when the state–loss mapping is stochastic, the regret improves to $\widetilde{\mathcal{O}}\big(\sqrt{(K+\min\{|\mathcal{S}|, d\})T}\big)$, implying that intermediate observations help whenever the state space is smaller than the delay. We extend these bounds to non-uniform delays. The work of Vernade et al. (2020) also considers a non-stationary action-state mapping and derive regret bounds for the switching regret. Preliminary results suggest that, as long as there is an algorithm that can provide bounds on the switching regret with delayed feedback, our ideas also transfer to this setting. Unfortunately, there is currently no algorithm that can provide bounds on the switching regret with delayed feedback and we leave this as a promising direction for future work.

In Chapter 4, we investigated a variant of DOCO where agents communicate only when simultaneously active. We proposed a distributed FTRL algorithm and established an expected individual regret bound of order $(\kappa/p^{3/4})N^{1/4}\sqrt{T}$. This result is supported by a lower bound indicating that the dependence on the activation probabilities is not significantly improvable. While we considered applying block-based techniques similar to those in Wan et al. (2024a), the stochastic nature of the communication graph necessitates longer block lengths to ensure convergence, likely negating the potential benefits in this setting. Finally, two key directions for future work remain: extending the analysis to settings where active agents perform multiple local updates before communicating (Scaman et al., 2019), and developing adaptive algorithms that eliminate the need for preliminary knowledge of p or $p_{\min}$ for parameter tuning.

In Chapter 5, we study D-OCO in the presence of unknown, time- and agent-varying feedback delays. We introduce a novel algorithm that significantly improves the existing theoretical guarantees,

achieving a regret bound of $\widetilde{\mathcal{O}}\Big(\sqrt{N^3 d_{\text{tot}}} + \sqrt{\frac{N^3 T}{\sqrt{1-\sigma_2}}}\Big)$. Here, N is the number of agents, T is the time horizon, d_{tot} denotes the average total delay across agents, and $1-\sigma_2$ represents the spectral gap of the network adjacency matrix. Crucially, we also demonstrate the tightness of our result by providing a matching lower bound, confirming the necessity of the dependencies on N, T, d_{tot}, and $1-\sigma_2$. Future work could consider a setting incorporating communication delays, where signal communication across each network link is subject to a fixed bounded delay. It remains an important and non-obvious question how these communication delays fundamentally influence the performance of the learning process, specifically their impact on each agent's regret. We hypothesize that the effect of communication delays is intricately linked to the network topology, which dictates the speed at which local information propagates across the network. A rigorous study into the impact of these communication delays on D-OCO performance is a compelling direction for future research.

In Chapter 6, we study the multi-agent multi-armed bandit (MA-MAB) problem under general Erdős–Rényi random networks with heterogeneous rewards. To the best of our knowledge, we are the first to formulate MA-MAB with Erdős–Rényi random networks, where the communication graph is induced by a base graph and each edge in the base graph appears in the communication graph with probability p. This formulation generalizes the classical Erdős–Rényi model, in which the base graph is complete. We propose an algorithmic framework that incorporates the gossip communication protocol into arm elimination. Importantly, we analyze the regret bound of the algorithm and show that it improves the regret even under the classical Erdős–Rényi model. Moreover, our regret bound holds for any p, explicitly characterizes its dependency on p and the algebraic connectivity of the base graph. This naturally reveals a trade-off between regret and communication efficiency. Moving forward, while we focus extensively on the stochastic setting, it would be valuable and exciting to explore other reward models—such as contextual bandits, where rewards depend on dynamic, non-stationary contexts. In addition, achieving the optimal trade-off among regret, communication, and privacy, as previously studied in homogeneous MA-MAB settings, points out a meaningful direction for future research.

Bibliography

Jacob Abernethy, Peter L Bartlett, Alexander Rakhlin, and Ambuj Tewari. Optimal strategies and minimax lower bounds for online convex games. In *Proceedings of the 21st annual conference on learning theory*, pages 414–424, 2008.

Juliette Achddou, Nicolò Cesa-Bianchi, and Hao Qiu. Distributed online optimization with stochastic agent availability. *arXiv preprint arXiv:2411.16477*, 2024.

Mridul Agarwal, Vaneet Aggarwal, and Kamyar Azizzadenesheli. Multi-agent multi-armed bandits with limited communication. *The Journal of Machine Learning Research*, 23(1):9529–9552, 2022.

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

Mohammad Mohammadi Amiri, Deniz Gündüz, Sanjeev R Kulkarni, and H Vincent Poor. Convergence of update aware device scheduling for federated learning at the wireless edge. *IEEE Transactions on Wireless Communications*, 20(6):3643–3658, 2021.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3), 2002a.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002b.

Katy S Azoury and Manfred K Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine learning*, 43:211–246, 2001.

Yogev Bar-On and Yishay Mansour. Non-stochastic bandits with evolving observations. In Gautam Kamath and Po-Ling Loh, editors, *Proceedings of The 36th International Conference on Algorithmic Learning Theory*, volume 272 of *Proceedings of Machine Learning Research*, pages 204–227. PMLR, 2025.

Sasmita Barik, Ravindra B Bapat, and Sukanta Pati. On the laplacian spectra of product graphs. *Applicable Analysis and Discrete Mathematics*, pages 39–58, 2015.

Dimitri P Bertsekas and John N Tsitsiklis. Some aspects of parallel and distributed iterative algorithms—a survey. *Automatica*, 27(1):3–21, 1991.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations research*, 57(6):1407–1420, 2009.

Ilai Bistritz, Zhengyuan Zhou, Xi Chen, Nicholas Bambos, and Jose H. Blanchet. Online EXP3 learning in adversarial bandits with delayed feedback. In *Advances in Neural Information Processing Systems*, pages 11345–11354, 2019.

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, 2022a.

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, 2022b.

Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Gossip algorithms: Design, analysis and applications. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 3, pages 1653–1664. IEEE, 2005.

Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE transactions on information theory*, 52(6):2508–2530, 2006.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Xuanyu Cao and Tamer Basar. Decentralized online convex optimization with feedback delays. *IEEE Trans. Autom. Control.*, 67(6):2889–2904, 2022.

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

Nicolò Cesa-Bianchi, Tommaso Cesari, and Claire Monteleoni. Cooperative online learning: Keeping your neighbors updated. In *Algorithmic learning theory*, pages 234–250. PMLR, 2020.

Nicolò Cesa-Bianchi, Claudio Gentile, Yishay Mansour, and Alberto Minora. Delay and cooperation in nonstochastic bandits. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 605–622. PMLR, 2016a.

Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Delay and cooperation in nonstochastic bandits. *Journal of Machine Learning Research*, 20, 2019.

Nicol'o Cesa-Bianchi, Claudio Gentile, Yishay Mansour, and Alberto Minora. Delay and cooperation in nonstochastic bandits. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 605–622, Columbia University, New York, New York, USA, 23–26 Jun 2016b. PMLR.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):1–27, 2011.

Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, 24, 2011.

Ronshee Chawla, Abishek Sankararaman, Ayalvadi Ganesh, and Sanjay Shakkottai. The gossiping insert-eliminate algorithm for multi-agent bandits. In *International conference on artificial intelligence and statistics*, pages 3471–3481. PMLR, 2020.

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

Ashok Cutkosky. Artificial constraints and hints for unbounded online learning. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 874–894. PMLR, 2019.

John C Duchi, Alekh Agarwal, and Martin J Wainwright. Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Transactions on Automatic control*, 57(3):592–606, 2011.

Hubert Eichner, Tomer Koren, Brendan McMahan, Nathan Srebro, and Kunal Talwar. Semi-cyclic stochastic gradient descent. In *International Conference on Machine Learning*, pages 1764–1773. PMLR, 2019.

Stephen G. Eick. The two-armed bandit with delayed responses. *The Annals of Statistics*, 1988.

Paul Erdos, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. math. inst. hung. acad. sci*, 5(1):17–60, 1960.

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.

Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.

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.

Pierre Gaillard, Yannig Goude, and Raphaël Nedellec. Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting. *International Journal of forecasting*, 32(3):1038–1050, 2016.

Deepak Ganesan, Alberto Cerpa, Wei Ye, Yan Yu, Jerry Zhao, and Deborah Estrin. Networking issues in wireless sensor networks. *Journal of parallel and distributed computing*, 64(7):799–814, 2004.

Khashayar Gatmiry and Jon Schneider. Adversarial online learning with temporal feedback graphs. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 4548–4572. PMLR, 2024.

Saeed Ghoorchian and Setareh Maghsudi. Multi-armed bandit for energy-efficient and delay-sensitive edge computing in dynamic networks with uncertainty. *IEEE Transactions on Cognitive Communications and Networking*, 7(1):279–293, 2020.

Xinran Gu, Kaixuan Huang, Jingzhao Zhang, and Longbo Huang. Fast federated learning in the presence of arbitrary device unavailability. *Advances in Neural Information Processing Systems*, 34:12052–12064, 2021.

András György and Pooria Joulani. Adapting to delays and data in adversarial multi-armed bandits. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, 2021.

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

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

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

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.

Dirk van der Hoeven and Nicolò Cesa-Bianchi. Nonstochastic bandits and experts with arm-dependent delays. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 2022–2044. PMLR, 28–30 Mar 2022a.

Dirk van der Hoeven and Nicolò Cesa-Bianchi. Nonstochastic bandits and experts with arm-dependent delays. In *International Conference on Artificial Intelligence and Statistics*, 2022b.

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 Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 1285–1321. PMLR, 12–15 Jul 2023a.

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 *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 1285–1321. PMLR, 2023b.

Derek Allan Holton and John Sheehan. *The petersen graph*, volume 7. Cambridge University Press, 1993.

Saghar Hosseini, Airlie Chapman, and Mehran Mesbahi. Online distributed optimization via dual averaging. In *52nd IEEE Conference on Decision and Control*, pages 1484–1489. IEEE, 2013.

Saghar Hosseini, Airlie Chapman, and Mehran Mesbahi. Online distributed convex optimization on dynamic networks. *IEEE Transactions on Automatic Control*, 61(11):3545–3550, 2016.

Tiancheng Jin, Tal Lancewicki, Haipeng Luo, Yishay Mansour, and Aviv Rosenberg. Near-optimal regret for adversarial mdp with delayed bandit feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 33469–33481. Curran Associates, Inc., 2022.

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

Pooria Joulani, András György, and Csaba Szepesvári. Delay-tolerant online convex optimization: Unified analysis and adaptive-gradient algorithms. In Dale Schuurmans and Michael P. Wellman, editors, *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 1744–1750. AAAI Press, 2016a. doi: 10.1609/AAAI.V30I1.10320. URL `https://doi.org/10.1609/aaai.v30i1.10320`.

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, 2016b.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605. IEEE, 2003.

Tal Lancewicki, Shahar Segal, Tomer Koren, and Yishay Mansour. Stochastic multi-armed bandits with unrestricted delay distributions. In *International Conference on Machine Learning*, pages 5969–5978. PMLR, 2021.

Tal Lancewicki, Aviv Rosenberg, and Yishay Mansour. Learning adversarial markov decision processes with delayed feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7281–7289, 2022.

Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. Distributed cooperative decision-making in multiarmed bandits: Frequentist and bayesian algorithms. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 167–172. IEEE, 2016a.

Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. On distributed cooperative decision-making in multiarmed bandits. In *2016 European Control Conference (ECC)*, pages 243–248. IEEE, 2016b.

John Langford, Alexander Smola, and Martin Zinkevich. Slow learners are fast. *arXiv preprint arXiv:0911.0491*, 2009.

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

Jinlong Lei, Peng Yi, Yiguang Hong, Jie Chen, and Guodong Shi. Online convex optimization over Erdos-Rényi random networks. *Advances in neural information processing systems*, 33:15591–15601, 2020.

Dan Li, Kerry D Wong, Yu Hen Hu, and Akbar M Sayeed. Detection, classification, and tracking of targets. *IEEE signal processing magazine*, 19(2):17–29, 2002.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

Xiuxian Li, Lihua Xie, and Na Li. A survey on distributed online optimization and online games. *Annual Reviews in Control*, 56:100904, 2023.

Nick Littlestone. *Mistake bounds and logarithmic linear-threshold learning algorithms*. PhD thesis, University of California, Santa Cruz, 1990. UMI Order No: GAX89-26506.

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

Jingyuan Liu, Hao Qiu, Lin Yang, and Mengfan Xu. Distributed multi-agent bandits over Erdős–Rényi random networks. *preprint*, 2025.

Timothy Arthur Mann, Sven Gowal, András György, Huiyi Hu, Ray Jiang, Balaji Lakshminarayanan, and Prav Srinivasan. Learning from delayed outcomes via proxies with applications to recommender systems. In *International Conference on Machine Learning*, pages 4324–4332, 2019.

Shuai Mao, Wei Du, Yu-Chu Tian, Juping Gu, and Yang Tang. Online distributed convex optimization for unbalanced varying graphs with delayed feedback. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2025.

David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

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*, 2022a.

Saeed Masoudian, Julian Zimmert, and Yevgeny Seldin. A best-of-both-worlds algorithm for bandits with delayed feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 11752–11762. Curran Associates, Inc., 2022b.

Saeed Masoudian, Julian Zimmert, and Yevgeny Seldin. A best-of-both-worlds algorithm for bandits with delayed feedback with robustness to excessive delays. *arXiv preprint*, arXiv:2308.10675, 2023. URL `https://arxiv.org/abs/2308.10675`.

Saeed Masoudian, Julian Zimmert, and Yevgeny Seldin. A best-of-both-worlds algorithm for bandits with delayed feedback with robustness to excessive delays. *arXiv preprint*, arXiv:2308.10675, 2024a. URL `https://arxiv.org/abs/2308.10675`.

Saeed Masoudian, Julian Zimmert, and Yevgeny Seldin. A best-of-both-worlds algorithm for bandits with delayed feedback with robustness to excessive delays. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b.

David Mateos-Núnez and Jorge Cortés. Distributed online convex optimization over jointly connected digraphs. *IEEE Transactions on Network Science and Engineering*, 1(1):23–37, 2014.

Jack J. Mayo, Hedi Hadiji, and Tim van Erven. Scale-free unconstrained online learning for curved losses. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 4464–4497. PMLR, 2022.

Brendan McMahan and Matthew Streeter. Delay-tolerant algorithms for asynchronous distributed online learning. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014a.

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

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

Angelia Nedić and Alex Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 60(3):601–615, 2014.

Angelia Nedic and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009a. doi: 10.1109/TAC.2008.2009515.

Angelia Nedic and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009b.

Angelia Nedic, Asuman Ozdaglar, and Pablo A Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.

Gergely Neu. Explore no more: Improved high-probability regret bounds for non-stochastic bandits. In *Advances in Neural Information Processing Systems*, 2015.

Gergely Neu, András György, Csaba Szepesvári, and András Antos. Online markov decision processes under bandit feedback. In *Advances in Neural Information Processing Systems*, 2010.

Gergely Neu, András György, Csaba Szepesvári, and András Antos. Online markov decision processes under bandit feedback. *IEEE Transactions on Automatic Control*, 59:676–691, 2014.

Tuan-Anh Nguyen, Nguyen Kim Thang, and Denis Trystram. Handling delayed feedback in distributed online optimization: A projection-free approach. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 197–211. Springer, 2024.

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

Francesco Orabona. A modern introduction to online learning. *arXiv preprint*, arXiv:1912.13213, 2025. URL `https://arxiv.org/abs/1912.13213`.

Ciara Pike-Burke, Shipra Agrawal, Csaba Szepesvari, and Steffen Grunewalder. Bandits with delayed, aggregated anonymous feedback. In *International Conference on Machine Learning*, pages 4105–4113. PMLR, 2018.

Hao Qiu, Emmanuel Esposito, and Mengxiao Zhang. Exploiting curvature in online convex optimization with delayed feedback. In *Forty-second International Conference on Machine Learning*, 2025a.

Hao Qiu, Mengxiao Zhang, and Juliette Achddou. Decentralized online convex optimization with unknown feedback delays. *preprint*, 2025b.

Kent Quanrud and Daniel Khashabi. Online learning with adversarial delays. *Advances in neural information processing systems*, 28, 2015.

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

Maxim Raginsky, Nooshin Kiarashi, and Rebecca Willett. Decentralized online convex programming with local information. In *Proceedings of the 2011 American Control Conference*, pages 5363–5369. IEEE, 2011.

Alexander Ryabchenko, Idan Attias, and Daniel M. Roy. Capacity-constrained online learning with delays: Scheduling frameworks and regret trade-offs. *arXiv preprint arXiv:2503.19856v1*, 2025.

Abishek Sankararaman, Ayalvadi Ganesh, and Sanjay Shakkottai. Social learning in multi agent multi armed bandits. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3 (3):1–35, 2019.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal convergence rates for convex distributed optimization in networks. *Journal of Machine Learning Research*, 20(159):1–31, 2019. URL `http://jmlr.org/papers/v20/19-543.html`.

Ofir Schlisselberg, Ido Cohen, Tal Lancewicki, and Yishay Mansour. Delay as payoff in MAB. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(19):20310–20317, 2025.

Steven L Scott. Multi-armed bandit experiments in the online service economy. *Applied Stochastic Models in Business and Industry*, 31(1):37–45, 2015.

Shahin Shahrampour and Ali Jadbabaie. Distributed online optimization in dynamic environments using mirror descent. *IEEE Transactions on Automatic Control*, 63(3):714–725, 2018. doi: 10.1109/TAC.2017.2743462.

Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Ola Shorinwa, Trevor Halsted, Javier Yu, and Mac Schwager. Distributed optimization methods for multi-robot systems: Part 1—a tutorial [tutorial]. *IEEE Robotics & Automation Magazine*, 31(3): 121–138, 2024.

Richard Simon. Adaptive treatment assignment methods and clinical trials. *Biometrics*, 33, 1977.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2), 2019.

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

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25, 1933.

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.

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

Dirk Van Der Hoeven, Ciara Pike-Burke, Hao Qiu, and Nicolò Cesa-Bianchi. Trading-off payments and accuracy in online classification with paid stochastic experts. In *International Conference on Machine Learning*, pages 34809–34830. PMLR, 2023.

Claire Vernade, András György, and Timothy A. Mann. Non-stationary delayed bandits with intermediate observations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, 2020.

Volodimir G Vovk. Aggregating strategies. In *Proceedings of the third annual workshop on Computational learning theory*, pages 371–386, 1990.

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

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

Yuanyu Wan, Wei-Wei Tu, and Lijun Zhang. Online frank-wolfe with arbitrary delays. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 19703–19715. Curran Associates, Inc., 2022b.

Yuanyu Wan, Yibo Wang, Chang Yao, Wei-Wei Tu, and Lijun Zhang. Projection-free online learning with arbitrary delays, 2023.

Yuanyu Wan, Tong Wei, Mingli Song, and Lijun Zhang. Nearly optimal regret for decentralized online convex optimization. In *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 4862–4888. PMLR, 30 Jun–03 Jul 2024a.

Yuanyu Wan, Tong Wei, Mingli Song, and Lijun Zhang. Nearly optimal regret for decentralized online convex optimization. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 4862–4888. PMLR, 2024b.

Yuanyu Wan, Tong Wei, Bo Xue, Mingli Song, and Lijun Zhang. Optimal and efficient algorithms for decentralized online convex optimization, 2024c. URL https://arxiv.org/abs/2402.09173.

Yuanyu Wan, Chang Yao, Mingli Song, and Lijun Zhang. Improved regret for bandit convex optimization with delayed feedback. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 169–196. Curran Associates, Inc., 2024d.

Po-An Wang, Alexandre Proutiere, Kaito Ariu, Yassir Jedra, and Alessio Russo. Optimal algorithms for multiplayer multi-armed bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 4120–4129. PMLR, 2020.

Shiqiang Wang and Mingyue Ji. A unified analysis of federated learning with arbitrary client participation. *Advances in Neural Information Processing Systems*, 35:19124–19137, 2022.

Xiong Wang, Jiancheng Ye, and John CS Lui. Decentralized task offloading in edge computing: A multi-user multi-armed bandit approach. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 1199–1208. IEEE, 2022a.

Xuchuang Wang, Lin Yang, Yu-Zhen Janice Chen, Xutong Liu, Mohammad Hajiesmaili, Don Towsley, and John CS Lui. Achieving near-optimal individual regret & low communications in multi-agent bandits. In *The Eleventh International Conference on Learning Representations*, 2022b.

Yu Wang, Shiwen Mao, and R Mark Nelms. Distributed online algorithm for optimal real-time energy distribution in the smart grid. *IEEE Internet of Things Journal*, 1(1):70–80, 2014.

Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *International Conference on Learning Representations*, 2021.

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

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

Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004a.

Lin Xiao and Stephen P. Boyd. Fast linear iterations for distributed averaging. *Syst. Control. Lett.*, 53(1):65–78, 2004b. doi: 10.1016/J.SYSCONLE.2004.02.022. URL https://doi.org/10.1016/j.sysconle.2004.02.022.

Menghui Xiong, Daniel WC Ho, Baoyong Zhang, Deming Yuan, and Shengyuan Xu. Distributed online mirror descent with delayed subgradient and event-triggered communications. *IEEE Transactions on Network Science and Engineering*, 11(2):1702–1715, 2023a.

Menghui Xiong, Baoyong Zhang, Deming Yuan, Yijun Zhang, and Jun Chen. Event-triggered distributed online convex optimization with delayed bandit feedback. *Applied Mathematics and Computation*, 445:127865, 2023b.

Mengfan Xu and Diego Klabjan. Decentralized randomly distributed multi-agent multi-armed bandit with heterogeneous rewards. *Advances in Neural Information Processing Systems*, 36:74799–74855, 2023.

Mengfan Xu and Diego Klabjan. Multi-agent multi-armed bandit regret complexity and optimality. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

Yunbei Xu and Assaf Zeevi. Bayesian design principles for frequentist sequential learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 38768–38800. PMLR, 23–29 Jul 2023.

Feng Yan, Shreyas Sundaram, SVN Vishwanathan, and Yuan Qi. Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2483–2493, 2013.

Yikai Yan, Chaoyue Niu, Yucheng Ding, Zhenzhe Zheng, Shaojie Tang, Qinya Li, Fan Wu, Chengfei Lyu, Yanghe Feng, and Guihai Chen. Federated optimization under intermittent client availability. *INFORMS Journal on Computing*, 36(1):185–202, 2024.

Tao Yang, Xinlei Yi, Junfeng Wu, Ye Yuan, Di Wu, Ziyang Meng, Yiguang Hong, Hong Wang, Zongli Lin, and Karl H Johansson. A survey of distributed optimization. *Annual Reviews in Control*, 47:278–305, 2019.

Haishan Ye, Luo Luo, Ziang Zhou, and Tong Zhang. Multi-consensus decentralized accelerated gradient descent. *Journal of machine learning research*, 24(306):1–50, 2023a.

Mang Ye, Xiuwen Fang, Bo Du, Pong C Yuen, and Dacheng Tao. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, 56(3):1–44, 2023b.

Jialin Yi and Milan Vojnovic. Doubly adversarial federated bandits. In *International Conference on Machine Learning*, pages 39951–39967. PMLR, 2023.

Deming Yuan, Alexandre Proutiere, and Guodong Shi. Distributed online linear regressions. *IEEE Transactions on Information Theory*, 67(1):616–639, 2020.

Deming Yuan, Alexandre Proutiere, and Guodong Shi. Distributed online optimization with long-term constraints. *IEEE Transactions on Automatic Control*, 67(3):1089–1104, 2021.

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

Mengxiao Zhang, Yingfei Wang, and Haipeng Luo. Contextual linear bandits with delay as payoff. *arXiv preprint arXiv:2502.12528v2*, 2025.

Jiang Zhu, Yonghui Song, Dingde Jiang, and Houbing Song. Multi-armed bandit channel access scheme with cognitive radio technology in wireless sensor networks for the internet of things. *IEEE access*, 4:4609–4617, 2016.

Jingxuan Zhu and Ji Liu. Distributed multiarmed bandits. *IEEE Transactions on Automatic Control*, 68(5):3025–3040, 2023.

Jingxuan Zhu, Romeil Sandhu, and Ji Liu. A distributed algorithm for sequential decision making in multi-armed bandit with homogeneous rewards. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3078–3083. IEEE, 2020.

Jingxuan Zhu, Ethan Mulle, Christopher S Smith, Alec Koppel, and Ji Liu. Decentralized upper confidence bound algorithms for homogeneous multi-agent multi-armed bandits. *IEEE Transactions on Automatic Control*, 2025.

Zhaowei Zhu, Jingxuan Zhu, Ji Liu, and Yang Liu. Federated bandit: A gossiping approach. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(1):1–29, 2021.

Julian Zimmert and Yevgeny Seldin. An optimal algorithm for adversarial bandits with arbitrary delays. In *Proceedings on the International Conference on Artificial Intelligence and Statistics*, 2020a.

Julian Zimmert and Yevgeny Seldin. An optimal algorithm for adversarial bandits with arbitrary delays. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 3285–3294. PMLR, 2020b.

Julian Zimmert and Yevgeny Seldin. Tsallis-INF: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(1):1310–1358, 2021.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936, 2003.

Appendix A

Proof Details for Chapter 2

A.1 Auxiliary results

In this section, we show several auxiliary lemmas that will be helpful throughout the paper.

A.1.1 General results for the regret analysis

The following lemma is a standard result for the regret of FTRL.

Lemma A.1 (Orabona (2025, Lemma 7.1)). *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed and non-empty. Denote by $F_t(x) = \psi_t(x) + \sum_{\tau=1}^{t-1} \ell_\tau(x)$. Assume that $\arg\min_{x\in\mathcal{X}} F_t(x)$ is not empty and $x_t \in \arg\min_{x\in\mathcal{X}} F_t(x)$. Then, for any $u \in \mathcal{X}$,*

$$\sum_{t=1}^{T} \big(\ell_t(x_t) - \ell_t(u)\big) = \psi_{T+1}(u) - \min_{x\in\mathcal{X}} \psi_1(x) + \sum_{t=1}^{T} \big[F_t(x_t) - F_{t+1}(x_{t+1}) + \ell_t(x_t)\big] + F_{T+1}(x_{T+1}) - F_{T+1}(u) \; .$$

The next lemma bounds the distance between two FTRL iterates with different linear losses and possibly different regularizers. It also shows a simplified upper bound in the case when the two iterates have the same regularizer.

Lemma A.2 (Stability lemma). *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed and non-empty. Let $A_1, A_2 \succeq 0$ be two positive semidefinite matrices, $b_1, b_2 \in \mathbb{R}^n$, and $c_1, c_2 \in \mathbb{R}$. Define $\psi_1(x) = x^\top A_1 x + b_1^\top x + c_1$ and $\psi_2(x) = x^\top A_2 x + b_2^\top x + c_2$. Suppose that $z_1 \in \arg\min_{x\in\mathcal{X}} \big\{\langle w_1, x\rangle + \psi_1(x)\big\}$ and $z_2 \in \arg\min_{x\in\mathcal{X}} \big\{\langle w_2, x\rangle + \psi_2(x)\big\}$. Then, we have*

$$\|z_1 - z_2\|_{A_1}^2 + \|z_1 - z_2\|_{A_2}^2 \leq \langle w_1 - w_2, z_2 - z_1\rangle + (\psi_1(z_2) - \psi_2(z_2)) - (\psi_1(z_1) - \psi_2(z_1)) \; .$$

Furthermore, if $\psi_1(x) = \psi_2(x) = x^\top A x + b^\top x + c$ with positive definite $A \succ 0$, we have

$$\|z_1 - z_2\|_A \leq \frac{1}{2}\|w_1 - w_2\|_{A^{-1}} \; .$$

Proof. Let $h_1(x) = \langle w_1, x\rangle + \psi_1(x)$ and $h_2(x) = \langle w_2, x\rangle + \psi_2(x)$ be twice-differentiable functions with Hessians $A_1 + A_1^\top$ and $A_2 + A_2^\top$, respectively. Note that $z_1 \in \arg\min_{x\in\mathcal{X}} h_1(x)$ and $z_2 \in \arg\min_{x\in\mathcal{X}} h_2(x)$. By Taylor's theorem and first-order optimality conditions, we know that

$$(\langle w_1, z_2\rangle + \psi_1(z_2)) - (\langle w_1, z_1\rangle + \psi_1(z_1)) = h_1(z_2) - h_1(z_1) \geq \|z_1 - z_2\|_{A_1}^2 \; ,$$

$$(\langle w_2, z_1\rangle + \psi_2(z_1)) - (\langle w_2, z_2\rangle + \psi_2(z_2)) = h_2(z_1) - h_2(z_2) \geq \|z_1 - z_2\|_{A_2}^2 .$$

Summing up the above two inequalities, we obtain

$$\|z_1 - z_2\|_{A_1}^2 + \|z_1 - z_2\|_{A_2}^2 \leq \langle w_1 - w_2, z_2 - z_1\rangle + (\psi_1(z_2) - \psi_2(z_2)) - (\psi_1(z_1) - \psi_2(z_1)) .$$

The second result is directly obtained by applying the Cauchy-Schwarz inequality when $\psi_1(x) = \psi_2(x)$. □

The following lemma is the quadratic bound of α-exp-concave functions.

Lemma A.3 (Hazan et al. (2007, Lemma 2)). *Let $\ell\colon \mathcal{X} \to \mathbb{R}$ be an α-exp-concave function. Then, under Assumption 2.1 and Assumption 2.2, we have that*

$$\ell(x) \geq \ell(y) + \langle \nabla\ell(y), x - y\rangle + \frac{\beta}{2}\big(\langle \nabla\ell(y), x - y\rangle\big)^2$$

for any $x, y \in \mathcal{X}$, where $\beta = \frac{1}{2}\min\{\frac{1}{4LD}, \alpha\}$.

The following lemma is the link of the Bregman divergences between 3 points.

Lemma A.4 (Wei et al. (2021, Lemma 10)). *Let $\mathcal{A}$ be a convex set and $x_2 = \arg\min_{x\in\mathcal{A}}\{\langle g, x\rangle + D_\psi(x, x_1)\}$. Then, for any $u \in \mathcal{A}$,*

$$\langle x_2 - u, g\rangle \leq D_\psi(u, x_1) - D_\psi(u, x_2) - D_\psi(x_2, x_1) .$$

The following lemma is the general bound on $\langle g, v\rangle - \frac{\lambda}{2}\|v\|^2$, which related to the one achievable via the Fenchel-Young inequality but strengthened thanks to a norm constraint on v.

Lemma A.5 (Flaspohler et al. (2021, Lemma 18)). *Let $\|\cdot\|$ be a norm over $\mathbb{R}^n$ and let $\|\cdot\|_*$ be its dual norm. For any constants $\lambda, c, b > 0$ and any $g \in \mathbb{R}^n$,*

$$\sup_{v\in\mathbb{R}^n:\|v\|\leq\min\{\frac{c}{\lambda}, b\}}\left(\langle g, v\rangle - \frac{\lambda}{2}\|v\|^2\right) \leq \min\left\{\frac{1}{2\lambda}\|g\|_*^2,\ \frac{c}{\lambda}\|g\|_*,\ b\|g\|_*\right\} .$$

A.1.2 Results for delay-related quantities

The following three lemmas quantify the relationship between $\sigma_{\max}$, $d_{\max}$, and d_{tot}.

Lemma A.6 (Masoudian et al. (2022b, Lemma 3)). *Let $d_{\max}(S) = \max_{\tau\in S} d_\tau$ and $\bar{S} = [T] \setminus S$ for any $S \subseteq [T]$. Then,*

$$\sigma_{\max} \leq \min_{S\subseteq[T]}\big(|S| + d_{\max}(\bar{S})\big) .$$

Lemma A.7. *Let $d_{\text{tot}}(S) = \sum_{\tau\in S} d_\tau$ and $\bar{S} = [T] \setminus S$ for any $S \subseteq [T]$. Then,*

$$\sigma_{\max} \leq 2\sqrt{2}\min_{S\subseteq[T]}\left(|S| + \sqrt{d_{\text{tot}}(\bar{S})}\right) .$$

Proof. First, observe that $d_{\text{tot}}(S) = \sum_{t=1}^T |m_t \cap S|$ for any $S \subseteq [T]$. Also note that the bound trivially holds if $\sigma_{\max} = 0$; hence, assume $\sigma_{\max} \geq 1$ without loss of generality. Let t^* be any round

such that $|m_{t^*}| = \sigma_{\max}$. Consider any $S \subseteq [T]$, and define $A = m_{t^*} \cap S$ and $B = m_{t^*} \cap \bar{S}$. If $|A| \geq (\sqrt{2}-1)|m_{t^*}|$, then

$$|S| + \sqrt{d_{\text{tot}}(\bar{S})} \geq |S| \geq |A| \geq (\sqrt{2}-1)\sigma_{\max} .$$

Otherwise, we have that $|B| > (2-\sqrt{2})|m_{t^*}|$. Hence, denote $B = \{t_1, \dots, t_{|B|}\}$ such that $t_1 < \dots < t_{|B|}$ and observe that $|m_{t_i+1} \cap B| \geq i$ for any $t_i \in B$. We can consequently prove that

$$|S| + \sqrt{d_{\text{tot}}(\bar{S})} \geq \sqrt{d_{\text{tot}}(\bar{S})} = \sqrt{\sum_{t=1}^{T} |m_t \cap \bar{S}|} \geq \sqrt{\sum_{t \in B} |m_{t+1} \cap B|} \geq \sqrt{\sum_{i=1}^{|B|} i} \geq \frac{|B|}{\sqrt{2}} > (\sqrt{2}-1)\sigma_{\max} ,$$

which concludes the proof as $\frac{1}{\sqrt{2}-1} \leq 2\sqrt{2}$. □

Lemma A.8. *Let $\sigma^S_{\max} = \max_{\tau \in [T]} |m_\tau \cap S|$ and $\bar{S} = [T] \setminus S$ for any $S \subseteq [T]$. Then,*

$$\sigma_{\max} = \min_{S \subseteq [T]} \left(|S| + \sigma^{\bar{S}}_{\max} \right) .$$

Proof. First, it trivially holds that

$$\sigma_{\max} \geq \min_{S \subseteq [T]} \left(|S| + \sigma^{\bar{S}}_{\max} \right) .$$

We now only need to show the inequality in the other direction. Consider any $S \subseteq [T]$ and let t^* be any round such that $|m_{t^*}| = \sigma_{\max}$. Then,

$$|S| + \sigma^{\bar{S}}_{\max} \geq |S| + |m_{t^*} \cap \bar{S}| = |S| + |m_{t^*} \setminus S| \geq |m_{t^*}| = \sigma_{\max} ,$$

which concludes the proof. □

The following lemma further illustrates the relationship between $\sigma_{\max}$ and $\sqrt{d_{\text{tot}}}$ in a more concrete way.

Lemma A.9. *There exists a delay sequence $(d_t)_{t \in [T]}$ such that $\sigma_{\max} \geq \sqrt{1.5 \cdot d_{\text{tot}}}$. In addition, there also exists a delay sequence such that $\sigma_{\max} = 1$ and $\sqrt{d_{\text{tot}}} = \sqrt{T}$.*

Proof. Given a positive integer $N > 5$, consider the sequence $(d_t)_{t \in [T]}$, where $d_t = N - t$ for all $t \leq N$ and $d_t = 0$ for all $t > N$. In this case, $\sigma_{\max} = \sigma_{N-1} = N - 1$ and $\sqrt{1.5 \cdot d_{\text{tot}}} = \sqrt{\frac{3N(N-1)}{4}} \leq N - 1$. On the other hand, consider the sequence where $d_t = 1$ for all $t \in [T]$. In this case, $\sigma_{\max} = 1$ and $\sqrt{d_{\text{tot}}} = \sqrt{T}$. □

On a similar note, we show another similar result depicting the relationship between $d_{\max}$ and $\sqrt{d_{\text{tot}}}$.

Lemma A.10. *There exists a delay sequence $(d_t)_{t \in [T]}$ such that $d_{\max} = T$ and $\sqrt{d_{\text{tot}}} = \sqrt{T}$. In addition, there also exists a delay sequence such that $d_{\max} = 1$ and $\sqrt{d_{\text{tot}}} = \sqrt{T}$.*

Proof. Consider the sequence $(d_t)_{t \in [T]}$ where one round $t_0 \leq T/2$ with $d_{t_0} = T - t_0$ and all the other rounds $d_t = 0$ for $t \neq t_0$, then we can choose $t_0 = 1$ and have $d_{\max} = T$ and $\sqrt{d_{\text{tot}}} = \sqrt{T}$. On the other hand, consider the sequence where $d_t = 1$ for all $t \in [T]$, then $d_{\max} = 1$ and $\sqrt{d_{\text{tot}}} = \sqrt{T}$. □

A.2 Omitted details in Section 2.3

In this section, we show the omitted details in Section 2.3. For completeness, we restate the theorem and provide its proof.

Theorem 2.1. *Assume that $\ell_1, \dots, \ell_T$ are λ-strongly convex with respect to the Euclidean norm $\|\cdot\|_2$. Then, under Assumption 2.1, Algorithm 2.1 guarantees that*

$$\mathrm{Reg}_T = \mathcal{O}\left(\frac{L^2}{\lambda}\left(\ln T + \min\left\{\sigma_{\max}\ln T, \sqrt{d_{\mathrm{tot}}}\right\}\right)\right).$$

Proof. First of all, define

$$F_t(x) = \sum_{\tau\in o_t}\langle g_\tau, x\rangle + \frac{\lambda}{2}\sum_{\tau=1}^{t-1}\|x - x_\tau\|_2^2 \qquad \text{and} \qquad F_t^\star(x) = \sum_{\tau=1}^{t-1}\left(\langle g_\tau, x\rangle + \frac{\lambda}{2}\|x - x_\tau\|_2^2\right)$$

for any $t \geq 1$. Observe that $x_t \in \arg\min_{x\in\mathcal{X}} F_t(x)$ and additionally define $x_t^\star \in \arg\min_{x\in\mathcal{X}} F_t^\star(x)$ for $t \geq 2$, while $x_1^\star = x_1$ (since $F_1^\star(x) = F_1(x)$). The sequence $(x_t^\star)_{t\geq 1}$ represents the "cheating" sequence that uses the gradients from all rounds up to $t-1$, including those from rounds in m_t that are yet to be received because of the delays. As mentioned in Section 2.3, we decompose the regret as follows:

$$\begin{aligned}\mathrm{Reg}_T(u) = \sum_{t=1}^T\big(\ell_t(x_t) - \ell_t(u)\big) &\leq \sum_{t=1}^T\left(\langle g_t, x_t - u\rangle - \frac{\lambda}{2}\|x_t - u\|_2^2\right)\\ &= \underbrace{\sum_{t=1}^T\langle g_t, x_t^\star - u\rangle}_{\mathrm{Reg}_T^\star(u)} + \underbrace{\sum_{t=1}^T\langle g_t, x_t - x_t^\star\rangle}_{\texttt{Drift}_T} - \frac{\lambda}{2}\sum_{t=1}^T\|x_t - u\|_2^2\,, \end{aligned} \tag{A.1}$$

where the first inequality follows from the λ-strong convexity of ℓ_t. Next, we analyze the cheating term $\mathrm{Reg}_T^\star(u)$ and the drift term $\texttt{Drift}_T$ individually, and their respective upper bounds will then be combined to derive the final regret bound.

To analyze $\mathrm{Reg}_T^\star(u)$, first define $\psi_t(x) = \frac{\lambda}{2}\sum_{\tau=1}^{t-1}\|x - x_\tau\|_2^2$ for $t \geq 1$. We can therefore rewrite both $F_t(x) = \sum_{\tau\in o_t}\langle g_\tau, x\rangle + \psi_t(x)$ and $F_t^\star(x) = \sum_{\tau=1}^{t-1}\langle g_\tau, x\rangle + \psi_t(x)$. Hence, applying Lemma A.1, we can bound $\mathrm{Reg}_T^\star(u)$ as follows:

$$\begin{aligned}\mathrm{Reg}_T^\star(u) &= \sum_{t=1}^T\langle g_t, x_t^\star - u\rangle\\ &= \psi_{T+1}(u) - \min_{x\in\mathcal{X}}\psi_1(x) + \sum_{t=1}^T\left[F_t^\star\big(x_t^\star\big) - F_{t+1}^\star\big(x_{t+1}^\star\big) + \langle g_t, x_t^\star\rangle\right] + F_{T+1}^\star\big(x_{T+1}^\star\big) - F_{T+1}^\star(u)\\ &\leq \psi_{T+1}(u) + \sum_{t=1}^T\Big[\big(F_t^\star(x_t^\star) + \langle g_t, x_t^\star\rangle\big) - \big(F_t^\star(x_{t+1}^\star) + \langle g_t, x_{t+1}^\star\rangle\big) - \psi_{t+1}(x_{t+1}^\star) + \psi_t(x_{t+1}^\star)\Big]\,,\end{aligned} \tag{A.2}$$

where the last inequality holds because $F_{T+1}^\star(x_{T+1}^\star) \leq F_{T+1}^\star(u)$ by optimality of $x_{T+1}^\star$, together with the non-negativity of ψ_1.

Focus on the difference between the terms $F_t^\star(x_t^\star) + \langle g_t, x_t^\star \rangle$ and $F_t^\star(x_{t+1}^\star) + \langle g_t, x_{t+1}^\star \rangle$ within the sum in the right-hand side of Equation (A.2). Applying Lemma A.2 for $z_1 = x_{t+1}^\star$ with $A_1 = \frac{\lambda t}{2} I$ and $w_1 = \sum_{\tau \le t} g_\tau$, and $z_2 = x_t^\star$ with $A_2 = \frac{\lambda(t-1)}{2} I$ and $w_2 = \sum_{\tau \le t-1} g_\tau$, we have that

$$\begin{aligned}(2t-1)\frac{\lambda}{2}\|x_t^\star - x_{t+1}^\star\|_2^2 &= \|x_t^\star - x_{t+1}^\star\|_{A_1}^2 + \|x_t^\star - x_{t+1}^\star\|_{A_2}^2 \\ &\le \langle g_t, x_t^\star - x_{t+1}^\star \rangle + \frac{\lambda}{2}\|x_t^\star - x_t\|_2^2 - \frac{\lambda}{2}\|x_{t+1}^\star - x_t\|_2^2 \\ &\le \|g_t\|_2 \|x_t^\star - x_{t+1}^\star\|_2 + \frac{\lambda}{2}\|x_t^\star - x_t\|_2^2 \,,\end{aligned}$$

where we used the Cauchy-Schwarz inequality in the last step. By straightforward calculations, we can show that the above inequality implies that

$$\|x_t^\star - x_{t+1}^\star\|_2 \le \frac{2\|g_t\|_2}{\lambda(2t-1)} + \frac{\|x_t^\star - x_t\|_2}{\sqrt{2t-1}} \le \frac{2\|g_t\|_2}{\lambda(2t-1)} + \|x_t^\star - x_t\|_2 \,. \tag{A.3}$$

We can leverage this inequality to show that

$$\begin{aligned}\big(F_t^\star(x_t^\star) + \langle g_t, x_t^\star \rangle\big) - \big(F_t^\star(x_{t+1}^\star) + \langle g_t, x_{t+1}^\star \rangle\big) &\le \langle g_t, x_t^\star - x_{t+1}^\star \rangle && (F_t^\star(x_t^\star) \le F_{t+1}^\star(x_{t+1}^\star)) \\ &\le \|g_t\|_2 \|x_t^\star - x_{t+1}^\star\|_2 && \text{(Cauchy-Schwarz)} \\ &\le \frac{2\|g_t\|_2^2}{\lambda(2t-1)} + \|g_t\|_2 \|x_t^\star - x_t\|_2 \,, && \text{(Equation (A.3))}\end{aligned}$$

where the first inequality is due to the optimality of $x_t^\star$ with respect to $F_t^\star$. Plugging the above into the bound on $\mathrm{Reg}_T^\star(u)$ from Equation (A.2), we obtain

$$\begin{aligned}\mathrm{Reg}_T^\star(u) &\le \psi_{T+1}(u) + \sum_{t=1}^T \left[\frac{2\|g_t\|_2^2}{\lambda(2t-1)} + \|g_t\|_2\|x_t^\star - x_t\|_2 + \psi_t(x_{t+1}^\star) - \psi_{t+1}(x_{t+1}^\star)\right] \\ &= \frac{\lambda}{2}\sum_{t=1}^T \|x_t - u\|_2^2 + \sum_{t=1}^T \left[\frac{2\|g_t\|_2^2}{\lambda(2t-1)} + \|g_t\|_2\|x_t^\star - x_t\|_2 - \frac{\lambda}{2}\|x_{t+1}^\star - x_t\|_2^2\right] \\ &\le \frac{\lambda}{2}\sum_{t=1}^T \|x_t - u\|_2^2 + \frac{L^2}{\lambda}\sum_{t=1}^T \frac{2}{2t-1} + L\sum_{t=1}^T \|x_t^\star - x_t\|_2 \\ &\le \frac{\lambda}{2}\sum_{t=1}^T \|x_t - u\|_2^2 + \frac{L^2}{\lambda}\ln(2T+1) + L\sum_{t=1}^T \|x_t^\star - x_t\|_2 \,,\end{aligned} \tag{A.4}$$

where the equality is due to the definition of ψ_t, while the second inequality follows from $\|g_t\|_2 \le L$ by Assumption 2.1.

Observe that, given such a bound on the cheating term, we now have to consider three different terms as shown in Equation (A.4). While the second one is a desirable logarithmic term, and the first one is negligible since it will be canceled when plugging this bound on $\mathrm{Reg}_T^\star(u)$ into Equation (A.1), the third one needs some further analysis. Interestingly enough, this latter term involves a difference between $x_t^\star$ and x_t, in an analogous way as in the drift term $\texttt{Drift}_T$. We indeed show that we can handle both terms in the same way.

We thus move to the analysis of the $\texttt{Drift}_T$ term. One can immediately observe that, by

Cauchy-Schwarz and by Assumption 2.1,

$$\texttt{Drift}_T = \sum_{t=1}^{T} \langle g_t, x_t - x_t^\star \rangle \leq \sum_{t=1}^{T} \|g_t\|_2 \|x_t^\star - x_t\|_2 \leq L \sum_{t=1}^{T} \|x_t^\star - x_t\|_2 \,. \tag{A.5}$$

While it immediately follows that $\|x_1^\star - x_1\|_2 = 0$ by definition of $x_1^\star$, we require some additional effort when studying the other norms $\|x_t^\star - x_t\|_2$ for $t \geq 2$. To this end, we rely once more on Lemma A.2 for $z_1 = x_t^\star$ with $w_1 = \sum_{\tau \leq t-1} g_\tau$ and $z_2 = x_t$ with $w_2 = \sum_{\tau \in o_t} g_\tau$, using $A = (t-1)\frac{\lambda}{2} I$, and show that

$$\frac{\lambda(t-1)}{2} \|x_t^\star - x_t\|_2^2 = \|x_t^\star - x_t\|_A^2 \leq \frac{1}{4} \left\| \sum_{\tau \in m_t} g_\tau \right\|_{A^{-1}}^2 = \frac{1}{2\lambda(t-1)} \left\| \sum_{\tau \in m_t} g_\tau \right\|_2^2 .$$

We can thus rewrite this inequality in the following way:

$$\|x_t^\star - x_t\|_2 \leq \frac{1}{\lambda(t-1)} \left\| \sum_{\tau \in m_t} g_\tau \right\|_2 \leq \frac{1}{\lambda(t-1)} \sum_{\tau \in m_t} \|g_\tau\|_2 \leq \frac{L|m_t|}{\lambda(t-1)} \,, \tag{A.6}$$

where we used once again that $\|g_\tau\|_2 \leq L$ by Assumption 2.1. The above considerations consequently imply that the sum of interest for bounding $\texttt{Drift}_T$ satisfies

$$\sum_{t=1}^{T} \|x_t^\star - x_t\|_2 \leq \frac{L}{\lambda} \sum_{t=2}^{T} \frac{|m_t|}{t-1} \,. \tag{A.7}$$

The sum on the right-hand side of the above inequality can be immediately bounded as

$$\sum_{t=2}^{T} \frac{|m_t|}{t-1} \leq \sigma_{\max} \sum_{t=2}^{T} \frac{1}{t-1} \leq \sigma_{\max} \ln(2T) \tag{A.8}$$

by definition of $\sigma_{\max}$. Furthermore, by using the fact that $\sum_{\tau \leq t} |m_\tau| \leq (t-1)^2$ since $m_\tau \subseteq [\tau - 1]$ for any τ, we can prove at the same time that

$$\sum_{t=2}^{T} \frac{|m_t|}{t-1} = \sum_{t=2}^{T} \frac{|m_t|}{\sqrt{(t-1)^2}} \leq \sum_{t=2}^{T} \frac{|m_t|}{\sqrt{\sum_{\tau \leq t} |m_\tau|}} \leq 2\sqrt{\sum_{t=1}^{T} |m_t|} \leq 2\sqrt{d_{\text{tot}}} \,, \tag{A.9}$$

where the second inequality is due to Orabona (2025, Lemma 4.13).

Combining all the results gathered so far, we can finally derive the overall regret bound as follows. In particular, for any $u \in \mathcal{X}$, we have

$$\begin{aligned}
\text{Reg}_T(u) &\leq \text{Reg}_T^*(u) + \texttt{Drift}_T - \frac{\lambda}{2} \sum_{t=1}^{T} \|x_t - u\|_2^2 && \text{(Equation (A.1))} \\
&\leq \frac{L^2}{\lambda} \ln(2T+1) + L \sum_{t=1}^{T} \|x_t^\star - x_t\|_2 + \texttt{Drift}_T && \text{(Equation (A.4))} \\
&\leq \frac{L^2}{\lambda} \ln(2T+1) + 2L \sum_{t=1}^{T} \|x_t^\star - x_t\|_2 && \text{(Equation (A.5))}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{L^2}{\lambda}\ln(2T+1) + \frac{2L^2}{\lambda}\sum_{t=2}^{T}\frac{|m_t|}{t-1} && \text{(Equation (A.7))}\\
&\leq \frac{L^2}{\lambda}\ln(2T+1) + \frac{2L^2}{\lambda}\min\left\{\sigma_{\max}\ln(2T), 2\sqrt{d_{\text{tot}}}\right\} && \text{(Equations (A.8) and (A.9))}\\
&= \mathcal{O}\left(\frac{L^2}{\lambda}\left(\ln T + \min\left\{\sigma_{\max}\ln T, \sqrt{d_{\text{tot}}}\right\}\right)\right) . && \square
\end{aligned}
$$

A.3 Omitted details in Section 2.4

In this section, we show the omitted details from Section 2.4. To do so, we first introduce the following useful lemma that will be crucial in the regret analysis of Algorithm 2.2. It essentially corresponds to the standard elliptical potential lemma, but here adapted to the presence of delays.

Lemma A.11. *Let $\phi > 0$, $L > 0$, and $0 < \eta_0 \leq \eta_1 \leq \cdots \leq \eta_N$. For any $t \in [N]$, let $a_t \in \mathbb{R}^n$ such that $\|a_t\|_2 \leq L$ and define $A_t = \eta_t I + \phi\sum_{\tau\leq t} a_\tau a_\tau^\top$. Then, it holds that*

$$
\sum_{t=1}^{N}\|a_t\|_{A_{t-1}^{-1}}\left(\sum_{\tau\in m_t}\|a_\tau\|_{A_{t-1}^{-1}}\right) \leq \frac{2nd_{\max}^{\leq N}}{\phi}\left(\frac{\phi L^2}{\eta_0}+1\right)\ln\left(1+\frac{\phi L^2 N}{\eta_0 n}\right),
$$

and that

$$
\sum_{t=1}^{N}\|a_t\|_{A_{t}^{-1}}\left(\sum_{\tau\in m_t}\|a_\tau\|_{A_{t}^{-1}}\right) \leq \frac{2nd_{\max}^{\leq N}}{\phi}\ln\left(1+\frac{\phi L^2 N}{\eta_0 n}\right).
$$

Proof. Define $B_t = \frac{1}{\phi}A_t$ and $C_t = B_t - \frac{\eta_t-\eta_0}{\phi}I \preceq B_t$ for any $t \in [N]$. By the AM-GM inequality, we first show that

$$
\begin{aligned}
\sum_{t=1}^{N}\|a_t\|_{A_{t-1}^{-1}}\sum_{\tau\in m_t}\|a_\tau\|_{A_{t-1}^{-1}} &\leq \sum_{t=1}^{N}\left(\frac{|m_t|}{2}\|a_t\|^2_{A_{t-1}^{-1}} + \frac{1}{2}\sum_{\tau\in m_t}\|a_\tau\|^2_{A_{t-1}^{-1}}\right)\\
&\leq \sum_{t=1}^{N}\left(\frac{|m_t|}{2}\|a_t\|^2_{A_{t-1}^{-1}} + \frac{1}{2}\sum_{\tau\in m_t}\|a_\tau\|^2_{A_{\tau-1}^{-1}}\right)\\
&= \frac{1}{\phi}\sum_{t=1}^{N}\left(\frac{|m_t|}{2}\|a_t\|^2_{B_{t-1}^{-1}} + \frac{1}{2}\sum_{\tau\in m_t}\|a_\tau\|^2_{B_{\tau-1}^{-1}}\right),
\end{aligned}
$$

where we also used the fact that $A_{\tau-1} \preceq A_{t-1}$ for any $\tau < t$. Now observe that

$$
\sum_{t=1}^{N}|m_t|\cdot\|a_t\|^2_{B_{t-1}^{-1}} \leq d_{\max}^{\leq N}\sum_{t=1}^{N}\|a_t\|^2_{B_{t-1}^{-1}}
$$

since $|m_t| \leq d_{\max}^{\leq N}$ for $t \leq N$. Similarly, we can show that

$$
\sum_{t=1}^{N}\sum_{\tau\in m_t}\|a_\tau\|^2_{B_{\tau-1}^{-1}} = \sum_{t=1}^{N} d_t\|a_t\|^2_{B_{t-1}^{-1}} \leq d_{\max}^{\leq N}\sum_{t=1}^{N}\|a_t\|^2_{B_{t-1}^{-1}}
$$

as for any $\tau \in [N]$ there are no more than d_τ rounds t such that $\tau \in m_t$. Putting these results

together, we obtain that

$$\sum_{t=1}^{N}\left(\frac{|m_t|}{2}\|a_t\|^2_{B_{t-1}^{-1}}+\frac{1}{2}\sum_{\tau\in m_t}\|a_\tau\|^2_{B_{\tau-1}^{-1}}\right)\le d_{\max}^{\le N}\sum_{t=1}^{N}\|a_t\|^2_{B_{t-1}^{-1}}\le d_{\max}^{\le N}\sum_{t=1}^{N}\|a_t\|^2_{C_{t-1}^{-1}}\,.$$

By the fact that $\|a_t\|^2_{C_{t-1}^{-1}}\le\frac{\phi L^2}{\eta_0}$, we can use Lemma 19.4 in Lattimore and Szepesvári (2020) and show that

$$\sum_{t=1}^{N}\|a_t\|^2_{C_{t-1}^{-1}}\le\left(\frac{\phi L^2}{\eta_0}+1\right)\sum_{t=1}^{N}\min\Big\{1,\|a_t\|^2_{C_{t-1}^{-1}}\Big\}\le 2n\left(\frac{\phi L^2}{\eta_0}+1\right)\ln\left(1+\frac{L^2N}{\eta_0 n}\right).$$

Concatenating all the above results concludes the proof of the first inequality.

For the second inequality, similar steps suffice to prove it, but with a different observation that now $\|a_t\|^2_{C_t^{-1}}\le\min\Big\{1,\|a_t\|^2_{C_{t-1}^{-1}}\Big\}$ because

$$\|a_t\|^2_{C_t^{-1}}\le a_t^\top\left(\upsilon I+a_ta_t^\top\right)^{-1}a_t=a_t^\top\left(\frac{1}{\upsilon}I-\frac{a_ta_t^\top}{\upsilon^2+\upsilon\|a_t\|_2^2}\right)a_t=\frac{\|a_t\|_2^2}{\upsilon}-\frac{\|a_t\|_2^4}{\upsilon^2+\upsilon\|a_t\|_2^2}=\frac{\|a_t\|_2^2}{\upsilon+\|a_t\|_2^2}\le 1\,,$$

where we used the Sherman-Morrison formula in the first equality with $\upsilon=\eta_0/\phi$, and since $\|a_t\|_{C_t^{-1}}\le\|a_t\|_{C_{t-1}^{-1}}$ given that $C_{t-1}\preceq C_t$. □

For completeness, we restate Theorem 2.2, the main result of Section 2.4.1, and provide its proof.

Theorem 2.2. *Assume that $\ell_1,\dots,\ell_T$ are α-exp-concave and let $\beta=\frac{1}{2}\min\{\frac{1}{4LD},\alpha\}$. Then, under Assumption 2.1 and Assumption 2.2, Algorithm 2.2 with $0<\eta_0\le\eta_1\le\cdots\le\eta_T$ guarantees that*

$$\mathrm{Reg}_T=\mathcal{O}\Big(\frac{n}{\beta}\ln\Big(1+\frac{\beta L^2T}{\eta_0 n}\Big)+\eta_TD^2+\min\{B_1,B_2\}\Big),$$

where $B_1=\left(\frac{L^2}{\eta_0}+\frac{1}{\beta}\right)nd_{\max}\ln\left(1+\frac{\beta L^2T}{\eta_0 n}\right)$ and $B_2=L^2\sum_{t=1}^{T}\frac{|m_t|}{\eta_{t-1}}$.

Proof. First, in a similar way as in the proof of Theorem 2.1, we define

$$F_t(x)=\sum_{\tau\in o_t}\langle g_\tau,x\rangle+\psi_t(x)\qquad\text{and}\qquad F_t^\star(x)=\sum_{\tau=1}^{t-1}\langle g_\tau,x\rangle+\psi_t^\star(x),$$

where $\psi_t(x)=\frac{\eta_{t-1}}{2}\|x\|_2^2+\frac{\beta}{2}\sum_{\tau\in o_t}\left(\langle g_\tau,x-x_\tau\rangle\right)^2$ and $\psi_t^\star(x)=\frac{\eta_{t-1}}{2}\|x\|_2^2+\frac{\beta}{2}\sum_{\tau=1}^{t-1}\left(\langle g_\tau,x-x_\tau\rangle\right)^2$. Observe that $x_t\in\arg\min_{x\in\mathcal{X}}F_t(x)$, and define $x_t^\star\in\arg\min_{x\in\mathcal{X}}F_t^\star(x)$ for $t\ge1$ to be the predictions following a similar update rule while using all the information up to round $t-1$. Similarly to the regret decomposition for the strongly convex case shown in Appendix A.2, we decompose the regret as follows:

$$\mathrm{Reg}_T(u)=\sum_{t=1}^{T}(\ell_t(x_t)-\ell_t(u))\le\sum_{t=1}^{T}\left(\langle g_t,x_t-u\rangle-\frac{\beta}{2}\langle x_t-u,g_t\rangle^2\right)$$

$$= \underbrace{\sum_{t=1}^{T} \langle g_t, x_t^\star - u \rangle}_{\mathrm{Reg}_T^\star(u)} + \underbrace{\sum_{t=1}^{T} \langle g_t, x_t - x_t^\star \rangle}_{\texttt{Drift}_T} - \frac{\beta}{2} \sum_{t=1}^{T} (\langle x_t - u, g_t \rangle)^2 \,, \quad \text{(A.10)}$$

where the inequality holds thanks to Lemma A.3.

Let us begin the analysis of the "linearized" regret by first focusing on the cheating term $\mathrm{Reg}_T^\star(u)$. Let $F_t'(x) = F_t^\star(x) + \langle g_t, x \rangle$ and define $x_t' \in \arg\min_{x \in \mathcal{X}} F_t'(x)$. Leveraging Lemma A.1 with $\ell_t(\cdot) = \langle g_t, \cdot \rangle$, we show that

$$\begin{aligned}
\mathrm{Reg}_T^\star(u) &= \sum_{t=1}^{T} \langle g_t, x_t^\star - u \rangle \\
&= \psi_{T+1}^\star(u) - \min_{x \in \mathcal{X}} \psi_1^\star(x) + \sum_{t=1}^{T} \left[F_t^\star\left(x_t^\star\right) - F_{t+1}^\star\left(x_{t+1}^\star\right) + \langle g_t, x_t^\star \rangle \right] + F_{T+1}^\star\left(x_{T+1}^\star\right) - F_{T+1}^\star(u) \\
&\le \psi_{T+1}^\star(u) + \sum_{t=1}^{T} \left[\left(F_t^\star(x_t^\star) + \langle g_t, x_t^\star \rangle \right) - \left(F_t^\star(x_{t+1}^\star) + \langle g_t, x_{t+1}^\star \rangle \right) - \psi_{t+1}^\star(x_{t+1}^\star) + \psi_t^\star(x_{t+1}^\star) \right] \\
&\le \psi_{T+1}^\star(u) + \sum_{t=1}^{T} \left[F_t'(x_t^\star) - F_t'(x_t') + \psi_t^\star(x_{t+1}^\star) - \psi_{t+1}^\star(x_{t+1}^\star) \right] \qquad \text{(definition of } F_t' \text{ and } x_t') \\
&\le \psi_{T+1}^\star(u) + \sum_{t=1}^{T} \left(F_t'(x_t^\star) - F_t'(x_t') \right) , \qquad \text{(A.11)}
\end{aligned}$$

where in the first inequality we used the facts that $F_{T+1}^\star(x_{T+1}^\star) \le F_{T+1}^\star(u)$ and that $\psi_1^\star$ is nonnegative, while the last inequality is due to $\psi_t^\star(x_{t+1}^\star) \le \psi_{t+1}^\star(x_{t+1}^\star)$. Applying now Lemma A.2, we have $\|x_t^\star - x_t'\|_{A_{t-1}} \le \|g_t\|_{A_{t-1}^{-1}}$, where $A_{t-1} = \eta_{t-1} I + \beta \sum_{\tau=1}^{t-1} g_\tau g_\tau^\top$. This further means that

$$\begin{aligned}
F_t'(x_t^\star) - F_t'(x_t') &\le \left\langle \nabla F_t'(x_t^\star), x_t^\star - x_t' \right\rangle && \text{(convexity of } F_t') \\
&= \left\langle \nabla F_t^\star(x_t^\star) + g_t, x_t^\star - x_t' \right\rangle && \text{(definition of } F_t') \\
&\le \left\langle g_t, x_t^\star - x_t' \right\rangle && \text{(first-order optimality)} \\
&\le \min\left\{ \|g_t\|_2 \|x_t^\star - x_t'\|_2 \,, \|g_t\|_{A_{t-1}^{-1}} \|x_t^\star - x_t'\|_{A_{t-1}} \right\} && \text{(Cauchy-Schwarz inequality)} \\
&\le \min\left\{ LD, \|g_t\|_{A_{t-1}^{-1}} \|x_t^\star - x_t'\|_{A_{t-1}} \right\} && \text{(Assumption 2.1 and Assumption 2.2)} \\
&\le \min\left\{ LD, \|g_t\|_{A_{t-1}^{-1}}^2 \right\} . && \text{(A.12)}
\end{aligned}$$

We now focus on the sum of terms on the right-hand side of Equation (A.11). Because η_t is non-decreasing by assumption, we have

$$\begin{aligned}
\sum_{t=1}^{T} \left(F_t'(x_t^\star) - F_t'(x_t') \right) &\le \sum_{t=1}^{T} \min\left\{ LD, \|g_t\|_{A_{t-1}^{-1}}^2 \right\} && \text{(Equation (A.12))} \\
&\le \sum_{t=1}^{T} \min\left\{ LD, \frac{1}{\beta} \|g_t\|_{(\frac{\eta_0}{\beta} I + \sum_{\tau<t} g_\tau g_\tau^\top)^{-1}}^2 \right\} && (\eta_{t-1} I \succeq \eta_0 I) \\
&\le \max\left\{ LD, \frac{1}{\beta} \right\} \sum_{t=1}^{T} \min\left\{ 1, \|g_t\|_{(\frac{\eta_0}{\beta} I + \sum_{\tau<t} g_\tau g_\tau^\top)^{-1}}^2 \right\}
\end{aligned}$$

$$\leq \left(LD + \frac{1}{\beta}\right) n \ln\left(1 + \frac{\beta L^2 T}{n\eta_0}\right) , \tag{A.13}$$

where the last inequality follows by Lattimore and Szepesvári (2020, Lemma 19.4). Combining the previous inequalities, we can show that $\mathrm{Reg}_T^\star(u)$ satisfies

$$\begin{aligned}
\mathrm{Reg}_T^\star(u) &\leq \psi_{T+1}^\star(u) + \sum_{t=1}^{T} \left(F_t'(x_t^\star) - F_t'(x_t')\right) && \text{(Equation (A.11))} \\
&\leq \psi_{T+1}^\star(u) + \frac{\beta}{2}\sum_{t=1}^{T} \left(\langle g_t, u - x_t\rangle\right)^2 + \left(LD + \frac{1}{\beta}\right) n \ln\left(1 + \frac{\beta L^2 T}{n\eta_0}\right) && \text{(Equation (A.13))} \\
&= \frac{\eta_T}{2}\|u\|_2^2 + \frac{\beta}{2}\sum_{t=1}^{T} \left(\langle g_t, u - x_t\rangle\right)^2 + \left(LD + \frac{1}{\beta}\right) n \ln\left(1 + \frac{\beta L^2 T}{n\eta_0}\right) , && \text{(A.14)}
\end{aligned}$$

where we simply replace $\psi_{T+1}^\star$ with its definition in the last step.

We thus move to the analysis of the $\texttt{Drift}_T$ term. Using the Cauchy-Schwarz inequality, we have

$$\texttt{Drift}_T = \sum_{t=1}^{T} \langle g_t, x_t - x_t^\star\rangle \leq \sum_{t=1}^{T} \|g_t\|_{A_{t-1}^{-1}} \cdot \|x_t - x_t^\star\|_{A_{t-1}} . \tag{A.15}$$

Applying Lemma A.2, we obtain that

$$F_t^\star(x_t) - F_t^\star(x_t^\star) \geq \frac{1}{2}\|x_t - x_t^\star\|_{A_{t-1}}^2 \qquad \text{and} \qquad F_t(x_t^\star) - F_t(x_t) \geq \frac{1}{2}\|x_t - x_t^\star\|_{A_{o_t}}^2 ,$$

where $A_{o_t} = \eta_{t-1} I + \beta \sum_{\tau \in o_t} g_\tau g_\tau^\top$. Summing the above inequalities, and replacing $F_t^\star$ and F_t with their definitions, it follows that

$$\begin{aligned}
&\frac{1}{2}\|x_t - x_t^\star\|_{A_{o_t}}^2 + \frac{1}{2}\|x_t - x_t^\star\|_{A_{t-1}}^2 \\
&\quad\leq \sum_{\tau=1}^{t-1} \langle g_\tau, x_t\rangle - \sum_{\tau \in o_t} \langle g_\tau, x_t^\star\rangle + \sum_{\tau=1}^{t-1} \langle g_\tau, x_t^\star\rangle - \sum_{\tau \in o_t} \langle g_\tau, x_t\rangle \\
&\qquad + \frac{\beta}{2}\left(\sum_{\tau=1}^{t-1} \left(\langle g_\tau, x_t - x_\tau\rangle\right)^2 - \sum_{\tau=1}^{t-1} \left(\langle g_\tau, x_t^\star - x_\tau\rangle\right)^2 + \sum_{\tau \in o_t} \left(\langle g_\tau, x_t^\star - x_\tau\rangle\right)^2 - \sum_{\tau \in o_t} \left(\langle g_\tau, x_t - x_\tau\rangle\right)^2\right) \\
&\quad\leq \sum_{\tau \in m_t} \langle g_\tau, x_t - x_t^\star\rangle + \frac{\beta}{2}\left(\sum_{\tau \in m_t} \left(\langle g_\tau, x_t - x_\tau\rangle\right)^2 - \sum_{\tau \in m_t} \left(\langle g_\tau, x_t^\star - x_\tau\rangle\right)^2\right) \\
&\quad= \sum_{\tau \in m_t} \langle g_\tau, x_t - x_t^\star\rangle + \frac{\beta}{2}\left(\sum_{\tau \in m_t} \langle g_\tau, x_t - x_t^\star\rangle \cdot \langle g_\tau, x_t + x_t^\star - 2x_\tau\rangle\right) \\
&\quad\leq \sum_{\tau \in m_t} |\langle g_\tau, x_t - x_t^\star\rangle| + \frac{\beta}{2}\sum_{\tau \in m_t} |\langle g_\tau, x_t - x_t^\star\rangle|\,|\langle g_\tau, x_t + x_t^\star - 2x_\tau\rangle| \\
&\quad\leq (1 + 2LD\beta) \sum_{\tau \in m_t} |\langle g_\tau, x_t - x_t^\star\rangle| && \text{(Assumptions 2.1 and 2.2)} \\
&\quad\leq (1 + 2LD\beta)\left(\sum_{\tau \in m_t} \|g_\tau\|_{A_{t-1}^{-1}}\right)\|x_t - x_t^\star\|_{A_{t-1}} && \text{(Cauchy-Schwarz inequality)}
\end{aligned}$$

$$\leq \frac{5}{4}\left(\sum_{\tau\in m_t}\|g_\tau\|_{A_{t-1}^{-1}}\right)\|x_t - x_t^\star\|_{A_{t-1}} \qquad (\beta \leq \tfrac{1}{8LD})$$
$$\leq 2\left(\sum_{\tau\in m_t}\|g_\tau\|_{A_{t-1}^{-1}}\right)\|x_t - x_t^\star\|_{A_{t-1}}\,.$$

Rearranging the terms, we can obtain that $\|x_t - x_t^\star\|_{A_{t-1}} \leq 4\sum_{\tau\in m_t}\|g_\tau\|_{A_{t-1}^{-1}}$. Plugging this inequality into $\texttt{Drift}_T$, we have

$$\begin{aligned}
\texttt{Drift}_T &\leq \sum_{t=1}^{T}\|g_t\|_{A_{t-1}^{-1}}\cdot\|x_t - x_t^\star\|_{A_{t-1}} && \text{(Equation (A.15))}\\
&\leq 4\sum_{t=1}^{T}\|g_t\|_{A_{t-1}^{-1}}\left(\sum_{\tau\in m_t}\|g_\tau\|_{A_{t-1}^{-1}}\right)\\
&\leq 8d_{\max}^{\leq T} n\left(\frac{L^2}{\eta_0}+\frac{1}{\beta}\right)\ln\left(1+\frac{\beta T L^2}{n\eta_0}\right)\,, && \text{(A.16)}
\end{aligned}$$

where the last inequality is due to Lemma A.11. On the other hand, we can also bound $\texttt{Drift}_T$ in a different way:

$$\begin{aligned}
\texttt{Drift}_T &\leq \sum_{t=1}^{T}\|g_t\|_{A_{t-1}^{-1}}\cdot\|x_t - x_t^\star\|_{A_{t-1}}\\
&\leq 4\sum_{t=1}^{T}\|g_t\|_{A_{t-1}^{-1}}\left(\sum_{\tau\in m_t}\|g_\tau\|_{A_{t-1}^{-1}}\right)\\
&\leq 4L^2\sum_{t=1}^{T}\frac{|m_t|}{\eta_{t-1}}\,, && \text{(A.17)}
\end{aligned}$$

where in the last step we use the fact that $\|g_s\|^2_{A_{t-1}^{-1}} \leq \frac{L^2}{\eta_{t-1}}$ for any $s\in[T]$, also due to Assumption 2.1. Combining all bounds together, we finally obtain that

$$\begin{aligned}
\mathrm{Reg}_T(u) &\leq \mathrm{Reg}_T^\star(u) + \texttt{Drift}_T - \frac{\beta}{2}\sum_{t=1}^{T}(\langle x_t - u, g_t\rangle)^2 && \text{(Equation (A.10))}\\
&\leq \frac{\eta_T}{2}\|u\|_2^2 + \left(LD+\frac{1}{\beta}\right)n\ln\left(1+\frac{\beta L^2 T}{n\eta_0}\right) + \texttt{Drift}_T && \text{(Equation (A.14))}\\
&\leq \frac{\eta_T}{2}\|u\|_2^2 + \left(LD+\frac{1}{\beta}\right)n\ln\left(1+\frac{\beta L^2 T}{n\eta_0}\right)\\
&\quad + 4\min\left\{2d_{\max}^{\leq T} n\left(\frac{L^2}{\eta_0}+\frac{1}{\beta}\right)\ln\left(1+\frac{\beta L^2 T}{\eta_0 n}\right), L^2\sum_{t=1}^{T}\frac{|m_t|}{\eta_{t-1}}\right\}\\
&&& \text{(Equations (A.16) and (A.17))}\\
&= \mathcal{O}\left(\frac{n}{\beta}\ln\left(1+\frac{\beta L^2 T}{\eta_0 n}\right) + \eta_T D^2 + \min\{B_1, B_2\}\right)\,, && \text{(Assumption 2.2)}
\end{aligned}$$

where

$$B_1 = \left(\frac{L^2}{\eta_0}+\frac{1}{\beta}\right)n d_{\max}\ln\left(1+\frac{\beta L^2 T}{\eta_0 n}\right) \qquad \text{and} \qquad B_2 = L^2\sum_{t=1}^{T}\frac{|m_t|}{\eta_{t-1}}$$

are defined as in the theorem statement, and we used the fact that $GD \leq \frac{1}{\beta}$. □

The following corollary is a restatement of Corollary 2.1, which shows that via an adaptive tuning of the learning rate used by Algorithm 2.2, we are able to guarantee $\mathcal{O}(\min\{d_{\max} \ln T, \sqrt{d_{\text{tot}}}\})$ regret.

Corollary 2.1. *Assume that $\ell_1, \ldots, \ell_T$ are α-exp-concave and let $\beta = \frac{1}{2}\min\{\frac{1}{4LD}, \alpha\}$. Then, under Assumption 2.1 and Assumption 2.2, Algorithm 2.2 with the adaptive learning rate $\eta_t = \min\{a_t, b_t\}+1$, where a_t and b_t are defined in Equations (2.6) and (2.7), guarantees that*

$$\text{Reg}_T = \mathcal{O}\left(\frac{n}{\beta}\ln\left(1 + \frac{\beta L^2 T}{n}\right) + D^2 + \min\{C_1, C_2\}\right),$$

where $C_1 = \left(\frac{D}{L} + 1\right)\left(L^2 + \frac{1}{\beta}\right) n d_{\max} \ln\left(1 + \frac{\beta L^2 T}{n}\right)$ and $C_2 = \left(L^2 + LD\right)\left(\sqrt{d_{\text{tot}}} + 1\right)$.

Proof. The adaptive learning rate is given by $\eta_0 = 1$ and $\eta_t = \min\{a_t, b_t\} + 1$ for all $t \geq 1$, where we recall that

$$a_t = \frac{2}{LD}\left(L^2 + \frac{1}{\beta}\right) n d_{\max}^{\leq t} \ln\left(1 + \frac{\beta L^2 T}{n}\right) \qquad \text{and} \qquad b_t = \frac{L}{D}\sqrt{\sum_{s=1}^{t} |m_s| + |m_t| + 1}\,,$$

Note that η_t is non-decreasing since a_t and b_t are non-decreasing. When $a_T \leq b_T$, we have

$$\text{Reg}_T(u) \leq \left(LD + \frac{1}{\beta}\right) n \ln\left(1 + \frac{\beta G T}{n}\right) + D^2 + \left(\frac{2D}{G} + 8\right)\left(L^2 + \frac{1}{\beta}\right) n d_{\max} \ln\left(1 + \frac{\beta L^2 T}{n}\right), \tag{A.18}$$

where $\|u\|_2 \leq D$ by Assumption 2.2. When $a_T \geq b_T$, we instead have

$$\begin{aligned}\text{Reg}_T(u) \leq \left(LD + \frac{1}{\beta}\right) n \ln\left(1 + \frac{\beta L^2 T}{n}\right) + D^2 + GD\left(\sqrt{\sum_{t=1}^{T} |m_t|} + 1\right) \\ + \sum_{t=1}^{\tau^\star} \|g_t\|_{A_{t-1}^{-1}} \cdot \|x_t - x_t^\star\|_{A_{t-1}} + \sum_{t=\tau^\star+1}^{T} \|g_t\|_{A_{t-1}^{-1}} \cdot \|x_t - x_t^\star\|_{A_{t-1}},\end{aligned}$$

where $\tau^\star$ is last round $a_{\tau^\star} \leq b_{\tau^\star}$. Hence, we have

$$\begin{aligned}\sum_{t=1}^{\tau^\star} \|g_t\|_{A_{t-1}^{-1}} \cdot \|x_t - x_t^\star\|_{A_{t-1}} &\leq 8\left(L^2 + \frac{1}{\beta}\right) n d_{\max}^{\leq \tau^\star} \ln\left(1 + \frac{\beta L^2 T}{n}\right) && \text{(Equation (A.16))} \\ &\leq 8L^2\sqrt{\sum_{t=1}^{\tau^\star} |m_t| + |m_{\tau^\star}| + 1} \\ &\leq 8L^2\left(\sqrt{\sum_{t=1}^{T} |m_t|} + 1\right) && \text{(A.19)}\end{aligned}$$

Regarding the remaining rounds until T, we can also show that

$$\sum_{t=\tau^\star+1}^{T} \|g_t\|_{A_{t-1}^{-1}} \cdot \|x_t - x_t^\star\|_{A_{t-1}} \leq 4L^2 \sum_{t=\tau^\star+1}^{T} \frac{|m_t|}{\eta_{t-1}} \qquad \text{(Equation (A.17))}$$

$$
\begin{aligned}
&\leq 4L^2 \sum_{t=\tau^\star+1}^{T} \frac{D|m_t|}{G\sqrt{\sum_{s=1}^{t-1}|m_s| + |m_{t-1}| + 1}} \\
&\leq 8L^2 \sum_{t=\tau^\star+1}^{T} \frac{D|m_t|}{G\sqrt{\sum_{s=\tau^\star+1}^{t}|m_s|}} \\
&\leq 8LD\sqrt{\sum_{t=\tau^\star+1}^{T}|m_t|} \\
&\leq 8LD\sqrt{\sum_{t=1}^{T}|m_t|}, \qquad \text{(A.20)}
\end{aligned}
$$

where the last inequality is due to Orabona (2025, Lemma 4.13). Combining the above three inequalities together, we have

$$
\mathrm{Reg}_T(u) \leq \left(LD + \frac{1}{\beta}\right) n \ln\left(1 + \frac{\beta L^2 T}{n}\right) + D^2 + \left(8L^2 + 9LD\right)\left(\sqrt{\sum_{t=1}^{T}|m_t|} + 1\right).
$$

Finally, we obtain

$$
\begin{aligned}
\mathrm{Reg}_T(u) &\leq \left(LD + \frac{1}{\beta}\right) n \ln\left(1 + \frac{\beta L^2 T}{n}\right) + D^2 \\
&\quad + \min\left\{\left(\frac{2D}{L} + 8\right)\left(L^2 d_{\max}^{\leq T} + \frac{d_{\max}^{\leq T}}{\beta}\right) n \ln\left(1 + \frac{\beta L^2 T}{n}\right), \left(8L^2 + 9LD\right)\left(\sqrt{d_{\mathrm{tot}}} + 1\right)\right\}. \\
&= \mathcal{O}\left(\frac{1}{\beta}\ln\left(1 + \frac{\beta L^2 T}{n}\right) + D^2 + \min\left\{C_1, C_2\right\}\right),
\end{aligned}
$$

where

$$
C_1 = \left(\frac{D}{L} + 1\right)\left(L^2 + \frac{1}{\beta}\right) n d_{\max} \ln\left(1 + \frac{\beta L^2 T}{n}\right)
$$

and

$$
C_2 = \left(L^2 + LD\right)\left(\sqrt{d_{\mathrm{tot}}} + 1\right)
$$

as in the theorem statement. □

A.4 Omitted details in Section 2.5

Here we present the omitted details from Section 2.5. For completeness, we restate the main result (Theorem 2.3) and provide its proof.

Theorem 2.3. *In the OLR problem with delayed labels under Assumption 2.3, Algorithm 2.3 guarantees for any* $0 < \eta_0 \leq \eta_1 \leq \cdots \leq \eta_T$ *that*

$$
\mathrm{Reg}_T(u) \leq \frac{\eta_T}{2}\|u\|_2^2 + nY^2 \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right) + \mathcal{O}\Big(Y^2\big(\sigma_{\max} + \min\left\{M_1, M_2\right\}\big)\Big),
$$

where $M_1 = n d_{\max} \ln\big(1 + \frac{Z^2 T}{\eta_0 n}\big)$ *and* $M_2 = Z^2 \sum_{t=1}^{T} \frac{|m_t|}{\eta_t}$.

Proof. We begin by defining

$$F_t(x) = \sum_{\tau \in o_t} -y_\tau \langle z_\tau, x\rangle + \psi_t(x) \qquad \text{and} \qquad F_t^*(x) = \sum_{\tau=1}^{t-1} -y_\tau \langle z_\tau, x\rangle + \psi_t(x),$$

where $\psi_t(x) = \frac{1}{2}\sum_{\tau=1}^{t} \left(\langle z_\tau, x\rangle\right)^2 + \frac{\eta_t}{2}\|x\|_2^2$ for $t \in [T]$, and we let $\psi_{T+1} = \psi_T$. Observe that $x_t \in \arg\min_{x\in\mathbb{R}^n} F_t(x)$, and define $x_t^\star \in \arg\min_{x\in\mathbb{R}^n} F_t^\star(x)$ for $t \geq 1$ to be the predictions following a similar update rule while using all the information up to round $t-1$, including the labels y_τ for rounds $\tau \in m_t$ that the algorithm is missing because of the delays.

Similarly to the regret decomposition for the strongly convex case shown in Appendix A.2, we rewrite the regret as follows:

$$\mathrm{Reg}_T(u) = \sum_{t=1}^{T}\big(\ell_t(\widetilde{x}_t) - \ell_t(u)\big) = \underbrace{\sum_{t=1}^{T}\big(\ell_t(x_t^\star) - \ell_t(u)\big)}_{\mathrm{Reg}_T^\star(u)} + \underbrace{\sum_{t=1}^{T}\big(\ell_t(\widetilde{x}_t) - \ell_t(x_t^\star)\big)}_{\texttt{Drift}_T}, \tag{A.21}$$

where $\mathrm{Reg}_T^\star(u)$ is the cheating regret for the iterates $x_1^\star, \dots, x_T^\star$, while $\texttt{Drift}_T$ is a drift term that quantifies the influence of the missing labels on the regret because of the delayed feedback. Note that, contrarily to other regret analyses in this work, here $\texttt{Drift}_T$ is also affected by the clipping in the definition of $\widetilde{x}_t$.

Let us first analyze the cheating regret $\mathrm{Reg}_T^\star(u)$. By the definition of the loss $\ell_t(x) = \frac{1}{2}\big(\langle z_t, x\rangle - y_t\big)^2$, we can rewrite the regret in the following way:

$$\mathrm{Reg}_T^\star(u) = \sum_{t=1}^{T}\big(\ell_t(x_t^\star) - \ell_t(u)\big) = \frac{1}{2}\sum_{t=1}^{T}\big(\langle z_t, x_t^\star\rangle\big)^2 + \sum_{t=1}^{T}\big(-y_t\langle z_t, x_t^\star\rangle + y_t\langle z_t, u\rangle\big) - \frac{1}{2}\sum_{t=1}^{T}\big(\langle z_t, u\rangle\big)^2. \tag{A.22}$$

We can now move our focus on the central sum, which essentially corresponds to the regret of the same sequence $\big(x_t^\star\big)_{t\geq 1}$ against the comparator $u \in \mathbb{R}^n$, but with respect to the linear losses $x \mapsto -y_t\langle z_t, x\rangle$. Additionally define $F_t'(x) = F_t^\star(x) - y_t\langle z_t, x\rangle$ for notational convenience. Hence, we analyze the above-mentioned term by applying Lemma A.1, which yields

$$\begin{aligned}
&\sum_{t=1}^{T}\big(-y_t\langle z_t, x_t^\star\rangle + y_t\langle z_t, u\rangle\big) \\
&\quad= \psi_{T+1}(u) - \min_{x\in\mathbb{R}^n}\psi_1(x) + \sum_{t=1}^{T}\Big[F_t^\star(x_t^\star) - F_{t+1}^\star(x_{t+1}^\star) - y_t\langle z_t, x_{t+1}^\star\rangle\Big] + F_{T+1}^\star(x_{T+1}^\star) - F_{T+1}^\star(u) \\
&\quad\leq \psi_{T+1}(u) + \sum_{t=1}^{T}\Big[F_t^\star(x_t^\star) - F_{t+1}^\star(x_{t+1}^\star) - y_t\langle z_t, x_{t+1}^\star\rangle\Big] \\
&\quad= \psi_{T+1}(u) + \sum_{t=1}^{T}\big(F_t'(x_t^\star) - F_t'(x_{t+1}^\star)\big) - \sum_{t=1}^{T}\big(\psi_{t+1}(x_{t+1}^\star) - \psi_t(x_{t+1}^\star)\big) \\
&\quad= \psi_T(u) + \sum_{t=1}^{T}\big(F_t'(x_t^\star) - F_t'(x_{t+1}^\star)\big) - \frac{1}{2}\sum_{t=1}^{T}\big(\langle z_t, x_t^\star\rangle\big)^2
\end{aligned}$$

$$\leq \psi_T(u) + \sum_{t=1}^{T}\big(F_t'(x_t^\star) - F_t'(x_t')\big) - \frac{1}{2}\sum_{t=1}^{T}\big(\langle z_t, x_t^\star\rangle\big)^2 \,, \tag{A.23}$$

where we let $x_t' \in \arg\min_{x\in\mathbb{R}^n} F_t'(x)$; in particular, the first inequality is due to the fact that $F_{T+1}^\star(x_{T+1}^\star) \leq F_{T+1}^\star(u)$ and that ψ_1 is non-negative, whereas the last equality follows by definition of ψ_t and $x_1^\star = 0$.

Consider now any term $F_t'(x_t^\star) - F_t'(x_t')$ in the sum after the last inequality and let $A_t = \eta_t I + \sum_{\tau=1}^{t} z_\tau z_\tau^\top$. Applying Lemma A.2 for $z_1 = x_t'$ and $z_2 = x_t^\star$ with $A = A_t$, we derive that

$$\|x_t^\star - x_t'\|_{A_t} \leq \frac{|y_t|}{2}\|z_t\|_{A_t^{-1}} \,. \tag{A.24}$$

We can now use this fact to show that

$$\begin{aligned}
F_t'(x_t^\star) - F_t'(x_t') &\leq \langle \nabla F_t'(x_t^\star), x_t^\star - x_t'\rangle && \text{(convexity of } F_t'\text{)}\\
&= \langle \nabla F_t^\star(x_t^\star) - y_t z_t, x_t^\star - x_t'\rangle && \text{(definition of } F_t'\text{)}\\
&\leq y_t\langle z_t, x_t' - x_t^\star\rangle && \text{(first-order optimality)}\\
&\leq |y_t|\|z_t\|_{A_t^{-1}}\|x_t^\star - x_t'\|_{A_t} && \text{(Cauchy-Schwarz inequality)}\\
&\leq \frac{|y_t|^2}{2}\|z_t\|_{A_t^{-1}}^2 && \text{(Equation (A.24))}\\
&\leq \frac{Y^2}{2}\|z_t\|_{A_t^{-1}}^2 \,, && \text{(A.25)}
\end{aligned}$$

where the last step is a consequence of $|y_t| \leq Y$ by Assumption 2.3. Further notice that $\|z_t\|_{A_t^{-1}}^2 \leq \|z_t\|_{A_{t-1}^{-1}}^2$ since $A_{t-1} \preceq A_t$, as well as

$$\|z_t\|_{A_t^{-1}}^2 \leq z_t^\top\big(\eta_t I + z_t z_t^\top\big)^{-1} z_t = z_t^\top\left(\frac{1}{\eta_t} I - \frac{z_t z_t^\top}{\eta_t^2 + \eta_t\|z_t\|_2^2}\right) z_t = \frac{\|z_t\|_2^2}{\eta_t} - \frac{\|z_t\|_2^4}{\eta_t^2 + \eta_t\|z_t\|_2^2} = \frac{\|z_t\|_2^2}{\eta_t + \|z_t\|_2^2} \leq 1 \,,$$

using the Sherman-Morrison formula at the first equality. Therefore, we show that the sum of the terms involving F_t' is

$$\begin{aligned}
\sum_{t=1}^{T}\big(F_t'(x_t^\star) - F_t'(x_t')\big) &\leq \frac{Y^2}{2}\sum_{t=1}^{T}\|z_t\|_{A_t^{-1}}^2 && \text{(Equation (A.25))}\\
&\leq \frac{Y^2}{2}\sum_{t=1}^{T}\min\Big\{1, \|z_t\|_{A_{t-1}^{-1}}^2\Big\}\\
&\leq nY^2\ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right) \,, && \text{(A.26)}
\end{aligned}$$

using Lemma 19.4 in Lattimore and Szepesvári (2020) at the last step. Then, combining together all these observations, we can bound $\mathrm{Reg}_T^\star(u)$ from above and obtain that

$$\mathrm{Reg}_T^\star(u) \leq \sum_{t=1}^{T}\big(F_t'(x_t^\star) - F_t'(x_t')\big) + \psi_T(u) - \frac{1}{2}\sum_{t=1}^{T}\big(\langle z_t, u\rangle\big)^2 \quad \text{(Equations (A.22) and (A.23))}$$

$$\leq nY^2 \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right) + \psi_T(u) - \frac{1}{2}\sum_{t=1}^{T}\big(\langle z_t, u\rangle\big)^2 \qquad \text{(Equation (A.26))}$$

$$= \frac{\eta_T}{2}\|u\|_2^2 + nY^2 \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right). \qquad \text{(definition of } \psi_T\text{)} \qquad \text{(A.27)}$$

Let us now consider the drift term $\texttt{Drift}_T$ from the decomposition in Equation (A.21). Define $\mathcal{T} = \{t \in [T] : \ell_t(\widetilde{x}_t) > \ell_t(x_t^\star)\}$ to be the rounds when $\widetilde{x}_t$ is worse than $x_t^\star$ with respect to the square loss ℓ_t. Moreover, recall the definition of $\rho_t = \max_{\tau \in o_t} |y_\tau|$ as the threshold used for clipping in the definition of $\widetilde{x}_t$. By the convexity of ℓ_t, we immediately have that

$$\texttt{Drift}_T \leq \sum_{t\in\mathcal{T}}\big(\ell_t(\widetilde{x}_t) - \ell_t(x_t^\star)\big) \leq \sum_{t\in\mathcal{T}}\langle \nabla \ell_t(\widetilde{x}_t), \widetilde{x}_t - x_t^\star\rangle = \sum_{t\in\mathcal{T}}\big(\langle z_t, \widetilde{x}_t\rangle - y_t\big)\big(\langle z_t, \widetilde{x}_t\rangle - \langle z_t, x_t^\star\rangle\big)\,. \quad \text{(A.28)}$$

Now, we distinguish the two following cases for any $t \in \mathcal{T}$:

- $\ell_t(\widetilde{x}_t) \leq \ell_t(x_t)$: thus, if $\langle z_t, \widetilde{x}_t\rangle \leq y_t$ it must be the case that $\langle z_t, x_t\rangle \leq \langle z_t, \widetilde{x}_t\rangle$, otherwise if $\langle z_t, \widetilde{x}_t\rangle > y_t$ then $\langle z_t, x_t\rangle \geq \langle z_t, \widetilde{x}_t\rangle$; in either case we have that

$$\begin{aligned}
\big(\langle z_t, \widetilde{x}_t\rangle - y_t\big)\big(\langle z_t, \widetilde{x}_t\rangle - \langle z_t, x_t^\star\rangle\big) &\leq \big(\langle z_t, \widetilde{x}_t\rangle - y_t\big)\big(\langle z_t, x_t\rangle - \langle z_t, x_t^\star\rangle\big) \\
&\leq \big(|\rho_t| + |y_t|\big)|\langle z_t, x_t - x_t^\star\rangle| && \text{(triangle inequality, definition of } \widetilde{x}_t\text{)} \\
&\leq 2Y|\langle z_t, x_t - x_t^\star\rangle| && \text{(Assumption 2.3)} \\
&\leq 2Y\|z_t\|_{A_t^{-1}}\|x_t - x_t^\star\|_{A_t}\,. && \text{(Cauchy-Schwarz)}
\end{aligned} \quad \text{(A.29)}$$

- $\ell_t(\widetilde{x}_t) > \ell_t(x_t)$: here it must be the case that $\widetilde{x}_t \neq x_t$, $y_t\langle z_t, \widetilde{x}_t\rangle \geq 0$, and $|y_t| > \rho_t$ (otherwise, clipping would have only decreased the square loss ℓ_t); since $t \in \mathcal{T}$ implies that $|\langle z_t, x_t^\star\rangle - y_t| \leq |\langle z_t, \widetilde{x}_t\rangle - y_t|$, it follows that

$$\begin{aligned}
\big(\langle z_t, \widetilde{x}_t\rangle - y_t\big)\big(\langle z_t, \widetilde{x}_t\rangle - \langle z_t, x_t^\star\rangle\big) &\leq |\langle z_t, \widetilde{x}_t\rangle - y_t|\big(|\langle z_t, \widetilde{x}_t\rangle - y_t| + |\langle z_t, x_t^\star\rangle - y_t|\big) && \text{(triangle inequality)} \\
&\leq 2\big(\langle z_t, \widetilde{x}_t\rangle - y_t\big)^2 \\
&= 2\big(|y_t| - |\langle z_t, \widetilde{x}_t\rangle|\big)^2 && (y_t\langle z_t, \widetilde{x}_t\rangle \geq 0) \\
&= 2\big(|y_t| - \rho_t\big)^2 && (|\langle z_t, \widetilde{x}_t\rangle| = \rho_t) \\
&< 2|y_t|^2\,. && (0 \leq \rho_t < |y_t|)
\end{aligned} \quad \text{(A.30)}$$

Given the above remarks, let $\mathcal{T}_1 = \{t \in \mathcal{T} : \ell_t(\widetilde{x}_t) \leq \ell_t(x_t)\}$ be the subset of rounds in $\mathcal{T}$ when clipping does not worsen the value of ℓ_t, and let $\mathcal{T}_2 = \mathcal{T} \setminus \mathcal{T}_1$ be the remaining rounds in $\mathcal{T}$. Then,

$$\texttt{Drift}_T \leq \sum_{t\in\mathcal{T}}\big(\langle z_t, \widetilde{x}_t\rangle - y_t\big)\big(\langle z_t, \widetilde{x}_t\rangle - \langle z_t, x_t^\star\rangle\big) \leq 2Y\sum_{t\in\mathcal{T}_1}\|z_t\|_{A_t^{-1}}\|x_t - x_t^\star\|_{A_t} + 2\sum_{t\in\mathcal{T}_2}|y_t|^2\,. \quad \text{(A.31)}$$

At this point, for any round $t \in \mathcal{T}_1$ we are interested in understanding the behavior of $\|z_t\|_{A_t^{-1}}\|x_t - x_t^\star\|_{A_t}$. Applying Lemma A.2, we have that

$$\|x_t - x_t^\star\|_{A_t}^2 \leq \frac{1}{2}\sum_{\tau\in m_t} y_\tau\langle z_\tau, x_t^\star - x_t\rangle \leq \frac{1}{2}\sum_{\tau\in m_t}|y_\tau|\|z_\tau\|_{A_t^{-1}}\|x_t^\star - x_t\|_{A_t} \leq \frac{Y}{2}\sum_{\tau\in m_t}\|z_\tau\|_{A_t^{-1}}\|x_t^\star - x_t\|_{A_t}\,,$$

where the second inequality follows by Cauchy-Schwarz, while the last one comes from Assumption 2.3. By rearranging terms in the previous inequality, we obtain that

$$\|x_t - x_t^\star\|_{A_t} \leq \frac{Y}{2} \sum_{\tau \in m_t} \|z_\tau\|_{A_t^{-1}} \,. \tag{A.32}$$

Recall that we define $d_{\max}^{\leq t} = \max_{\tau \leq t} \min\{d_\tau, t - \tau\}$ as the maximum delay that has been perceived up to round t. Hence, we can now bound the sum relative to rounds in $\mathcal{T}_1$ from above as

$$\begin{aligned} 2Y \sum_{t \in \mathcal{T}_1} \|z_t\|_{A_t^{-1}} \|x_t - x_t^\star\|_{A_t} &\leq Y^2 \sum_{t \in \mathcal{T}_1} \|z_t\|_{A_t^{-1}} \sum_{\tau \in m_t} \|z_\tau\|_{A_t^{-1}} \qquad \text{(Equation (A.32))} \\ &\leq Y^2 \sum_{t=1}^{T} \|z_t\|_{A_t^{-1}} \sum_{\tau \in m_t} \|z_\tau\|_{A_t^{-1}} \,. \end{aligned}$$

If we now adopt Lemma A.11, we have that

$$\sum_{t=1}^{T} \|z_t\|_{A_t^{-1}} \sum_{\tau \in m_t} \|z_\tau\|_{A_t^{-1}} \leq 2n d_{\max}^{\leq T} \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right) ,$$

while at the same time we have

$$\sum_{t=1}^{T} \|z_t\|_{A_t^{-1}} \sum_{\tau \in m_t} \|z_\tau\|_{A_t^{-1}} \leq Z^2 \sum_{t=1}^{T} \frac{|m_t|}{\eta_t} \,,$$

where we used the fact that $\|z_s\|_{A_t^{-1}} \leq \frac{Z^2}{\eta_t}$ for any $s \in [T]$. Thus, we have that

$$2Y \sum_{t \in \mathcal{T}_1} \|z_t\|_{A_t^{-1}} \|x_t - x_t^\star\|_{A_t} \leq Y^2 \min\left\{ 2n d_{\max}^{\leq T} \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right), Z^2 \sum_{t=1}^{T} \frac{|m_t|}{\eta_t} \right\} . \tag{A.33}$$

If we instead consider the sum over rounds in $\mathcal{T}_2$, it is possible to further bound it from above and relate it to the rounds for which the corresponding label does not belong to our estimate for the label range given by ρ_t. Indeed, if we let $\mathcal{R}eg = \{t \in [T] : |y_t| > \rho_t\}$ and given our previous remarks about $\mathcal{T}_2$, we have that $\mathcal{T}_2 \subseteq \mathcal{R}eg$. Now let $q_1 = \min\{\lceil \log_2 \rho_t \rceil : \rho_t > 0, t \in [T+1]\}$ and $q_2 = \lceil \log_2 \rho_{T+1} \rceil$. For convenience, define $\mathcal{I}_j = [2^j, 2^{j+1})$ for any $j \in \{q_1, \dots, q_2\}$. Then, for any $t \in \mathcal{R}eg$, there exists $j_t \in \{q_1, \dots, q_2\}$ such that $|y_t| \in \mathcal{I}_{j_t}$. Moreover, if we denote by $\nu_j \in [T+1]$ as the first time when $\rho_{\nu_j} \in \mathcal{I}_j$ for any $j \in \{q_1, \dots, q_2\}$, we can further show that any $t \in \mathcal{R}eg$ has to be such that $t \in m_{\nu_{j_t} - 1}$; if it were not the case, y_t would have been observed before time ν_{j_t} which is a contradiction because $|y_t| > \rho_\tau$ for any $\tau < \nu_{j_t}$. All things considered, we can derive that

$$\begin{aligned} 2 \sum_{t \in \mathcal{T}_2} |y_t|^2 &\leq 2 \sum_{t \in \mathcal{R}eg} |y_t|^2 \leq 2 \sum_{j=q_1}^{q_2} \sum_{t \in m_{\nu_j - 1}} |y_t|^2 \leq 2 \sum_{j=q_1}^{q_2} 2^{2j} |m_{\nu_j - 1}| \\ &\leq \sigma_{\max} \sum_{j=q_1}^{q_2} 2^{2j+1} \leq \frac{8}{3} \sigma_{\max} 4^{q_2} \leq \frac{32}{3} \sigma_{\max} \rho_{T+1}^2 \leq 11 Y^2 \sigma_{\max} \,. \end{aligned} \tag{A.34}$$

Combining all the results gathered so far, we can finally derive the overall regret bound as follows:

$$\begin{aligned}
\mathrm{Reg}_T(u) &\leq \mathrm{Reg}_T^\star(u) + \texttt{Drift}_T \\
&\leq \frac{\eta_T}{2}\|u\|_2^2 + nY^2 \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right) + \texttt{Drift}_T && \text{(Equation (A.27))} \\
&\leq \frac{\eta_T}{2}\|u\|_2^2 + nY^2 \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right) + 11Y^2\sigma_{\max} + 2Y \sum_{t\in\mathcal{T}_1} \|z_t\|_{A_t^{-1}} \|x_t - x_t^\star\|_{A_t} \\
& && \text{(Equations (A.31) and (A.34))} \\
&\leq \frac{\eta_T}{2}\|u\|_2^2 + nY^2 \ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right) + 11Y^2\sigma_{\max} \\
&\quad + Y^2 \min\left\{2nd_{\max}\ln\left(1 + \frac{Z^2 T}{\eta_0 n}\right), Z^2 \sum_{t=1}^{T} \frac{|m_t|}{\eta_t}\right\} && \text{(Equation (A.33))}
\end{aligned}$$

□

The following corollary is a restatement of Corollary 2.2, which shows that we can further achieve a $\mathcal{O}(\min\{d_{\max}\ln T, \sqrt{d_{\mathrm{tot}}}\})$ regret guarantee via an adaptive tuning of the learning rate of Algorithm 2.3 similar to the one adopted for Algorithm 2.2.

Corollary 2.2. *In the OLR problem with delayed labels under Assumption 2.3, Algorithm 2.3 with the adaptive learning rate $\eta_t = \gamma(\min\{a_t, b_t\} + 1)$, where a_t and b_t are defined in Equation (2.15) for any $\gamma > 0$ guarantees that*

$$\mathrm{Reg}_T \leq \frac{\gamma\|u\|_2^2}{2} + nY^2 \ln\left(1 + \frac{Z^2 T}{\gamma n}\right) + \mathcal{O}\big(\min\{Q_1, Q_2\}\big),$$

where $Q_1 = \big(\gamma\|u\|_2^2 + Y^2\big)nd_{\max}\ln\left(1 + \frac{Z^2T}{\gamma n}\right)$ and $Q_2 = \big(\gamma Z\|u\|_2^2 + (Z+1)Y^2\big)\sqrt{d_{\mathrm{tot}}}$.

Proof. By performing a similar analysis as in the proof of Theorem 2.3 up to Equation (A.33), for any time threshold $\tau^\star \in [T]$ we can actually separately analyze the time ranges $\{1, \ldots, \tau^\star\}$ and $\{\tau^\star + 1, \ldots, T\}$ in an analogous way as in the proof of Corollary 2.1, and have a bound of the following form:

$$2Y \sum_{t\in\mathcal{T}_1} \|z_t\|_{A_t^{-1}} \|x_t - x_t^\star\|_{A_t} \leq Y^2\left(2nd_{\max}^{\leq\tau^\star}\ln\left(1 + \frac{Z^2T}{\eta_0 n}\right) + Z^2 \sum_{t=\tau^\star+1}^{T} \frac{|m_t|}{\eta_t}\right). \tag{A.35}$$

Then, we use an adaptive tuning of the learning rate in a similar way as performed for the proof of Corollary 2.1. In particular, we define

$$a_t = 2nd_{\max}^{\leq t}\ln\left(1 + \frac{Z^2T}{\gamma n}\right) \quad \text{and} \quad b_t = Z\sqrt{\sum_{s=1}^{t}|m_s|}\,,$$

and, for any $\gamma > 0$, we set $\eta_0 = \gamma$ and $\eta_t = \gamma\big(\min\{a_t, b_t\} + 1\big)$ for any $t \geq 1$. First, when $a_T \leq b_T$

we have that

$$\begin{aligned}
\mathrm{Reg}_T(u) &\le \frac{\eta_T}{2}\|u\|_2^2 + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + Y^2\left(11\sigma_{\max} + 2nd_{\max}\ln\left(1+\frac{Z^2T}{\gamma n}\right)\right) \\
&\qquad\qquad \text{(Equations (Equations (A.31) and (A.34)) and (A.33))} \\
&\le \frac{\|u\|_2^2}{2}\eta_T + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + Y^2 d_{\max}\left(11 + 2n\ln\left(1+\frac{Z^2T}{\gamma n}\right)\right) \qquad (\sigma_{\max}\le d_{\max}) \\
&\le \frac{\gamma\|u\|_2^2}{2} + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 11Y^2 d_{\max} + \left(\gamma\|u\|_2^2 + 2Y^2\right)nd_{\max}\ln\left(1+\frac{Z^2T}{\gamma n}\right) \\
&\le \frac{\gamma\|u\|_2^2}{2} + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + \left(\gamma\|u\|_2^2 + 13Y^2\right)nd_{\max}\ln\left(1+\frac{Z^2T}{\gamma n}\right).
\end{aligned}$$

On the contrary, when $a_T > b_T$, we let $\tau^\star$ be the last round such that $a_{\tau^\star} \le b_{\tau^\star}$ and show that

$$\begin{aligned}
\mathrm{Reg}_T(u) &\le \frac{\|u\|_2^2}{2}\eta_T + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 11Y^2\sigma_{\max} + Y^2\left(2nd_{\max}^{\le\tau^\star}\ln\left(1+\frac{Z^2T}{\gamma n}\right) + Z^2\sum_{t=\tau^\star+1}^{T}\frac{|m_t|}{\eta_t}\right) \\
&\qquad\qquad \text{(Equations (Equations (A.31) and (A.34)) and (A.35))} \\
&\le \frac{\|u\|_2^2}{2}\eta_T + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 11Y^2\sigma_{\max} + ZY^2\left(\sqrt{\sum_{t=1}^{\tau^\star}|m_t|} + Z\sum_{t=\tau^\star+1}^{T}\frac{|m_t|}{\eta_t}\right) \\
&\qquad\qquad (a_{\tau^\star}\le b_{\tau^\star}) \\
&\le \frac{\|u\|_2^2}{2}\eta_T + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 11Y^2\sigma_{\max} + \frac{ZY^2}{\gamma}\left(\sqrt{\sum_{t=1}^{\tau^\star}|m_t|} + \sum_{t=\tau^\star+1}^{T}\frac{|m_t|}{\sqrt{\sum_{s=1}^{t}|m_s|}}\right) \\
&\qquad\qquad \text{(definition of } \eta_t) \\
&\le \frac{\|u\|_2^2}{2}\eta_T + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 11Y^2\sigma_{\max} + ZY^2\left(\sqrt{\sum_{t=1}^{\tau^\star}|m_t|} + 2\sqrt{\sum_{t=\tau^\star+1}^{T}|m_s|}\right) \\
&\qquad\qquad \text{(Orabona (2025, Lemma 4.13))} \\
&\le \frac{\|u\|_2^2}{2}\eta_T + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 11Y^2\sigma_{\max} + 2ZY^2\sqrt{2d_{\mathrm{tot}}} \\
&\le \frac{\|u\|_2^2}{2}\eta_T + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 2(11+Z)Y^2\sqrt{2d_{\mathrm{tot}}} \qquad \text{(Lemma A.7)} \\
&\le \frac{\gamma\|u\|_2^2}{2}\left(1+Z\sqrt{d_{\mathrm{tot}}}\right) + nY^2\ln\left(1+\frac{Z^2T}{\gamma n}\right) + 2(11+Z)Y^2\sqrt{2d_{\mathrm{tot}}}\,. \qquad \text{(definition of } \eta_T)
\end{aligned}$$

Considering the conditions in each of the two cases together with the definitions of a_t and b_t, this concludes the proof. □

A.5 Online mirror descent for delayed OCO with strongly convex losses

In this section, we prove that the following online mirror descent (OMD) algorithm achieves a regret guarantee whose dependence on the delays is of order $\min\{\sigma_{\max}\ln T, \sqrt{d_{\mathrm{tot}}}\}$, similarly to Algorithm 2.1. To be precise, an OMD-based algorithm which handles delays was initially proposed

by Wu et al. (2024) in their Algorithm 6. However, Wu et al. (2024) only manage to show that this algorithm achieves regret $\mathcal{O}\left(\frac{d_{\max}(L^2+D)}{\lambda}\ln T + \frac{d_{\max}G}{\lambda^2}\right)$ under Assumption 2.1 and Assumption 2.2. Here, we report its pseudocode in Algorithm A.1 and we provide an improved regret analysis for it. Not only do we provide a significantly better guarantee, but we also manage to lift Assumption 2.2 and only require the boundedness of the gradient norms via Assumption 2.1. The key to achieve these improvements simultaneously is a fundamentally different and more careful regret analysis.

Algorithm A.1: Delayed OMD for strongly convex functions

input: strong convexity parameter $\lambda > 0$, learning rates $\eta_t = \frac{2}{t\lambda}$ for all $t \in [T]$
initialize: $x_1 \in \mathcal{X}$
for $t = 1, 2, \dots$ **do**
    Play x_t
    Receive $g_\tau = \nabla \ell_\tau(x_\tau)$ for all $\tau \in o_{t+1} \setminus o_t$
    Update $x_{t+1} = \arg\min_{x \in \mathcal{X}} \sum_{\tau \in o_{t+1} \setminus o_t} \langle g_\tau, x \rangle + \frac{1}{\eta_t}\|x - x_t\|_2^2$.

Theorem A.1. *Assume that $\ell_1, \dots, \ell_T$ are λ-strongly convex functions with respect to the Euclidean norm $\|\cdot\|_2$. Then, under Assumption 2.1, Algorithm A.1 guarantees*

$$\mathrm{Reg}_T = \mathcal{O}\left(\frac{L^2}{\lambda}\left(\ln T + \min\left\{\sigma_{\max}\ln T, \sqrt{d_{\mathrm{tot}}}\right\}\right)\right) .$$

Proof. We begin with a decomposition of the regret that, similarly to the proof of Theorem 2.1, leverages the strong convexity of losses $\ell_1, \dots, \ell_T$ and attempts to isolate the discrepancy in the information available to the learner because of the delayed gradients. However, this decomposition differs from the one in Theorem 2.1 since the algorithm updates its predictions differently via mirror descent. Our approach follows the idea of framing such an information discrepancy via optimism (Flaspohler et al., 2021). For notational convenience, define $\widetilde{g}_1 = 0$ and $\widetilde{g}_{t+1} = \widetilde{g}_t + \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau - g_t$ for any $t \geq 1$. Note that, by definition, each $\widetilde{g}_t$ is equal to

$$\widetilde{g}_t = \sum_{\tau=1}^{t-1}\left(\widetilde{g}_{\tau+1} - \widetilde{g}_\tau\right) = \sum_{s=1}^{t-1}\left(\sum_{\tau \in o_{s+1} \setminus o_s} g_\tau - g_s\right) = \sum_{s \in o_t} g_s - \sum_{s=1}^{t-1} g_s = -\sum_{s \in m_t} g_s \tag{A.36}$$

and consequently $\widetilde{g}_{T+1} = 0$ since $m_{T+1} = \emptyset$. This definition of $\widetilde{g}_t$ allows to rewrite the "linearized" regret as

$$\sum_{t=1}^{T}\langle g_t, x_t - u\rangle = \sum_{t=1}^{T}\left\langle \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau, x_t - u \right\rangle + \sum_{t=1}^{T}\langle \widetilde{g}_t - \widetilde{g}_{t+1}, x_t\rangle \tag{A.37}$$

and to have that, for every round t,

$$\left\langle \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau, x_t - x_{t+1} \right\rangle = \langle g_t - \widetilde{g}_t + \widetilde{g}_{t+1}, x_t - x_{t+1}\rangle = \langle g_t - \widetilde{g}_t, x_t - x_{t+1}\rangle + \langle \widetilde{g}_{t+1}, x_t - x_{t+1}\rangle \ . \tag{A.38}$$

Moreover, according to the standard regret analysis of OMD (Lemma A.4), we know that

$$\left\langle \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau, x_t - u \right\rangle \leq \frac{1}{\eta_t}\Big(\|u - x_t\|_2^2 - \|u - x_{t+1}\|_2^2 - \|x_t - x_{t+1}\|_2^2 \Big) + \left\langle \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau, x_t - x_{t+1} \right\rangle . \tag{A.39}$$

The above observations then make it possible to bound the first sum in the right-hand side of Equation (A.37) as

$$\begin{aligned}
\sum_{t=1}^{T} \left\langle \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau, x_t - u \right\rangle &\leq \sum_{t=1}^{T} \frac{1}{\eta_t}\Big(\|u - x_t\|_2^2 - \|u - x_{t+1}\|_2^2 - \|x_t - x_{t+1}\|_2^2 \Big) \\
&\quad + \sum_{t=1}^{T} \left\langle \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau, x_t - x_{t+1} \right\rangle && \text{(Equation (A.39))} \\
&= \sum_{t=1}^{T} \frac{1}{\eta_t}\Big(\|u - x_t\|_2^2 - \|u - x_{t+1}\|_2^2 - \|x_t - x_{t+1}\|_2^2 \Big) \\
&\quad + \sum_{t=1}^{T} \langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle + \sum_{t=1}^{T} \langle \widetilde{g}_{t+1}, x_t - x_{t+1} \rangle && \text{(Equation (A.38))} \\
&= \sum_{t=1}^{T} \frac{1}{\eta_t}\Big(\|u - x_t\|_2^2 - \|u - x_{t+1}\|_2^2 - \|x_t - x_{t+1}\|_2^2 \Big) \\
&\quad + \sum_{t=1}^{T} \langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle + \sum_{t=1}^{T} \langle \widetilde{g}_{t+1} - \widetilde{g}_t, x_t \rangle \\
&\quad + \langle \widetilde{g}_1, x_1 \rangle - \langle \widetilde{g}_{T+1}, x_{T+1} \rangle \\
&= \sum_{t=1}^{T} \frac{1}{\eta_t}\Big(\|u - x_t\|_2^2 - \|u - x_{t+1}\|_2^2 - \|x_t - x_{t+1}\|_2^2 \Big) \\
&\quad + \sum_{t=1}^{T} \langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle + \sum_{t=1}^{T} \langle \widetilde{g}_{t+1} - \widetilde{g}_t, x_t \rangle , && \text{(A.40)}
\end{aligned}$$

where the second equality follows by carefully rearranging the terms in the sum $\sum_{t=1}^{T} \langle \widetilde{g}_{t+1}, x_t - x_{t+1} \rangle$, while the last equality is due to $\widetilde{g}_1 = \widetilde{g}_{T+1} = 0$ by definition.

At this point, we can rewrite the regret in the following way:

$$\begin{aligned}
\mathrm{Reg}_T(u) &= \sum_{t=1}^{T} \big(\ell_t(x_t) - \ell_t(u) \big) \\
&\leq \sum_{t=1}^{T} \langle g_t, x_t - u \rangle - \frac{\lambda}{2} \sum_{t=1}^{T} \|x_t - u\|_2^2 \\
&= \sum_{t=1}^{T} \left\langle \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau, x_t - u \right\rangle + \sum_{t=1}^{T} \langle \widetilde{g}_t - \widetilde{g}_{t+1}, x_t \rangle - \frac{\lambda}{2} \sum_{t=1}^{T} \|x_t - u\|_2^2 && \text{(Equation (A.37))} \\
&\leq \sum_{t=1}^{T} \frac{\|u - x_t\|_2^2 - \|u - x_{t+1}\|_2^2 - \|x_t - x_{t+1}\|_2^2}{\eta_t} + \sum_{t=1}^{T} \langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle - \frac{\lambda}{2} \sum_{t=1}^{T} \|x_t - u\|_2^2 \\
& && \text{(Equation (A.40))}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{T} \left(\frac{\|u - x_t\|_2^2 - \|u - x_{t+1}\|_2^2}{\eta_t} - \frac{\lambda}{2} \sum_{t=1}^{T} \|x_t - u\|_2^2 \right) + \sum_{t=1}^{T} \left(\langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle - \frac{\|x_t - x_{t+1}\|_2^2}{\eta_t} \right) \\
&= \frac{\lambda}{2} \sum_{t=1}^{T} \Big(\big(\|x_t - u\|_2^2 - \|x_{t+1} - u\|_2^2 \big) t - \|x_t - u\|_2^2 \Big) + \sum_{t=1}^{T} \left(\langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle - \frac{\|x_t - x_{t+1}\|_2^2}{\eta_t} \right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(definition of } \eta_t\text{)} \\
&= -\frac{\lambda T}{2} \|x_{T+1} - u\|_2^2 + \sum_{t=1}^{T} \left(\langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle - \frac{\|x_t - x_{t+1}\|_2^2}{\eta_t} \right) \\
&\leq \sum_{t=1}^{T} \left(\langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle - \frac{\|x_t - x_{t+1}\|_2^2}{\eta_t} \right) , \qquad \text{(A.41)}
\end{aligned}
$$

where the first inequality holds because of the λ-strong convexity of ℓ_t.

We now focus on the right-hand side of Equation (A.41). Applying Lemma A.2, we can bound from above the distance between subsequent iterates:

$$
\|x_t - x_{t+1}\|_2 \leq \eta_t \|g_t + \widetilde{g}_{t+1} - \widetilde{g}_t\|_2 = \eta_t \left\| \sum_{\tau \in o_{t+1} \setminus o_t} g_\tau \right\|_2 \leq L \eta_t \big(|o_{t+1}| - |o_t| \big) , \qquad \text{(A.42)}
$$

where the last inequality follows by jointly using the triangle inequality, the bound on the gradient norm (Assumption 2.1), and the fact that $o_t \subseteq o_{t+1}$.

What remains to analyze now is the distance $\|g_t - \widetilde{g}_t\|_2$, and a direct calculation allows us to show that

$$
\|g_t - \widetilde{g}_t\|_2 = \left\| g_t + \sum_{\tau \in m_t} g_\tau \right\|_2 \leq L(|m_t| + 1) , \qquad \text{(A.43)}
$$

again by using the triangle inequality and Assumption 2.1.

Applying Lemma A.5 with Equation (A.42), we show that the each term of the sum in the right-hand side of Equation (A.41) satisfies

$$
\langle g_t - \widetilde{g}_t, x_t - x_{t+1} \rangle - \frac{\|x_t - x_{t+1}\|_2^2}{\eta_t} \leq \min \left\{ L \eta_t \|g_t - \widetilde{g}_t\|_2 (|o_{t+1}| - |o_t|) , \eta_t \|g_t - \widetilde{g}_t\|_2^2 \right\} . \qquad \text{(A.44)}
$$

Therefore, starting from Equation (A.41), we are able to derive the final regret bound:

$$
\begin{aligned}
\mathrm{Reg}_T(u) &\leq \sum_{t=1}^{T} \eta_t \|g_t - \widetilde{g}_t\|_2 \|g_t + \widetilde{g}_{t+1} - \widetilde{g}_t\|_2 && \text{(Equations (A.41) and (A.44))} \\
&= \frac{2L}{\lambda} \sum_{t=1}^{T} \frac{\|g_t - \widetilde{g}_t\|_2 (|o_{t+1}| - |o_t|)}{t} && \text{(definition of } \eta_t\text{)} \\
&\leq \frac{2L^2}{\lambda} \sum_{t=1}^{T} \frac{(|m_t| + 1)(|o_{t+1}| - |o_t|)}{t} . && \text{(Equation (A.43))}
\end{aligned}
$$

Crucially, what remains to analyze is the sum in the right-hand side of the above inequality. We can

first show that

$$\begin{aligned}
\sum_{t=1}^{T} \frac{(|m_t|+1)(|o_{t+1}|-|o_t|)}{t} &\leq (\sigma_{\max}+1)\sum_{t=1}^{T}\frac{|o_{t+1}|-|o_t|}{t} && \text{(definition of } \sigma_{\max}) \\
&\leq (\sigma_{\max}+1)\sum_{t=1}^{T}\frac{|o_{t+1}|-|o_t|}{|o_{t+1}|} && (o_{t+1}\subseteq[t]) \\
&= (\sigma_{\max}+1)\sum_{t=1}^{T}\frac{(|o_{t+1}|-|o_t|)}{\sum_{s=1}^{t}(|o_{s+1}|-|o_s|)} \\
&\leq (\sigma_{\max}+1)(1+\ln T)\,, && \text{(A.45)}
\end{aligned}$$

where the last inequality follows by Orabona (2025, Lemma 4.13) and the fact that $\sum_{t=1}^{T}(|o_{t+1}|-|o_t|) = |o_{T+1}| = T$. Second, we can also bound such a sum in an alternative way:

$$\begin{aligned}
\sum_{t=1}^{T} \frac{(|m_t|+1)(|o_{t+1}|-|o_t|)}{t} &= \sum_{t=1}^{T}\frac{|m_t|(|o_{t+1}|-|o_t|)}{t} + \sum_{t=1}^{T}\frac{(|o_{t+1}|-|o_t|)}{t} \\
&\leq \sum_{t=1}^{T}\frac{|m_t|(|o_{t+1}|-|o_t|)}{t} + \sum_{t=1}^{T}\frac{(|o_{t+1}|-|o_t|)}{\sum_{s=1}^{t}(|o_{s+1}|-|o_s|)} && \text{(definition of } o_t) \\
&\leq \sum_{t=1}^{T}\frac{|m_t|(|o_{t+1}|-|o_t|)}{t} + \ln T + 1 \\
&\leq \sum_{t=1}^{T}\frac{|m_t|(t-|m_{t+1}|-(t-1-|m_t|))}{t} + \ln T + 1 \\
& && (|o_t|+|m_t| = t-1 \text{ for all } t) \\
&= \sum_{t=1}^{T}\frac{|m_t|(1+|m_t|-|m_{t+1}|)}{t} + \ln T + 1 \\
&= \sum_{t=1}^{T}\frac{|m_t|}{t} + |m_1|^2 - \frac{|m_T||m_{T+1}|}{t} + \sum_{t=2}^{T}\left(\frac{|m_t|^2}{t} - \frac{|m_{t-1}||m_t|}{t-1}\right) + \ln T + 1 \\
&= \sum_{t=1}^{T}\frac{|m_t|}{t} + \sum_{t=2}^{T}\left(\frac{|m_t|^2}{t} - \frac{|m_{t-1}||m_t|}{t-1}\right) + \ln T + 1 \quad \text{(definition of } m_t) \\
&\leq \sum_{t=1}^{T}\frac{|m_t|}{t} + \sum_{t=2}^{T}\left(\frac{(|m_{t-1}|+1)|m_t|}{t-1} - \frac{|m_{t-1}||m_t|}{t-1}\right) + \ln T + 1 \\
& && (m_{t+1}\subseteq m_t\cup\{t\} \text{ for all } t) \\
&\leq \sum_{t=1}^{T}\frac{|m_t|}{t} + \ln T + 1 \\
&\leq 2\sqrt{d_{\text{tot}}} + \ln T + 1\,,
\end{aligned}$$

where the last inequality follows by Equation (A.9). Combing the above two inequalities, we finally obtain

$$\mathrm{Reg}_T(u) \leq \frac{2L^2}{\lambda}(1+\ln T) + \frac{2L^2}{\lambda}\min\left\{\sigma_{\max}(1+\ln T)\,, 2\sqrt{d_{\text{tot}}}\right\}$$

$$= \mathcal{O}\left(\frac{L^2}{\lambda}\left(\ln T + \min\left\{\sigma_{\max}\ln T, \sqrt{d_{\text{tot}}}\right\}\right)\right) . \qquad \square$$

A.6 Additional experiments

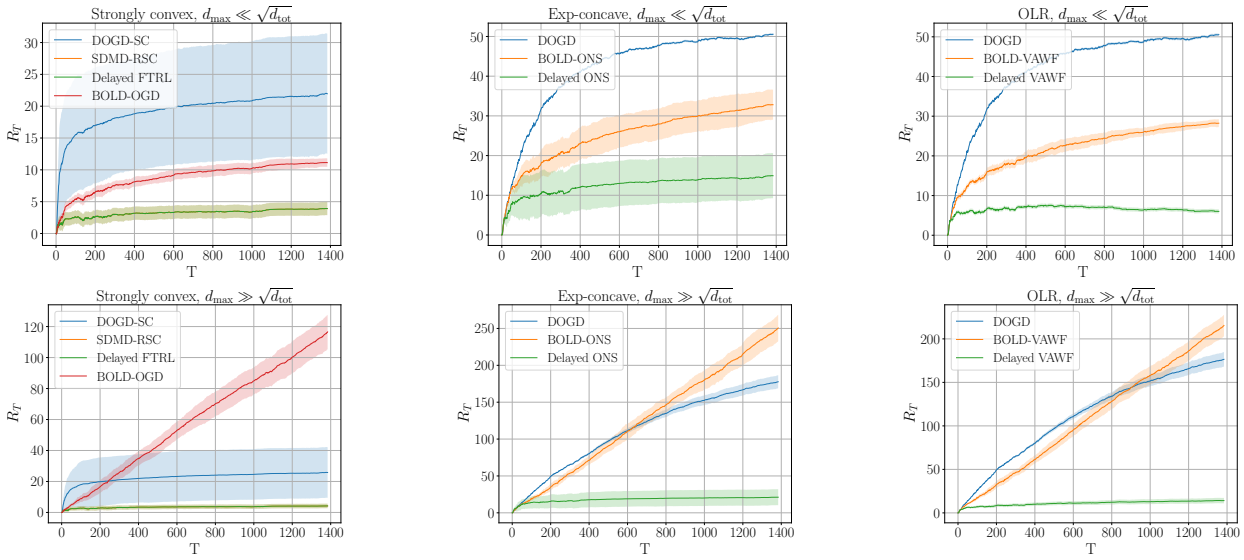


Figure A.1: Comparison with relevant baselines. The shaded areas consider a range centered around the mean with half-width corresponding to the empirical standard deviation over 20 repetitions.

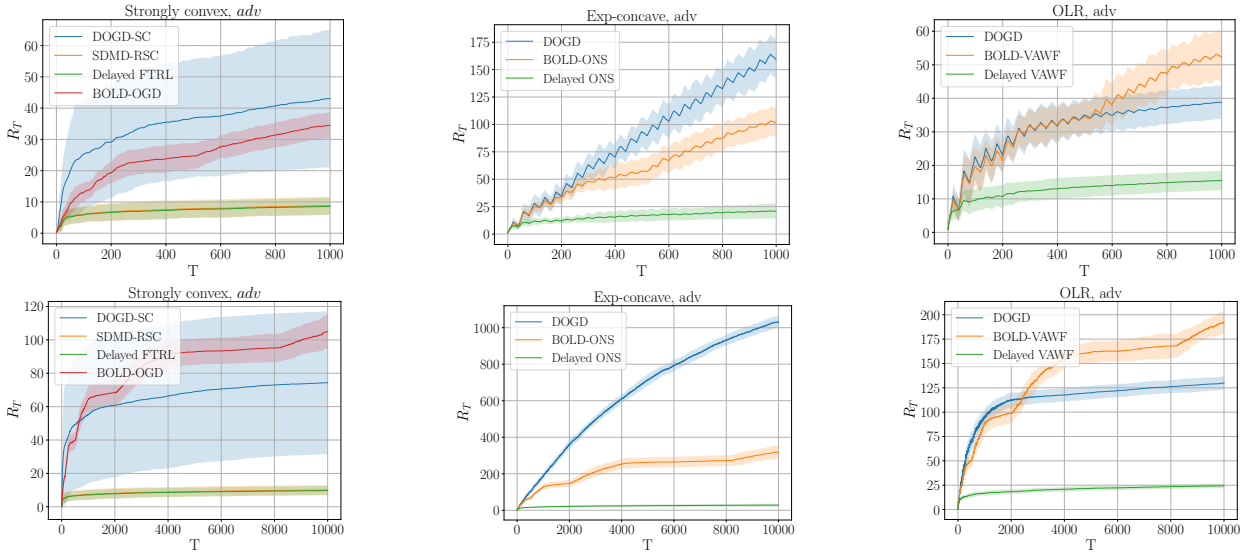


Figure A.2: Comparison with relevant baselines. The shaded areas consider a range centered around the mean with half-width corresponding to the empirical standard deviation over 20 repetitions. The top plots correspond to $T = 1000$, while the bottom plots correspond to $T = 10000$.

We consider a real-world dataset *mg_scale* from the LIBSVM repository (Chang and Lin, 2011). This dataset has 1385 samples and each sample has 6 features with values in $[-1, 1]$ and a label in $[0, 2]$. The experimental setup, including constructions of losses and delays, follows what already done for the experiments in Section 2.6. Figure A.1 shows a similar behaviour of the algorithms as already shown in Section 2.6.

We also designed a non-stationary environment as follows. The generation processes for the feature vectors, as well as the definition of the loss function, remain the same as the environment in Section 2.6. However, we modified the generation of the label y_t:

$$y_t = \langle z_t, \theta_t \rangle + \epsilon_t \ , \tag{A.46}$$

where the latent vector θ_t alternates every 30 rounds between the two vectors $\mathbf{1}$ and $\mathbf{0}$. This periodic change introduces non-stationarity, reflecting scenarios where the optimal action shifts over time. The delay d_t is independently sampled from a distribution that alternates every 30 rounds between a geometric distribution with success probability $T^{-1/3}$ and a uniform distribution over the set $\{0, 1, \ldots, 5\}$. Additionally, we also modify the noise term ϵ_t inspired by Xu and Zeevi (2023). Specifically, we flatten an abstract art piece by Jackson Pollock and take consecutive grayscale values in $[0, 1]$ as the noise ϵ_t. Figure A.2 shows that our algorithms again perform the best among all the benchmark algorithms.

Appendix B

Proof Details for Chapter 3

B.1 Auxiliary Results

Lemma B.1. *Consider any algorithm that picks actions $(A_t)_{t\in[T]}$ in the adversarial delayed bandits problem with intermediate feedback with arbitrary action-state mappings $(s_t)_{t\in[T]}$ and i.i.d. loss vectors $(\ell_t)_{t\in[T]}$. Then, for any given $\delta \in (0,1)$,*

$$\mathrm{Reg}_T - \mathcal{R}eg_T \leq \sqrt{2T\ln(2/\delta)} \qquad \textit{and} \qquad \mathcal{R}eg_T - \mathrm{Reg}_T \leq \sqrt{2T\ln(2K/\delta)}$$

individually hold with probability at least $1-\delta$.

Proof. First, observe that we can relate the two notions of regret as

$$\mathrm{Reg}_T = \mathcal{R}eg_T + \sum_{t=1}^{T}\big(\theta(S_t) - \ell_t(S_t)\big) + \underbrace{\min_{a\in\mathcal{A}}\sum_{t=1}^{T}\ell_t(s_t(a)) - \min_{a\in\mathcal{A}}\sum_{t=1}^{T}\theta(s_t(a))}_{(\triangle)} \ .$$

By Azuma-Hoeffding inequality, we can show that each side of

$$-\sqrt{\frac{T}{2}\ln\frac{1}{\delta'}} \leq \sum_{t=1}^{T}\big(\theta(S_t) - \ell_t(S_t)\big) \leq \sqrt{\frac{T}{2}\ln\frac{1}{\delta'}} \tag{B.1}$$

holds with probability at least $1-\delta'$. Now, define

$$a^*_\ell \in \operatorname*{arg\,min}_{a\in\mathcal{A}}\sum_{t=1}^{T}\ell_t(s_t(a)) \qquad \text{and} \qquad a^*_\theta \in \operatorname*{arg\,min}_{a\in\mathcal{A}}\sum_{t=1}^{T}\theta(s_t(a)) \ .$$

On the one hand, observe that

$$(\triangle) \leq \sum_{t=1}^{T}\ell_t(s_t(a^*_\theta)) - \sum_{t=1}^{T}\theta(s_t(a^*_\theta)) \leq \sqrt{\frac{T}{2}\ln\frac{1}{\delta'}} \ ,$$

where the last inequality holds with probability at least $1-\delta'$ by Azuma-Hoeffding inequality. On

the other hand, we can show that

$$(\triangle) \geq \sum_{t=1}^{T} \ell_t(s_t(a_\ell^*)) - \sum_{t=1}^{T} \theta(s_t(a_\ell^*)) =: (\diamond) \, .$$

However, in this case a_ℓ^* depends on the entire sequence $\ell_1, \ldots, \ell_T$. We thus need to use a union bound in order to show that

$$\mathbb{P}\left((\diamond) \leq -\sqrt{\frac{T}{2}\ln\frac{K}{\delta'}}\right) \leq \sum_{a\in\mathcal{A}} \mathbb{P}\left(\sum_{t=1}^{T} \ell_t(s_t(a)) - \sum_{t=1}^{T} \theta(s_t(a)) \leq -\sqrt{\frac{T}{2}\ln\frac{K}{\delta'}}\right) \leq \delta' \, ,$$

where the last inequality follows by Azuma-Hoeffding inequality. We conclude the proof by setting $\delta' = \delta/2$. □

Lemma B.2. *The estimates $(\widehat{\theta}_t)_{t=1}^T$ defined in Equation (3.3) are such that $|\widehat{\theta}_t(s) - \theta(s)| \leq \frac{1}{2}\varepsilon_t(s)$ simultaneously holds for all $t \in [T]$ and all $s \in \mathcal{S}$ with probability at least $1 - \delta/2$.*

Proof. In a similar way as in Vernade et al. (2020), define $X_m(s)$ to be the empirical mean estimate for $\theta(s)$ which uses the first $m \in [T]$ observed losses corresponding to state $s \in \mathcal{S}$. Notice that $\widehat{\theta}_t(s) = X_{N_t'(s)}(s)$, while we define $\varepsilon_m'(s) \coloneqq \sqrt{\frac{2}{m}\ln\frac{4ST}{\delta}}$ so that $\varepsilon_t(s) = \varepsilon'_{N_t'(s)}(s)$. We can additionally observe that $\mathbb{E}[X_m(s)] = \theta(s)$. Then, we can use Azuma-Hoeffding inequality to show that

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{s\in\mathcal{S}}\bigcap_{t\in[T]}\left\{|\widehat{\theta}_t(s) - \theta(s)| \leq \frac{1}{2}\varepsilon_t(s)\right\}\right) &\geq \mathbb{P}\left(\bigcap_{s\in\mathcal{S}}\bigcap_{m\in[T]}\left\{|X_m(s) - \theta(s)| \leq \frac{1}{2}\varepsilon_m'(s)\right\}\right) \\
&\geq 1 - 2\sum_{s\in\mathcal{S}}\sum_{m=1}^{T} e^{-\frac{1}{2}\varepsilon_m'(s)^2 m} \\
&= 1 - \frac{\delta}{2} \, ,
\end{aligned}$$

where we also used a union bound in the second inequality. □

Lemma B.3. *Consider any algorithm that picks actions $(A_t)_{t\in[T]}$ in the BIO setting with adversarial action-state mappings $(s_t)_{t\in[T]}$ and stochastic loss vectors $(\ell_t)_{t\in[T]}$. Assume that the losses for any fixed state are i.i.d., whereas pairs of losses $\ell_j(s), \ell_{j'}(s')$ of distinct states $s \neq s'$ might be correlated when $j > j'$ and $j - j' \leq d_{j'}$. Then, it holds that $\mathbb{E}[\mathrm{Reg}_T] \leq \mathbb{E}[\mathcal{R}eg_T]$, where the expectation is with respect to the stochasticity of the losses and the randomness of the algorithm.*

Proof. We know that $\mathbb{E}[\ell_t(s_t(a))] = \theta(s_t(a))$ for any fixed $a \in \mathcal{A}$ and all $t \in [T]$. We further observe that

$$\mathbb{E}[\ell_t(S_t)] = \mathbb{E}\Big[\mathbb{E}[\ell_t(s_t(A_t)) \mid A_t]\Big] = \mathbb{E}[\theta(S_t)]$$

holds for all $t \in [T]$, as A_t is independent of losses that can be correlated with ℓ_t. Now, define

$$a_\ell^* \in \arg\min_{a\in\mathcal{A}} \sum_{t=1}^{T} \ell_t(s_t(a)) \qquad \text{and} \qquad a_\theta^* \in \arg\min_{a\in\mathcal{A}} \sum_{t=1}^{T} \theta(s_t(a)) \, .$$

Then, we conclude the proof by showing that

$$\begin{aligned}\mathbb{E}\left[\mathcal{R}eg_T\right] &= \sum_{t=1}^{T}\mathbb{E}\left[\ell_t(S_t)\right] - \mathbb{E}\left[\sum_{t=1}^{T}\ell_t(s_t(a_\ell^*))\right] \\ &\geq \sum_{t=1}^{T}\mathbb{E}\left[\ell_t(S_t)\right] - \mathbb{E}\left[\sum_{t=1}^{T}\ell_t(s_t(a_\theta^*))\right] = \sum_{t=1}^{T}\mathbb{E}\left[\theta(S_t)\right] - \sum_{t=1}^{T}\theta(s_t(a_\theta^*)) = \mathbb{E}\left[\mathrm{Reg}_T\right] \ .\end{aligned}$$

□

B.2 Omitted Details in Section 3.4

B.2.1 Total Effective Delay Bound

Lemma 3.1 (Total effective delay)**.** *If* `MetaBIO` *is run with any algorithm* $\mathcal{B}$ *on delays* $(d_t)_{t\in[T]}$*, then its total effective delay is* $\widetilde{d}_{\mathrm{tot}} \leq d_\Phi$*.*

Proof of Lemma 3.1. For any $s \in \mathcal{S}$, we define $\mathcal{T}_s \coloneqq \{t \in [T] : S_t = s\}$ to be the set of all rounds when the state observed by the learner corresponds to s. Denote by t_s the last time step $t \in \mathcal{T}_s$ such that $N_t(s) < \sigma_t$ and let $\mathcal{C}_s \coloneqq \{t \in \mathcal{T}_s : t \leq t_s\}$ be those rounds in $\mathcal{T}_s$ that come no later than t_s. According to the choice of t_s, all the rounds in $\mathcal{T}_s$ for which learner waits for the respective delayed loss, must belong to $\mathcal{C}_s$, while the learner incurs $\widetilde{d}_t = 0$ delay for rounds $t \in \mathcal{T}_s \setminus \mathcal{C}_s$. Now we partition $\mathcal{C}_s$ into two sets: the observed set $\mathcal{C}_s^{\mathrm{obs}} \coloneqq \{t \in \mathcal{C}_s : t + d_t \leq t_s\}$ and the outstanding set $\mathcal{C}_s^{\mathrm{out}} \coloneqq \{t \in \mathcal{C}_s : t + d_t > t_s\}$. From the choice of t_s, we can see that the number of rounds in $\mathcal{C}_s^{\mathrm{obs}}$ is

$$|\mathcal{C}_s^{\mathrm{obs}}| \leq N_{t_s}(s) < \sigma_{t_s} \leq \sigma_{\max} \ ,$$

and the number of rounds in C_s^{out} is

$$|\mathcal{C}_s^{\mathrm{out}}| \leq \sigma_{t_s} \leq \sigma_{\max} \ .$$

Therefore, we have $|\mathcal{C}_s| \leq 2\sigma_{\max}$. So if we define $\mathcal{C}_{\mathrm{all}} \coloneqq \bigcup_{s\in\mathcal{S}} \mathcal{C}_s$, then $|\mathcal{C}_{\mathrm{all}}| \leq \min\{2S\sigma_{\max}, T\} = |\Phi|$. This also implies that

$$\sum_{t=1}^{T}\widetilde{d}_t \leq \sum_{t\in\mathcal{C}_{\mathrm{all}}} d_t \leq \sum_{t\in\Phi} d_t$$

by definition of Φ. □

B.2.2 Improved Regret for `DAda-Exp3` for Fixed δ

We follow the analysis of Theorem 4.1 in György and Joulani (2021, Appendix A) and our goal is to use the knowledge of $\delta \in (0, 1)$ to tune the learning rates $(\eta_t)_{t\in[T]}$ and the implicit exploration terms $(\gamma_t)_{t\in[T]}$, accordingly. Let $d_1, \ldots, d_T$ be the sequence of delays perceived by `DAda-Exp3`, and let $d_{\mathrm{tot}} \coloneqq \sum_{t=1}^{T} d_t$ be its total delay. Furthermore, let σ_t be the number of outstanding observations of `DAda-Exp3` at the beginning of round $t \in [T]$. Suppose that we take $\gamma_t = c\eta_t$ with $c > 0$ for all $t \in [T]$, then following the same analysis as in György and Joulani (2021, Appendix A), we end up

with the following regret bound that holds with probability at least $1 - 2\delta'$ for any $\delta' \in (0, 1/2)$:

$$\begin{aligned}\mathcal{R}eg_T &\leq \frac{\ln K}{\eta_T} + \sum_{t=1}^{T} \eta_t(\sigma_t + (c+1)K) + \frac{\ln(K/\delta')}{2c\eta_T} + \frac{\sigma_{\max} + c + 1}{2c} \ln(1/\delta') \\ &= \frac{1}{\eta_T}\left(\ln K + \frac{\ln(K/\delta')}{2c}\right) + \sum_{t=1}^{T} \eta_t(\sigma_{t-1} + (c+1)K) + \frac{\sigma_{\max} + 1}{2c} \ln(1/\delta') + \frac{\ln(1/\delta')}{2} \,.\end{aligned}$$

Therefore, by taking $\eta_t^{-1} = \sqrt{\frac{(c+1)Kt + \sum_{j=1}^{t} \sigma_j}{2\ln(K) + \frac{1}{c}\ln(K/\delta')}}$, we get the following bound with probability at least $1 - 2\delta'$:

$$\mathcal{R}eg_T \leq 2\sqrt{\left((c+1)KT + \sum_{t=1}^{T} \sigma_t\right)\left(2\ln(K) + \frac{\ln(K/\delta')}{c}\right)} + \frac{\sigma_{\max} + 1}{2c} \ln(1/\delta') + \frac{\ln(1/\delta')}{2} \,.$$

We know that $\sum_{t=1}^{T} \sigma_t = d_{\text{tot}}$ by definition of σ_t. Then, we can set $c = 1$ to obtain that the regret $\mathcal{R}eg_T$ (as per the original notion of regret used in György and Joulani (2021)) is

$$\mathcal{R}eg_T \leq 2\sqrt{2KT\left(3\ln(K) + \ln\left(1/\delta'\right)\right)} + 2\sqrt{d_{\text{tot}}\left(3\ln(K) + \ln\left(1/\delta'\right)\right)} + \frac{\sigma_{\max} + 2}{2} \ln(1/\delta') \tag{B.2}$$

with probability at least $1 - 2\delta'$.

From Lemma B.1, we have that

$$\text{Reg}_T \leq \mathcal{R}eg_T + \sqrt{2T\ln(2/\delta')} \tag{B.3}$$

holds with probability at least $1 - \delta'$. So, combining Equations (B.2) and (B.3), and setting $\delta \coloneqq 3\delta'$, we can upper bound our notion of regret Reg_T as

$$\text{Reg}_T \leq 2\sqrt{2KT\left(3\ln(K) + \ln\left(3/\delta\right)\right)} + \sqrt{2T\ln(6/\delta)} + 2\sqrt{d_{\text{tot}}\left(3\ln(K) + \ln\left(3/\delta\right)\right)} + \frac{\sigma_{\max} + 2}{2} \ln(3/\delta) \tag{B.4}$$

with probability at least $1 - \delta$.

B.2.3 Reduction to `DAda-Exp3` via `MetaBIO`

Based on the reduction via `MetaBIO`, we require that $\mathcal{B}$ guarantee a regret bound

$$\widehat{\mathcal{R}eg}_T^{\mathcal{B}} = \sum_{t=1}^{T} \widetilde{\theta}_t(S_t) - \min_{a \in \mathcal{A}} \sum_{t=1}^{T} \widetilde{\theta}_t(s_t(a)) \tag{B.5}$$

that holds with high probability when the losses experienced by $\mathcal{B}$ are of the form $\widetilde{\theta}_t\big(s_t(a)\big)$. Note that, even though the action-state mappings $s_1, \ldots, s_T$ are unknown to the learner, we can provide those losses as long as $\mathcal{B}$ requires bandit feedback only. Indeed, we can compute $\widetilde{\theta}_t(S_t)$ defined in Equations (3.1) and (3.3), while we cannot determine $s_t(a)$ for all actions $a \in \mathcal{A}$ that are not A_t. As mentioned in Section 3.4, in this work we consider `DAda-Exp3` (György and Joulani, 2021) as algorithm $\mathcal{B}$ used by `MetaBIO`. In what follows, we refer to this specific choice for the algorithm $\mathcal{B}$.

The analysis of `DAda-Exp3` for the high-probability bound (Theorem 3.1) is such that most steps

only require that the loss of each action is bounded in $[0,1]$. Then, those steps apply for any such sequence of loss vectors. However, the crucial part of that analysis that requires attention is the application of Lemma 1 from Neu (2015). We restate it below for reference.

Before that, we introduce the notation required for stating the result. We consider a learner choosing actions $A_1,\dots,A_T$ according to probability distributions $p_1,\dots,p_T$ over actions. We denote by $\mathcal{F}_{t-1}$ the observation history of the learner until the beginning of round t. The result uses importance-weighted estimates for the losses $\ell_1,\dots,\ell_T$ with implicit exploration, where the implicit exploration parameter is $\gamma_t \geq 0$ for each time t. These loss estimates are defined as

$$\widetilde{\ell}_t(a) = \frac{\mathbb{I}\{A_t = a\}}{p_t(a) + \gamma_t}\ell_t(a) \qquad \forall t \in [T], \forall a \in \mathcal{A}\,. \tag{B.6}$$

Lemma B.4 (Neu (2015, Lemma 1))**.** *Let γ_t and $\alpha_t(a)$ be nonnegative $\mathcal{F}_{t-1}$-measurable random variables such that $\alpha_t(a) \leq 2\gamma_t$, for all $t \in [T]$ and all $a \in \mathcal{A}$. Let $\widetilde{\ell}_t(a)$ be as in* (B.6)*. Then,*

$$\sum_{t=1}^{T}\sum_{a=1}^{K}\alpha_t(a)\big(\widetilde{\ell}_t(a) - \ell_t(a)\big) \leq \ln(1/\delta)$$

holds with probability at least $1-\delta$ for any $\delta \in (0,1)$.

In our case, we require an analogous result that work when loss vectors correspond with our estimates $\widetilde{\theta}_1,\dots,\widetilde{\theta}_T$. However, these estimate have a dependency with the past actions chosen by the learner. This requires some nontrivial changes in the proof of Neu (2015, Lemma 1).

Before that, we introduce some crucial definitions for this proof. Let $\rho(t) \coloneqq t + d_t$ be the arrival time for the realized loss $\ell_t(S_t)$ of the state S_t observed at time $t \in [T]$. Let $\widetilde{\rho}(t) \coloneqq t + \widetilde{d}_t$ be instead the arrival time perceived by algorithm $\mathcal{B}$ relative to its choice of A_t at time t, i.e., when $\mathcal{B}$ receives $\widetilde{\theta}_t(S_t)$. This also means that $\widetilde{\theta}_t(S_t)$ is only defined at time $\widetilde{\rho}(t) \leq \rho(t)$.

Let $\pi\colon [T] \to [T]$ be the permutation of $[T]$ that orders rounds according to their value of $\widetilde{\rho}$. In other words, π satisfies the following property:

$$\pi(r) < \pi(t) \quad \iff \quad \widetilde{\rho}(r) < \widetilde{\rho}(t) \vee (\widetilde{\rho}(r) = \widetilde{\rho}(t) \wedge r < t) \qquad \forall r,t \in [T]\,. \tag{B.7}$$

This permutation allows us to sort rounds according to the order in which `MetaBIO` feeds $\mathcal{B}$ with a respective estimate for the mean loss. In particular, the r-th round in this order corresponds with the round $t_r \coloneqq \pi^{-1}(r)$, for any $r \in [T]$. Hence, we can equivalently define the round t_r as the round such that its estimate $\widetilde{\theta}_{t_r}(S_{t_r})$ for the mean loss $\theta(S_{t_r})$ is the r-th estimate received by $\mathcal{B}$.

Define

$$\mathcal{F}_r \coloneqq \{(j, A_j, S_j, \ell_j(S_j)) \mid j \in [T], \pi(j) \leq r\} \qquad \forall r \in [T] \tag{B.8}$$

as the information observed by $\mathcal{B}$ by the end to the time step when we feed it the estimate relative to round t_r. Note that this defines a filtration, as $\mathcal{F}_{r-1} \subseteq \mathcal{F}_r$ for all $r \in [T]$, which has some desirable properties thanks to the ordering π we consider. In particular, we have that $\widetilde{d}_{t_r}, \varepsilon_{t_r}, p_{t_r}, N'_{t_r}$ are $\mathcal{F}_{r-1}$-measurable random variables by the way we define them. This property is also due to the fact that N_{t_r} and $\mathcal{L}'_{t_r}$ are determined when conditioning on $\mathcal{F}_{r-1}$. Moreover, we are now interested in

the following importance-weighted loss estimates with implicit exploration:

$$\widetilde{\ell}_t(a) \coloneqq \frac{\mathbb{I}\{A_t = a\}}{p_t(a) + \gamma_t}\widetilde{\theta}_t(s_t(a)) \qquad \forall t \in [T], \forall a \in \mathcal{A}\,. \tag{B.9}$$

Corollary B.1. *Let γ_{t_r} and $\alpha_{t_r}(a)$ be non-negative $\mathcal{F}_{r-1}$-measurable random variables such that $\alpha_{t_r}(a) \leq 2\gamma_{t_r}$, for all $r \in [T]$ and all $a \in \mathcal{A}$. Let $\widetilde{\ell}_t(a)$ be as in* (B.9)*. Then,*

$$\sum_{t=1}^{T}\sum_{a=1}^{K}\alpha_t(a)\big(\widetilde{\ell}_t(a) - \widetilde{\theta}_t(s_t(a))\big) \leq \ln(1/\delta)$$

holds with probability at least $1-\delta$ for any $\delta \in (0,1)$.

Proof. We follow the proof of Neu (2015, Lemma 1) by considering any realization $\ell_1, \ldots, \ell_T$ of the losses. The main difference is that, when defining the supermartingale as in the original proof, we need to consider the terms of the sum in the order denoted by π instead of the increasing order of t. For this reason, we rewrite the sum from the statement by following the order given by π:

$$\sum_{r=1}^{T}\sum_{a=1}^{K}\alpha_{t_r}(a)\big(\widetilde{\ell}_{t_r}(a) - \widetilde{\theta}_{t_r}(s_{t_r}(a))\big)\,.$$

At this point, we need prove that $\mathbb{E}\big[\widetilde{\ell}_{t_r}(a) \,\big|\, \mathcal{F}_{r-1}\big] \leq \widetilde{\theta}_{t_r}(s_{t_r}(a))$, where we recall that $t_r = \pi^{-1}(r)$. Also recall that ε_{t_r}, p_{t_r} and γ_{t_r} are $\mathcal{F}_{r-1}$-measurable. This property allows us to prove the inequality with the conditional expectation of $\widehat{\theta}_t$ instead of the one with the actual optimistic estimates $\widetilde{\theta}_t$, by the definition of the latter. In other words, we now need to prove that $\mathbb{E}\big[\widehat{\ell}_{t_r}(a) \,\big|\, \mathcal{F}_{r-1}\big] \leq \widehat{\theta}_{t_r}(s_{t_r}(a))$, where $\widehat{\ell}_t(a) = \frac{\mathbb{I}\{A_t=a\}}{p_t(a)+\gamma_t}\widehat{\theta}_t(s_t(a))$.

We can consider two cases depending on whether $\widetilde{d}_{t_r} < d_{t_r}$ is true or not (and, thus, we are in the case $\widetilde{d}_{t_r} = d_{t_r}$). In the first case, note that the realized losses used for computing $\widehat{\theta}_{t_r}(s_{t_r}(a))$ correspond to time steps in $\mathcal{L}'_{t_r}(s_{t_r}(a))$, for which there is a corresponding tuple in $\mathcal{F}_{r-1}$. Therefore, we have that $\widehat{\theta}_{t_r}(s_{t_r}(a))$ is $\mathcal{F}_{r-1}$-measurable, and we can show that

$$\mathbb{E}\left[\widehat{\ell}_{t_r}(a)\mathbb{I}\Big\{\widetilde{d}_{t_r} < d_{t_r}\Big\} \,\middle|\, \mathcal{F}_{r-1}\right] = \mathbb{E}\left[\frac{\mathbb{I}\{A_{t_r} = a\}}{p_{t_r}(a) + \gamma_{t_r}} \,\middle|\, \mathcal{F}_{r-1}\right]\frac{\mathbb{I}\Big\{\widetilde{d}_{t_r} < d_{t_r}\Big\}}{N'_{t_r}(s_{t_r}(a))}\sum_{j \in \mathcal{L}'_{t_r}(s_{t_r}(a))}\ell_j(s_{t_r}(a))\,.$$

In the second case, we have that $\widetilde{d}_{t_r} = d_{t_r}$, which implies that $t_r \in \mathcal{L}'_{t_r}(s_{t_r}(a))$ in the case $A_{t_r} = a$. This means that we have a corresponding tuple in $\mathcal{F}_{r-1}$ only for rounds in $\mathcal{L}'_{t_r}(s_{t_r}(a)) \setminus \{t_r\}$. Nonetheless, this does not pose an issue since we have the indicator $\mathbb{I}\{A_{t_r} = a\}$, and thus $S_{t_r} = s_t(a)$. Indeed, we have that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\ell}_{t_r}(a)\mathbb{I}\Big\{\widetilde{d}_{t_r} = d_{t_r}\Big\} \,\middle|\, \mathcal{F}_{r-1}\right] &= \mathbb{E}\left[\frac{\mathbb{I}\{A_{t_r} = a\}}{p_{t_r}(a) + \gamma_{t_r}} \cdot \frac{\mathbb{I}\Big\{\widetilde{d}_{t_r} = d_{t_r}\Big\}}{N'_{t_r}(s_{t_r}(a))}\sum_{j \in \mathcal{L}'_{t_r}(s_{t_r}(a))}\ell_j(s_{t_r}(a)) \,\middle|\, \mathcal{F}_{r-1}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{I}\{A_{t_r} = a\}}{p_{t_r}(a) + \gamma_{t_r}} \,\middle|\, \mathcal{F}_{r-1}\right]\frac{\mathbb{I}\Big\{\widetilde{d}_{t_r} = d_{t_r}\Big\}}{N'_{t_r}(s_{t_r}(a))}\sum_{\substack{j \in \mathcal{L}'_{t_r}(s_{t_r}(a)) \\ j \neq t_r}}\ell_j(s_{t_r}(a))
\end{aligned}$$

$$+ \mathbb{E}\left[\frac{\mathbb{I}\{A_{t_r}=a\}}{p_{t_r}(a)+\gamma_{t_r}} \,\middle|\, \mathcal{F}_{r-1}\right] \frac{\mathbb{I}\left\{\widetilde{d}_{t_r}=d_{t_r}\right\}}{N'_{t_r}(s_{t_r}(a))}\ell_{t_r}(s_{t_r}(a))$$

$$= \mathbb{E}\left[\frac{\mathbb{I}\{A_{t_r}=a\}}{p_{t_r}(a)+\gamma_{t_r}} \,\middle|\, \mathcal{F}_{r-1}\right] \frac{\mathbb{I}\left\{\widetilde{d}_{t_r}=d_{t_r}\right\}}{N'_{t_r}(s_{t_r}(a))} \sum_{j\in\mathcal{L}'_{t_r}(s_{t_r}(a))} \ell_j(s_{t_r}(a))$$

and therefore the inequality

$$\mathbb{E}\left[\widehat{\ell}_{t_r}(a) \,\middle|\, \mathcal{F}_{r-1}\right] = \mathbb{E}\left[\frac{\mathbb{I}\{A_{t_r}=a\}}{p_{t_r}(a)+\gamma_{t_r}} \,\middle|\, \mathcal{F}_{r-1}\right] \widehat{\theta}_{t_r}(s_{t_r}(a)) \le \widehat{\theta}_{t_r}(s_{t_r}(a))$$

is true because $\mathbb{I}\{\widetilde{d}_t < d_t\} + \mathbb{I}\{\widetilde{d}_t = d_t\} = 1$ for all $t \in [T]$, and by definition of $\widehat{\theta}_t$.

As already mentioned, this is equivalent to proving that $\mathbb{E}[\widetilde{\ell}_{t_r}(a) \,|\, \mathcal{F}_{r-1}] \le \widetilde{\theta}_{t_r}(s_{t_r}(a))$ holds. By using a notation similar to the original proof, if we define $\widetilde{\lambda}_r \coloneqq \sum_{a=1}^K \alpha_{t_r}(a)\widetilde{\ell}_{t_r}(a)$ and $\lambda_r \coloneqq \sum_{a=1}^K \alpha_{t_r}(a)\widetilde{\theta}_{t_r}(s_{t_r}(a))$, the process $(Z_r)_{r\in[T]}$ with $Z_r \coloneqq \exp\left(\sum_{j=1}^r(\widetilde{\lambda}_j - \lambda_j)\right)$ is a supermartingale with respect to $(\mathcal{F}_r)_{r\in[T]}$ which has the same properties as in the proof of Neu (2015, Lemma 1). This concludes the current proof by following a similar reasoning as in the original one. □

Thanks to this result, we can conclude that the adoption of `DAda-Exp3` for the reduction via `MetaBIO` can guarantee a high-probability regret bound on $\widehat{\mathcal{R}eg}_T^{\mathcal{B}}$ as stated in Theorem 3.1, but with total delay $\widetilde{d}_T = \sum_{t=1}^T \widetilde{d}_t$ instead of d_{tot}.

B.2.4 Regret of `MetaBIO`

By Lemma B.2, we have that

$$\mathrm{Reg}_T \le \sum_{t=1}^T \widetilde{\theta}_t(S_t) - \min_{a\in\mathcal{A}} \sum_{t=1}^T \widetilde{\theta}_t(s_t(a)) + \sum_{t=1}^T \varepsilon_t(S_t) = \widehat{\mathcal{R}eg}_T^{\mathcal{B}} + \sum_{t=1}^T \varepsilon_t(S_t) \tag{B.10}$$

with probability at least $1-\delta/2$, where $\widehat{R}_T^{\mathcal{B}}$ (Equation (B.5)) is the regret of algorithm $\mathcal{B}$ when fed with $(\widetilde{\theta}_t \circ s_t)_{t\in[T]}$ as losses.

Lemma B.5. *Conditioning on the event as stated in Lemma B.2, the sum of errors suffered from* `MetaBIO` *by using the loss estimates* $(\widetilde{\theta}_t)_{t\in[T]}$ *from Equations* (3.1) *and* (3.3) *is*

$$\sum_{t=1}^T \varepsilon_t(S_t) \le \left(4+2\sqrt{2}\right)\sqrt{ST\ln\frac{4ST}{\delta}}\,.$$

Proof. First, observe that we can rewrite the sum of errors as

$$\sum_{t=1}^T \varepsilon_t(S_t) = \sum_{t=1}^T \varepsilon_t(S_t)\mathbb{I}\{\widetilde{d}_t < d_t\} + \sum_{t=1}^T \varepsilon_t(S_t)\mathbb{I}\{\widetilde{d}_t = d_t\}\,.$$

We now provide an upper bound for the first sum of errors. For any $s \in \mathcal{S}$, we define $\mathcal{T}_s \coloneqq \{t \in [T] : S_t = s\}$ to be the set of all rounds when the state observed by the learner corresponds to s.

We can bound it as

$$\begin{aligned}
\sum_{t=1}^{T} \varepsilon_t(S_t)\mathbb{I}\{\widetilde{d}_t < d_t\} &= \sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}_s} \varepsilon_t(s)\mathbb{I}\{\widetilde{d}_t < d_t\} \\
&= \sqrt{2\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}_s}\sqrt{\frac{1}{N'_t(s)}}\mathbb{I}\{\widetilde{d}_t < d_t\} \\
&\le 2\sqrt{\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}_s}\sqrt{\frac{1}{M_t(s)}}\mathbb{I}\{\widetilde{d}_t < d_t\} \qquad \text{(because } N'_t(s) \ge \tfrac{1}{2}M_t(s)\text{)} \\
&\le 4\sqrt{\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sqrt{M_T(s)} \qquad \text{(since } M_t(s) \text{ is increasing over } \mathcal{T}_s\text{)} \\
&\le 4\sqrt{ST\ln\frac{4ST}{\delta}}\,,
\end{aligned}$$

where the second inequality holds because $N'_t(S_t) = N_t(S_t) \ge \frac{1}{2}M_t(S_t)$ when $\widetilde{d}_t < d_t$ since $M_t(S_t) \le N_t(S_t) + \sigma_t$, while the last one follows by Jensen's inequality and the fact that $\sum_{s\in\mathcal{S}} M_T(s) = T$.

As a last step, we provide an upper bound to the second sum. Let $J_s \coloneqq \{r \in \mathcal{T}_s : \widetilde{d}_r = d_r\}$ and notice that $|J_s| \le |\mathcal{T}_s| = M_T(s)$. Observe that $\rho(t) = \widetilde{\rho}(t)$ for each round t such that $\widetilde{d}_t = d_t$, and thus by Equation (B.7) we have that

$$\pi(r) < \pi(t) \iff \rho(r) < \rho(t) \vee (\rho(r) = \rho(t) \wedge r < t)$$

for all $r, t \in [T]$ such that $\widetilde{d}_r = d_r$ and $\widetilde{d}_t = d_t$. Define $\nu_s : J_s \to \big[|J_s|\big]$ by

$$\nu_s(t) \coloneqq |\{r \in J_s \,:\, \pi(r) \le \pi(t)\}| \qquad \forall t \in J_s\,.$$

Observe that $\nu_s(t) \le N'_t(s) = |\mathcal{L}'_t(s)|$ for all $s \in \mathcal{S}$ and all $t \in J_s$. This is due to the fact that $\nu_s(t)$ counts a subset of $\mathcal{L}'_t(s)$; to be precise, we have that $\nu_s(t) = |\mathcal{L}'_t(s) \cap J_s|$. Moreover, notice that the condition $\pi(r) \le \pi(t)$ defines a total order over J_s. Hence, $\nu_s(t)$ counts the number of elements of J_s preceding $t \in J_s$ (including t itself) in this total order. This implies that ν_s is a bijection between J_s and $\big[|J_s|\big]$. Then, using a similar reasoning as before, we show that

$$\begin{aligned}
\sum_{t=1}^{T} \varepsilon_t(S_t)\mathbb{I}\{\widetilde{d}_t = d_t\} &= \sqrt{2\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}_s}\sqrt{\frac{1}{N'_t(s)}}\mathbb{I}\{\widetilde{d}_t = d_t\} \\
&= \sqrt{2\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sum_{t\in J_s}\sqrt{\frac{1}{N'_t(s)}} \qquad \text{(by definition of } J_s\text{)} \\
&\le \sqrt{2\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sum_{t\in J_s}\sqrt{\frac{1}{\nu_s(t)}} \qquad \text{(since } \nu_s(t) \le N'_t(s) \text{ for } t \in J_s\text{)} \\
&\le 2\sqrt{2\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sqrt{|J_s|} \qquad \text{(since } \nu_s(t) \text{ is bijective)} \\
&\le 2\sqrt{2\ln\frac{4ST}{\delta}}\sum_{s\in\mathcal{S}}\sqrt{M_T(s)} \qquad \text{(since } |J_s| \le M_T(s)\text{)}
\end{aligned}$$

$$\leq 2\sqrt{2ST\ln\frac{4ST}{\delta}}\,. \qquad \text{(by Jensen's inequality)}$$

□

Theorem 3.2. *Let $\delta \in (0,1)$. If we run* `MetaBIO` *using* `DAda-Exp3`*, then the regret of* `MetaBIO` *in the BIO setting with adversarial action-state mappings and stochastic losses satisfies*

$$\mathrm{Reg}_T \leq 2\sqrt{2KTC_{K,3\delta}} + 7\sqrt{ST\ln\frac{4ST}{\delta}} + 2\sqrt{d_\Phi C_{K,3\delta}} + \frac{\sigma_{\max}+2}{2}\ln\frac{4}{\delta} \tag{3.8}$$

with probability at least $1-\delta$.

Proof of Theorem 3.2. By Equation (B.10), the regret Reg_T can be bounded as

$$\mathrm{Reg}_T \leq \widehat{\mathcal{R}eg}_T^{\mathcal{B}} + \sum_{t=1}^{T}\varepsilon_t(S_t) \leq \widehat{\mathcal{R}eg}_T^{\mathcal{B}} + 7\sqrt{ST\ln\frac{4ST}{\delta}}$$

with probability at least $1-\delta/2$, where the last inequality follows by Lemma B.5. From what we argued in Appendix B.2.3, we can upper bound $\widehat{\mathcal{R}eg}_T^{\mathcal{B}}$ using the high-probability regret bound of `DAda-Exp3`. Notice that the delays incurred by `DAda-Exp3` via `MetaBIO` are those given when providing the estimates $\left(\widetilde{\theta}_t\right)_{t\in[T]}$. We denote these delays by $\widetilde{d}_1,\dots,\widetilde{d}_T$, and the total delay perceived by `DAda-Exp3` is thus $\widetilde{d}_{\mathrm{tot}} = \sum_{t=1}^{T}\widetilde{d}_t$. Hence, from the improved bound for `DAda-Exp3` in Equation (B.2), we have that

$$\widehat{\mathcal{R}eg}_T^{\mathcal{B}} \leq 2\sqrt{2KT\left(3\ln(K)+\ln\left(4/\delta\right)\right)} + 2\sqrt{\widetilde{d}_{\mathrm{tot}}\left(3\ln(K)+\ln\left(4/\delta\right)\right)} + \frac{\sigma_{\max}+2}{2}\ln(4/\delta)$$

holds with probability at least $1-\delta/2$. The combination of the above two inequalities, together with Lemma 3.1, concludes the proof. □

B.2.5 Regret of `AdaMetaBIO`

Theorem 3.3. *Let $\delta \in (0,1)$. If we run* `AdaMetaBIO` *with* `DAda-Exp3`*, then the regret of* `AdaMetaBIO` *in the BIO setting with adversarial action-state mappings and stochastic losses satisfies*

$$\mathrm{Reg}_T \leq 3\min\left\{7\sqrt{ST\ln\frac{8ST}{\delta}}, \sqrt{d_{\mathrm{tot}}C_{K,2\delta}}\right\} + 6\sqrt{KTC_{K,2\delta}} + 2\sqrt{d_\Phi C_{K,2\delta}} + (\sigma_{\max}+2)\ln\frac{8}{\delta} \tag{3.9}$$

with probability at least $1-\delta$.

Proof of Theorem 3.3. Let $t^* \in [T]$ be the last round before `AdaMetaBIO` switches from `DAda-Exp3` to `MetaBIO`, i.e., the last round that satisfies $\mathfrak{D}_{t^*}C_{K,4\delta} \leq 49ST\ln\frac{8ST}{\delta}$. Then, define

$$a^* \in \arg\min_a \sum_{t=1}^{T}\theta(s_t(a))\,.$$

We may decompose regret as

$$\mathrm{Reg}_T = \sum_{t=1}^{t^*}\Big(\theta(S_t) - \theta(s_t(a^*))\Big) + \sum_{t=t^*+1}^{T}\Big(\theta(S_t) - \theta(s_t(a^*))\Big)$$

$$\leq \underbrace{\sum_{t=1}^{t^*} \theta(S_t) - \min_{a\in\mathcal{A}} \sum_{t=1}^{t^*} \theta(s_t(a))}_{\mathrm{Reg}_{t^*}} + \underbrace{\sum_{t=t^*+1}^{T} \theta(S_t) - \min_{a\in\mathcal{A}} \sum_{t=t^*+1}^{T} \theta(s_t(a))}_{\mathrm{Reg}_{t^*:T}} .$$

The incurred delay until time t^* is $\mathfrak{D}_{t^*}$. Thus, from Equation (B.4), we get that the following bound

$$\mathrm{Reg}_{t^*} \leq 2\sqrt{2Kt^*C_{K,2\delta}} + \sqrt{2t^* \ln \frac{12}{\delta}} + 2\sqrt{\mathfrak{D}_{t^*}C_{K,2\delta}} + \frac{\sigma_{\max}+2}{2} \ln \frac{6}{\delta} \tag{B.11}$$

holds with probability at least $1-\delta/2$, where we recall that $C_{K,\delta} = 3\ln K + \ln(12/\delta)$. If our algorithm never switches, then $t^* = T$ and we get the bound in (B.11) for Reg_T. Note that this is no greater than the upper bound in the statement as $\sqrt{\mathfrak{D}_T C_{K,2\delta}} \leq 7\sqrt{ST\ln(8ST/\delta)}$ by definition of t^* in this case.

Otherwise, we use the switching condition $\sqrt{\mathfrak{D}_{t^*}C_{K,2\delta}} \leq 7\sqrt{ST\ln(8ST/\delta)}$ along with the fact that $\sqrt{t^*\ln(12/\delta)} \leq \sqrt{Kt^*C_{K,2\delta}}$ to get

$$\mathrm{Reg}_{t^*} \leq 3\sqrt{2Kt^*C_{K,2\delta}} + 14\sqrt{ST\ln\frac{8ST}{\delta}} + \frac{\sigma_{\max}+2}{2}\ln\frac{6}{\delta} . \tag{B.12}$$

Furthermore, Theorem 3.2 directly gives us an upper bound for $\mathrm{Reg}_{t^*:T}$ since `AdaMetaBIO` runs `MetaBIO` for $t > t^*$ with the confidence parameter set to $\delta/2$. We just need to bound the total incurred delays of these rounds, namely $\widetilde{d}_{t^*:T}$. Let σ'_t be the outstanding observations for any round $t > t^*$ as perceived by the execution of `MetaBIO` starting after round t^*, that is, when considering only delays $(d_t)_{t>t^*}$. It is immediate to observe that $\sigma'_t \leq \sigma_t$ and thus $\max_{t>t^*}\sigma'_t \leq \max_{t>t^*}\sigma_t$. Moreover, from Lemma 3.1 we have

$$\widetilde{d}_{t^*:T} \leq d_{\Phi'} ,$$

where Φ' denotes a set of $\min\{T-t^*, 2S\sigma'_{\max}\}$ rounds with the largest delays among $(d_t)_{t>t^*}$, with $\sigma'_{\max} \coloneqq \max_{t>t^*}\sigma'_t$. So we have

$$d_{\Phi'} \leq d_{\Phi}$$

due to the fact that $|\Phi'| = \min\{T-t^*, 2S\sigma'_{\max}\} \leq \min\{T, 2S\sigma_{\max}\} = |\Phi|$. Therefore, from Theorem 3.2 we obtain

$$\mathrm{Reg}_{t^*:T} \leq 2\sqrt{2K(T-t^*)C_{K,3\delta}} + 7\sqrt{ST\ln\frac{8ST}{\delta}} + 2\sqrt{d_\Phi C_{K,3\delta}} + \frac{\sigma_{\max}+2}{2}\ln\frac{8}{\delta} \tag{B.13}$$

with probability at least $1-\delta/2$. We conclude the proof by combining Equations (B.12) and (B.13) along with the fact that $\sqrt{t^*} + \sqrt{T-t^*} \leq \sqrt{2T}$ to get that the bound

$$\mathrm{Reg}_T \leq 6\sqrt{KTC_{K,2\delta}} + 3\min\left\{7\sqrt{ST\ln\frac{8ST}{\delta}}, \sqrt{d_{\mathrm{tot}}C_{K,2\delta}}\right\} + 2\sqrt{d_\Phi C_{K,2\delta}} + (\sigma_{\max}+2)\ln\frac{8}{\delta}$$

holds with probability at least $1-\delta$. □

B.2.6 Expected Regret Analysis of `AdaMetaBIO` with `Tsallis-INF`

Proposition 3.1. *If we execute* `AdaMetaBIO` *with* `Tsallis-INF` *(Zimmert and Seldin, 2020a), and use the switching condition* $\sqrt{8\mathfrak{D}_t \ln K} > 6\sqrt{ST\ln(2ST)}$ *at each round* $t \in [T]$*, where* $\mathfrak{D}_t = \sum_{j=1}^{t} \sigma_j$*, then the regret of* `AdaMetaBIO` *in the BIO setting with adversarial action-state mappings and stochastic losses satisfies*

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq 4\sqrt{2KT} + 2\sqrt{2d_\Phi \ln K} + 4\min\left\{3\sqrt{ST\ln(2ST)}, \sqrt{2d_{\mathrm{tot}}\ln K}\right\}.$$

Proof of Proposition 3.1. We begin by studying of expected regret of `MetaBIO` and we then give a regret analysis of `AdaMetaBIO`. When running `MetaBIO`, we use the unbiased empirical mean estimators $(\widehat{\theta}_t)_{t\in[T]}$ as the mean loss estimates, rather than the lower confidence bounds $(\widetilde{\theta}_t)_{t\in[T]}$. The expected regret is defined as

$$\mathbb{E}[\mathrm{Reg}_T] = \sum_{t=1}^{T} \mathbb{E}\left[\theta(S_t)\right] - \sum_{t=1}^{T} \theta(s_t(a^*)) \,,$$

where we fix any $a^* \in \arg\min_{a\in\mathcal{A}} \sum_{t=1}^{T} \theta(s_t(a))$. Here we use a version of `Tsallis-INF` that is tailored for the delayed bandits problem (Zimmert and Seldin, 2020a), which guarantees a bound in expectation on the regret

$$\widehat{\mathcal{R}eg}_T^{\,\mathrm{Tsallis}}(a) \coloneqq \sum_{t=1}^{T} \widehat{\theta}_t(S_t) - \sum_{t=1}^{T} \widehat{\theta}_t(s_t(a))$$

against any fixed action $a \in \mathcal{A}$, using the loss estimates $\left(\widehat{\theta}_t\right)_{t\in[T]}$. Observe that this regret is defined in terms of our estimates, as required in our case. By Zimmert and Seldin (2020a, Theorem 1), `Tsallis-INF` guarantees that its expected regret is

$$\mathbb{E}\left[\widehat{\mathcal{R}eg}_T^{\,\mathrm{Tsallis}}(a^*)\right] = \mathbb{E}\left[\sum_{t=1}^{T} \widehat{\theta}_t(S_t) - \sum_{t=1}^{T} \widehat{\theta}_t(s_t(a^*))\right] \leq 4\sqrt{KT} + \sqrt{8\widetilde{d}_T \ln K} \leq 4\sqrt{KT} + \sqrt{8d_\Phi \ln K} \,,$$

where the last inequality uses Lemma 3.1. Then, we can focus on our notion of regret and use the above regret bound to obtain that

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}_T] &= \mathbb{E}\Big[\mathrm{Reg}_T - \widehat{\mathcal{R}eg}_T^{\,\mathrm{Tsallis}}(a^*)\Big] + \mathbb{E}\left[\widehat{\mathcal{R}eg}_T^{\,\mathrm{Tsallis}}(a^*)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\big(\theta(S_t) - \widehat{\theta}_t(S_t)\big)\right] + \mathbb{E}\left[\sum_{t=1}^{T}\big(\widehat{\theta}_t(s_t(a^*)) - \theta(s_t(a^*))\big)\right] + \mathbb{E}\left[\widehat{\mathcal{R}eg}_T^{\,\mathrm{Tsallis}}(a^*)\right] \\
&\leq \mathbb{E}\Bigg[\underbrace{\sum_{t=1}^{T}\big(\theta(S_t) - \widehat{\theta}_t(S_t)\big)}_{\Delta}\Bigg] + \mathbb{E}\left[\sum_{t=1}^{T}\big(\widehat{\theta}_t(s_t(a^*)) - \theta(s_t(a^*))\big)\right] + 4\sqrt{KT} + \sqrt{8d_\Phi \ln K} \,.
\end{aligned} \tag{B.14}$$

We know that our mean estimator is unbiased. Therefore, we have that $\mathbb{E}\left[\widehat{\theta}_t(s_t(a^*))\right] = \theta(s_t(a^*))$ for any $t \in [T]$, meaning that the second term in the right-hand side of (B.14) is equal to zero.

On the other hand, we can apply Lemma B.2 to get the following bound for Δ that holds with

probability at least $1 - \delta/2$ for any $\delta \in (0, 1)$:

$$\Delta \leq \min\left\{\frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(S_t), T\right\}, \tag{B.15}$$

where we recall that $\varepsilon_t(s) = \sqrt{\frac{2}{N_t'(s)}\ln\frac{4ST}{\delta}}$. In particular, the inequality $\Delta \leq T$ is true in general. By Lemma B.5, we can bound the right-hand side of (B.15) as

$$\frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(S_t) \leq \frac{7}{2}\sqrt{ST\ln\frac{4ST}{\delta}}$$

when conditioning on the event as in the statement of Lemma B.2. If we denote such an event as $\mathcal{E}$, we have that $\mathbb{P}\left(\overline{\mathcal{E}}\right) \leq \delta/2$ and that $\mathbb{E}\left[\Delta \mid \mathcal{E}\right] \leq \frac{7}{2}\sqrt{ST\ln\left(4ST/\delta\right)}$. As a consequence, we notice that

$$\mathbb{E}\left[\Delta\right] = \mathbb{E}\left[\Delta \mid \mathcal{E}\right]\mathbb{P}\left(\mathcal{E}\right) + \mathbb{E}\left[\Delta \mid \overline{\mathcal{E}}\right]\mathbb{P}\left(\overline{\mathcal{E}}\right) \leq \frac{7}{2}\sqrt{ST\ln\frac{4ST}{\delta}} + \frac{\delta}{2}T \leq 5\sqrt{ST\ln\left(2ST\right)} + 1$$

where in the last inequality we set $\delta = 2/T$. Since we assume that $S \geq 2$, we can easily observe that $\mathbb{E}\left[\Delta\right] \leq 6\sqrt{ST\ln\left(2ST\right)}$. Plugging this into Equation (B.14) gives us

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq 4\sqrt{KT} + \sqrt{8d_\Phi\ln K} + 6\sqrt{ST\ln(2ST)}\,. \tag{B.16}$$

At this point, we can proceed to the proof of the overall bound on the expected regret of `AdaMetaBIO`. The behaviour of `AdaMetaBIO` follows the same principle as before, but the switching condition is different:

$$\sqrt{8\mathfrak{D}_t\ln K} > 6\sqrt{ST\ln(2ST)}\,.$$

Similar to the analysis of `AdaMetaBIO` in Appendix B.2.5, we decompose the regret into

$$\mathbb{E}[\mathrm{Reg}_T] \leq \underbrace{\sum_{t=1}^{t^*}\mathbb{E}\left[\theta(S_t)\right] - \min_{a\in\mathcal{A}}\sum_{t=1}^{t^*}\theta(s_t(a))}_{\mathrm{Reg}_{t^*}} + \underbrace{\sum_{t=t^*+1}^{T}\mathbb{E}\left[\theta(S_t)\right] - \min_{a\in\mathcal{A}}\sum_{t=t^*+1}^{T}\theta(s_t(a))}_{\mathrm{Reg}_{t^*:T}}\,,$$

where t^* is the last round satisfying $\sqrt{8\mathfrak{D}_{t^*}} \leq 6\sqrt{ST\ln\left(2ST\right)}$. Then, we have

$$\mathbb{E}[\mathrm{Reg}_{t^*}] \leq 4\sqrt{Kt^*} + \sqrt{8\mathfrak{D}_{t^*}\ln K}\,. \tag{B.17}$$

If $t^* = T$ then $\mathrm{Reg}_{t^*} = \mathrm{Reg}_T$ and we get the bound in Equation (B.17), where we note that $\sqrt{8\mathfrak{D}_T\ln K} \leq 6\sqrt{ST\ln\left(2ST\right)}$ by definition of t^* in this case, and we can replace $\mathfrak{D}_T$ by d_T. Otherwise, $t^* < T$ and we can apply the bound for `MetaBIO` from Equation (B.16), along with the fact that the total incurred delay after round t^* is upper bounded by d_Φ, in order to derive an upper bound for $\mathbb{E}[\mathrm{Reg}_{t^*:T}]$ that is

$$\mathbb{E}[\mathrm{Reg}_{t^*:T}] \leq 4\sqrt{K(T-t^*)} + \sqrt{8d_\Phi\ln K} + 6\sqrt{ST\ln(2ST)}\,. \tag{B.18}$$

Finally, if we use the fact that $\sqrt{8\mathfrak{D}_{t^*}} \leq 6\sqrt{ST\ln(2ST)}$ (by definition of t^*) in Equation (B.17),

and combine it with Equation (B.18), we conclude that

$$\mathbb{E}[\mathrm{Reg}_T] \leq 4\sqrt{2KT} + \sqrt{8d_\Phi \ln K} + 2\min\Big\{6\sqrt{ST\ln(2ST)}, \sqrt{8d_{\mathrm{tot}}\ln K}\Big\}\,,$$

where we also used the fact that $\sqrt{t^*} + \sqrt{T-t^*} \leq \sqrt{2T}$. □

B.3 Omitted Details in Section 3.5

Theorem 3.5. *Suppose that the action-state mapping is adversarial and the losses are stochastic and that $d_t = d$ for all $t \in [T]$. If $T \geq \min\{S, d\}$ then there exists a distribution of losses and a sequence of action-state mappings such that any (possibly randomized) algorithm suffers regret $\mathbb{E}\left[\mathrm{Reg}_T\right] = \Omega\big(\sqrt{\min\{S,d\}T}\big)$.*

Proof of Theorem 3.5. Assume without loss of generality that $K = 2$ and let $\mathcal{S} \coloneqq \{h_1, \ldots, h_S\}$ be the finite set of possible states. Let $S' \coloneqq \lfloor \min\{S/2, d\} \rfloor$ and let $I_1, \ldots, I_T$ be the actions chosen by the considered algorithm. Split the T time steps into $m \coloneqq \lfloor T/S' \rfloor$ blocks $B_1, \ldots, B_m$ of equal size S', eventually leaving $\leq S' - 1$ extra time steps. We assume with no loss of generality that the last step corresponds to the end of the m-th block. The feedback formed by the losses of the actions chosen by the algorithm in a certain block is received only after the last time step of the same block since $S \leq 2d$. Define $b_i \coloneqq (i-1)S' + 1$ for all $i \in [m]$. We assume that the learner receives *all* the realized losses $\ell_t(s_t(A))$ for all $t \in B_i$ and all $A \in \{1, 2\}$ at the end of each block, which means that we are in a full information setting, as this only helps the algorithm.

Now, we define a specific sequence of assignments from actions to states, and construct losses so that the expected regret becomes sufficiently large. Let $s_t(A) \coloneqq h_{2(t-b_i)+A}$ for all $t \in B_i$, all $i \in [m]$ and all $A \in \{1, 2\}$; this means that, for the first time step of any block, actions 1 and 2 will be assigned to states h_1 and h_2 respectively, then to h_3 and h_4 respectively in the next time step of the same block, and so on. Let $\varepsilon \coloneqq \frac{1}{4}\sqrt{\frac{S'}{2T\ln(4/3)}} \in \left[0, \frac{1}{4}\right]$ and let $\theta^{(A)} \in \mathbb{R}^2$ be a vector of mean losses such that $\theta_i^{(A)} \coloneqq \frac{1}{2} - \mathbb{I}\{i = A\}\,\varepsilon$, for each $A \in \{1, 2\}$. We simplify the notation with $\mathbb{E}_A[\cdot] \coloneqq \mathbb{E}\left[\cdot \,\middle|\, \theta^{(A)}\right]$ and $\mathbb{P}_A(\cdot) \coloneqq \mathbb{P}\left(\cdot \,\middle|\, \theta^{(A)}\right)$, where the conditioning on $\theta^{(A)}$ means that we sample losses for each state assigned to $i \in \{1, 2\}$ such that they are Bernoulli random variables with mean $\theta_i^{(A)}$. In particular, conditioning on $\theta^{(A)}$, we sample independent Bernoulli random variables $X_1^i, \ldots, X_m^i$ with mean $\theta_i^{(A)}$, one for each block, for $i \in \{1, 2\}$. Then, the losses are defined as $\ell_t(s_t(i)) \coloneqq X_j^i$ for each $t \in B_j$ and each $j \in [m]$.

We can now proceed to show a lower bound for the expected pseudo-regret. Let T_i be the number of times the learner chooses action i over all T time steps. The expected pseudo-regret over the two instances determined by $\theta^{(k)}$ for $k \in \{1, 2\}$ adds up to

$$\mathbb{E}_1\left[\mathrm{Reg}_T\right] + \mathbb{E}_2\left[\mathrm{Reg}_T\right] = \varepsilon\left(2T - \mathbb{E}_1\left[T_1\right] - \mathbb{E}_2\left[T_2\right]\right)\,.$$

Following the standard analysis, we show that the difference $\mathbb{E}_2\left[T_2\right] - \mathbb{E}_1\left[T_2\right]$ is such that

$$\mathbb{E}_2\left[T_2\right] - \mathbb{E}_1\left[T_2\right] \leq T \cdot d_{\mathrm{TV}}(\mathbb{P}_2, \mathbb{P}_1) \leq T\sqrt{\frac{1}{2} D_{\mathrm{KL}}\left(\mathbb{P}_1 \,\|\, \mathbb{P}_2\right)}\,,$$

where the last step follows by Pinsker's inequality.

Let $\lambda_i \coloneqq \{(I_t, \ell_t(S_t(1)), \ell_t(S_t(2))) \mid t \in B_i\}$ be the feedback set known to the learner by the end of block B_i, and let $\lambda^i \coloneqq (\lambda_1, \ldots, \lambda_i)$ be the tuple of all feedback sets up to the end of block B_i. Denote by $\mathbb{P}_{k,i}(\cdot)$ the probability measure of feedback tuples λ^i conditioned on $\theta^{(A)}$. By the chain rule for the relative entropy, we can observe that

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_1 \,\|\, \mathbb{P}_2) &= \sum_{i=1}^{m} \sum_{\lambda^{i-1}} \mathbb{P}_1\left(\lambda^{i-1}\right) D_{\mathrm{KL}}\left(\mathbb{P}_{1,i}\left(\cdot \mid \lambda^{i-1}\right) \,\|\, \mathbb{P}_{2,i}\left(\cdot \mid \lambda^{i-1}\right)\right) \\ &\leq \sum_{i=1}^{m} \sum_{\lambda^{i-1}} \mathbb{P}_1\left(\lambda^{i-1}\right) 16\varepsilon^2 \ln(4/3) \\ &= 16m\varepsilon^2 \ln(4/3) \ , \end{aligned}$$

where we used the fact that each relative entropy $D_{\mathrm{KL}}\left(\mathbb{P}_{1,i}\left(\cdot \mid \lambda^{i-1}\right) \,\|\, \mathbb{P}_{2,i}\left(\cdot \mid \lambda^{i-1}\right)\right)$ corresponds to the sum of the relative entropy between two Bernoulli distributions with means $1/2$ and $1/2 - \varepsilon$ and that between Bernoulli distributions with means $1/2 - \varepsilon$ and $1/2$, respectively, which is upper bounded by $16\varepsilon^2 \ln(4/3)$ for $\varepsilon \in [0, 1/4]$. This follows by an application of the chain rule for the relative entropy, as well as from the fact that the distribution of I_t is the same under both $\mathbb{P}_{1,i}\left(\cdot \mid \lambda^{i-1}\right)$ and $\mathbb{P}_{2,i}\left(\cdot \mid \lambda^{i-1}\right)$, for all $t \in B_i$ and any λ^{i-1}. Therefore, we have that

$$\mathbb{E}_2[T_2] - \mathbb{E}_1[T_2] \leq 2\varepsilon T \sqrt{2m \ln(4/3)}$$

which also implies that

$$\mathbb{E}_1[\mathrm{Reg}_T] + \mathbb{E}_2[\mathrm{Reg}_T] \geq \varepsilon T \left(1 - 2\varepsilon \sqrt{2\frac{T}{S'} \ln(4/3)}\right) = \frac{\varepsilon T}{2} \geq \frac{1}{8}\sqrt{\frac{\lfloor S/2 \rfloor T}{2\ln(4/3)}} \geq \frac{1}{8}\sqrt{\frac{ST}{6\ln(4/3)}} \ ,$$

where we used the facts that $m \leq T/S'$ and that $\lfloor S/2 \rfloor \geq S/3$ for any integer $S \geq 2$. This means that the expected pseudo-regret of the learner has to be $\frac{1}{16}\sqrt{\frac{ST}{6\ln(4/3)}}$ at least in one of the two instances. Now, for $S > 2d$ we use the same construction, but now we only use $2d$ states, which leads to the promised $\Omega(\sqrt{\min\{S, d\}T})$ lower bound. □

Theorem 3.6. *Suppose that the action-state mapping is adversarial, the losses are stochastic, and that $d_t = d$ for all $t \in [T]$. If $T \geq d + 1$ then there exists a distribution of losses and a sequence of action-state mappings such that any (possibly randomized) algorithm suffers regret*

$$\mathbb{E}[\mathrm{Reg}_T] = \Omega\Big(\min\Big\{(d+1)\sqrt{S}, \sqrt{(d+1)T}\Big\}\Big) \ .$$

Proof of Theorem 3.6. Let $S' \coloneqq \min\left\{\lfloor \frac{S}{2} \rfloor, \lfloor \frac{T}{d+1} \rfloor\right\} \geq 1$. We consider the first $(d+1)S'$ rounds of the game and divide them into S' blocks $B_1, \ldots, B_{S'}$ of same length $d + 1$. In this way, we ensure that the feedback for any time step in some block is revealed to the learner only after its final round.

Without loss of generality, we can assume that the learner observes all the losses of one block immediately after its last time step; this only helps the learner since they would observe only the incurred losses at possibly later rounds otherwise. We can further simplify the problem by assuming that losses are deterministic functions of the states, i.e., $\ell_t \equiv \theta$ for every round t. This also means that the problem turns into an easier, full-information version of our problem with deterministic

losses. Now, let the adversary choose the action-state mappings such that for each block index i and each action $a \in \mathcal{A}$, $S_t(a) = S_{t'}(a) \in \{s_{2i-1}, s_{2i}\}$ for all $t, t' \in B_i$. Furthermore, we assume that the losses are chosen such that $\theta(s_{2i-1}) \in \{0, 1\}$ and $\theta(s_{2i}) = 1 - \theta(s_{2i-1})$ for all $i \in [S']$. In this construction, the learner cannot obtain any useful information from the states of a block because of the delays. Moreover, the states observed in one block are not observed again in the other blocks.

It thus suffices to prove a lower bound for a standard full information game with S' rounds and loss range $[0, d+1]$. Hence, we can conclude that the expected regret of any algorithm has to be

$$\mathbb{E}\left[\mathrm{Reg}_T\right] = \Omega\left((d+1)\sqrt{S'}\right) = \Omega\left(\min\left\{(d+1)\sqrt{S}, \sqrt{(d+1)T}\right\}\right) . \qquad \square$$

B.4 Action-State Mappings and Loss Means Used in the Experiments

Table B.1 and Table B.2 describe the instances used to generate the data for the experiments of Section 3.6.

Mean loss	$s = 1$	$s = 2$	$s = 3$
$\theta(s)$	0.2	0.4	0.8

Mapping	$P(1\|a)$	$P(2\|a)$	$P(3\|a)$
$a = 1$	0.8	0.1	0.1
$a = 2$	0.4	0.5	0.1
$a = 3$	0.3	0.7	0.0
$a = 4$	0.5	0.3	0.2

Table B.1: Mean losses and stochastic action-state mapping for Experiment 1 in Section 3.6.

Mean loss	$s = 1$	$s = 2$	$s = 3$
$\theta(s)$	0	1	1

Environment 1

Mapping	$P(1\|a)$	$P(2\|a)$	$P(3\|a)$
$a = 1$	0.06	0.47	0.47
$a = 2$	0	0.50	0.50
$a = 3$	0	0.50	0.50
$a = 4$	0	0.50	0.50

Environment 2

Mapping	$P(1\|a)$	$P(2\|a)$	$P(3\|a)$
$a = 1$	1	0	0
$a = 2$	0.94	0.03	0.03
$a = 3$	0.94	0.03	0.03
$a = 4$	0.94	0.03	0.03

Table B.2: Mean losses and stochastic action-state mappings for Experiment 2 in Section 3.6.

Appendix C

Proof Details for Chapter 4

C.1 Additional related works

Distributed optimization (DO), now central to federated learning, dates back to the work of Bertsekas and Tsitsiklis (1991), originally applied to parallel computation. Much of the research in DO focuses on gossip algorithms, introduced by Boyd et al. (2005, 2006) to address the distributed averaging problem, especially in scenarios where communication is expensive. These algorithms involve randomly selecting a neighbor for information exchange, or *gossiping*. The concept later expanded to include weighted averaging of information from all neighbors using weights collected in a gossip matrix. Nedic et al. (2010) extended gossip methods to distributed optimization, combining projected gradient descent with gossip-based averaging of iterates. Typically, the convergence rate of gossip algorithms is inversely related to the spectral gap of the gossip matrix.

A key constraint frequently considered in distributed optimization is that the algorithm should be robust to random network topologies. This can arise not only from unstable communication channels (Nedić and Olshevsky, 2014), but randomization can also be leveraged to reduce communication costs while preserving performance (Lei et al., 2020). Other constraints, considered in the literature but less relevant to this work, include event-related communication and time delays (Yang et al., 2019). Recent advances in DO also concern accelerated gossip algorithms that allow for accelerated rates with respect to the number of agents (Wan et al., 2024a).

In this work, we address the problem of device unavailability, a topic explored in federated learning from various angles. For example, availability patterns, such as diurnal cycles, can violate the assumption of data independence, as active agents may disproportionately represent certain populations (e.g., by geographic location) (Eichner et al., 2019, Amiri et al., 2021). Device unavailability has not been addressed in the context of DOO, except indirectly in (Hosseini et al., 2016), which focuses on time-varying graphs, thus tackling the case in which isolated devices are unavailable for communication. Our approach differs in that inactive agents do not have an associated loss function and do not contribute to the global loss. Raginsky et al. (2011) consider a notion of information structure replacing the communication network. However, their results are based on a specific linear structure and a horizon-dependent communication radius within which agents can freely exchange information.

In the online DOO setting, Yan et al. (2013) proposed a (sub)gradient descent algorithm with regret bounds of $\mathcal{O}(\sqrt{T})$ for the convex and $\mathcal{O}(\log T)$ for the strongly convex case. Hosseini et al.

(2013) later introduced an online dual-averaging algorithm, also achieving $\mathcal{O}(\sqrt{T})$ regret for convex losses. Yuan et al. (2021) extended this to long-term constraints.

C.2 Additional remark on Algorithm 4.1 and Notation

The instance of `Gossip-FTRL` run by agent v has two local variables: $g_t(v)$, corresponding to the local loss gradient for the prediction $x_t(v)$ of v at time t, and $z_t(v)$, corresponding to the estimate of the network loss gradient computed by agent v. Let $\Gamma_t \in \mathbb{R}^{N\times d}$ be the matrix whose v-th row is

$$g_t(v) = \begin{cases} 0 & \text{if } v \notin S_t, \\ \nabla \ell_t\big(v, x_t(v)\big) & \text{if } v \in S_t. \end{cases} \tag{C.1}$$

Correspondingly, we define Z_t, the matrix whose v-th row is $z_t(v)$ for all $v \in \mathcal{V}$. Let $\boldsymbol{e}_v$ be the canonical basis vector for coordinate $v \in [N]$. Let the weights $W_t(v,\cdot)$ computed by the instance of Algorithm 4.1 run by each $v \in S_t$ form a $N \times N$ gossip matrix W_t for S_t such that $W_t(v,\cdot) = \boldsymbol{e}_v$ for all $v \in \mathcal{V} \setminus S_t$. We may write the updates performed by the instance as $Z_{t+1} = W_t Z_t + \Gamma_t$. Note that the definitions of W_t and Γ_t imply that $z_{t+1}(v) = z_t(v)$ for all agents $v \in \mathcal{V} \setminus S_t$ that are inactive at time t. Moreover, any active agent $v \in S_t$ can compute $z_{t+1}(v)$ using the most recent value $z_s(j)$ (for some $s < t$) received by agents $j \in \mathcal{N}_v \setminus S_t$ (i.e., neighbors inactive at time t).

C.3 Preliminary results

Lemma C.1 (Individual Regret decomposition)**.** *Denoting by* $x^* = \arg\min_{x\in\mathcal{X}} \sum_{t\le T; u\in S_t} \ell_t^{\mathrm{net}}(S_t, x)$*, the individual regret*

$$\mathrm{Reg}_T(u) = \sum_{t\le T; u\in S_t} \ell_t^{\mathrm{net}}\big(S_t, x_t(u)\big) - \min_{x\in\mathcal{X}} \sum_{t\le T; u\in S_t} \ell_t^{\mathrm{net}}(S_t, x)$$

can be decomposed in the following way

$$\begin{aligned}\mathrm{Reg}_T(u) \le & \underbrace{2\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{u, v \in S_t\}}{|S_t|} L\,\|x_t(v) - y_t\|}_{(a)} + \underbrace{\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|} \langle \nabla \ell_t(x_t(v), v), y_t - x^*\rangle}_{(b)} \\ & + \underbrace{\sum_{t=1}^{T} \mathbb{I}\{u \in S_t\} L\,\|x_t(u) - y_t\|}_{(c)},\end{aligned}$$

for any $y_t \in \mathcal{X}$*, with the convention that* $0/0 = 0$ *when* $|S_t| = 0$*.*

This will be particularly useful when setting y_t as the prediction of an omniscient agent knowing the gradients of all incurred losses up to time $t-1$. In this case, Term (B) is the part of the regret related to the loss incurred by the prediction of the omniscient agent, and Term (A) and (C) constitute the part of the regret related to the deviations with respect to these predictions.

Proof. By definition of the regret,

$$\begin{aligned}
\mathrm{Reg}_T(u) &= \sum_t^T \sum_{v\in\mathcal{V}} \frac{1}{|S_t|}(\ell_t(x_t(u),v) - \ell_t(x^*,v))\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\} \\
&= \sum_t^T \sum_{v\in\mathcal{V}} \frac{1}{|S_t|}(\ell_t(x_t(u),v) - \ell_t(y_t,v) + \ell_t(y_t,v) - \ell_t(x^*,v)) \times \mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\} \\
&\leq \sum_t^T \sum_{v\in\mathcal{V}} \frac{1}{|S_t|}(L\,\|x_t(u)-y_t\| + \ell_t(y_t,v) - \ell_t(x^*,v)) \times \mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\}\,,
\end{aligned}$$

because $\ell_t(\cdot,v)$ is L-Lipschitz over the set $\mathcal{X}$ w.r.t the norm $\|\cdot\|$, i.e. $|\ell_t(x,v)-\ell_t(y,v)| \leq L\|x-y\|, \forall\, x,y\in X$. Next, because we need to introduce individual gradients, we add and remove each $\ell_t(x_t,v)$:

$$\begin{aligned}
\mathrm{Reg}_T(u) &\leq \sum_t^T \sum_{v\in\mathcal{V}} \frac{1}{|S_t|}\,(L\,\|x_t(u)-y_t\| + \ell_t(y_t,v) - \ell_t(x_t(v),v) + \ell_t(x_t(v),v) - \ell_t(x^*,v))\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\} \\
&\leq \sum_t^T \sum_{v\in\mathcal{V}} \frac{1}{|S_t|}\,(L\,\|x_t(u)-y_t\| + L\,\|x_t(v)-y_t\| + \ell_t(x_t(v),v) - \ell_t(x^*,v))\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\}\,,
\end{aligned}$$

where we used again the Lipschitzness of the loss functions. Then by convexity of $\ell_t(\cdot,v)$,

$$\mathrm{Reg}_T(u) \leq \sum_t^T \sum_{v\in\mathcal{V}} \frac{1}{|S_t|}(L\,\|x_t(u)-y_t\| + L\,\|x_t(v)-y_t\| + \langle\nabla\ell_t(x_t(v),v), x_t(v)-x^*\rangle)\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\}\,,$$

which can be rewritten as

$$\begin{aligned}
\mathrm{Reg}_T^{\mathrm{net}} &\leq \sum_{t=1}^T \mathbb{I}\{u\in S_t\} L\,\|x_t(u)-y_t\| + \sum_{t=1}^T \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}\mathbb{I}\{u\in S_t\}}{|S_t|} L\,\|x_t(v)-y_t\| \\
&\quad + \sum_{t=1}^T \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}\mathbb{I}\{u\in S_t\}}{|S_t|}\,\langle\nabla\ell_t(x_t(v),v), x_t(v)-x^*\rangle \\
&= \sum_{t=1}^T \mathbb{I}\{u\in S_t\} L\,\|x_t(u)-y_t\| + \sum_{t=1}^T \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}\mathbb{I}\{u\in S_t\}}{|S_t|} L\,\|x_t(v)-y_t\| \\
&\quad + \sum_{t=1}^T \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}\mathbb{I}\{u\in S_t\}}{|S_t|}\,\langle\nabla\ell_t(x_t(v),v), x_t(v)-y_t+y_t-x^*\rangle \\
&\leq \sum_{t=1}^T \mathbb{I}\{u\in S_t\} L\,\|x_t(u)-y_t\| + 2\sum_{t=1}^T \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}\mathbb{I}\{u\in S_t\}}{|S_t|} L\,\|x_t(v)-y_t\| \\
&\quad + \sum_{t=1}^T \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}\mathbb{I}\{u\in S_t\}}{|S_t|}\,\langle\nabla\ell_t(x_t(v),v), y_t-x^*\rangle\,,
\end{aligned}$$

again by the Lipschitzness of the losses. Finally,

$$\begin{aligned}\mathrm{Reg}_T^{\mathrm{net}} \leq & \underbrace{2\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\frac{\mathbb{I}\{u,v\in S_t\}}{|S_t|}L\left\|x_t(v)-y_t\right\|}_{(a)} + \underbrace{\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\frac{\mathbb{I}\{u,v\in S_t\}}{|S_t|}\left\langle\nabla\ell_t(x_t(v),v),y_t-x^*\right\rangle}_{(b)} \\ & + \underbrace{\sum_{t=1}^{T}\mathbb{I}\{u\in S_t\}L\left\|x_t(u)-y_t\right\|}_{(c)}, \end{aligned} \tag{C.2}$$

concluding the proof. □

Lemma C.2 (Network regret's decomposition)**.** *Denoting by* $x^* = \arg\min_{x\in\mathcal{X}}\sum_{t=1}^{T}\ell_t^{\mathrm{net}}(S_t,x)$*, the network regret*

$$\mathrm{Reg}_T^{\mathrm{net}} = \sum_{t=1}^{T}\frac{1}{|S_t|}\sum_{v\in S_t}\ell_t^{\mathrm{net}}\big(S_t,x_t(v)\big) - \min_{x\in\mathcal{X}}\sum_{t=1}^{T}\ell_t^{\mathrm{net}}(S_t,x)$$

can be decomposed in the following way

$$\mathrm{Reg}_T^{\mathrm{net}} \leq \underbrace{3\sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\frac{\mathbb{I}\{u\in S_t\}}{|S_t|}L\left\|x_t(u)-y_t\right\|}_{(A)} + \underbrace{\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\frac{\mathbb{I}\{v\in S_t\}}{|S_t|}\left\langle\nabla\ell_t(x_t(v),v),y_t-x^*\right\rangle}_{(B)},$$

for any $y_t\in\mathcal{X}$*, with the convention that* $0/0=0$ *when* $|S_t|=0$*.*

Proof. By definition of the regret,

$$\begin{aligned}\mathrm{Reg}_T^{\mathrm{net}} &= \sum_{t}^{T}\sum_{u\in\mathcal{V}}\sum_{v\in\mathcal{V}}\frac{1}{|S_t|^2}(\ell_t(x_t(u),v)-\ell_t(x^*,v))\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\} \\ &= \sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\sum_{v\in\mathcal{V}}\frac{1}{|S_t|^2}(\ell_t(x_t(u),v)-\ell_t(y_t,v)+\ell_t(y_t,v)-\ell_t(x^*,v))\times\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\} \\ &\leq \sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\sum_{v\in\mathcal{V}}\frac{1}{|S_t|^2}(L\left\|x_t(u)-y_t\right\|+\ell_t(y_t,v)-\ell_t(x^*,v))\times\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\},\end{aligned}$$

because $\ell_t(\cdot,v)$ is L-Lipschitz over the set $\mathcal{X}$ w.r.t the norm $\|\cdot\|$, i.e. $|\ell_t(x,v)-\ell_t(y,v)|\leq L\|x-y\|,\forall\, x,y\in X$. Next, because we need to introduce individual gradients, we add and remove each $\ell_t(x_t,v)$:

$$\begin{aligned}\mathrm{Reg}_T^{\mathrm{net}} &\leq \sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\sum_{v\in\mathcal{V}}(L\left\|x_t(u)-y_t\right\|+\ell_t(y_t,v)-\ell_t(x_t(v),v)+\ell_t(x_t(v),v)-\ell_t(x^*,v))\times\frac{1}{|S_t|^2}\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\} \\ &\leq \sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\sum_{v\in\mathcal{V}}(L\left\|x_t(u)-y_t\right\|+L\left\|x_t(v)-y_t\right\|+\ell_t(x_t(v),v)-\ell_t(x^*,v))\times\frac{1}{|S_t|^2}\mathbb{I}\{u\in S_t\}\mathbb{I}\{v\in S_t\},\end{aligned}$$

where we used again the Lipschitzness of the loss functions. Then by convexity of $\ell_t(\cdot, v)$,

$$\mathrm{Reg}_T^{\mathrm{net}} \leq \sum_{t=1}^{T} \sum_{u\in\mathcal{V}} \sum_{v\in\mathcal{V}} \frac{1}{|S_t|^2} (L \left\|x_t(u) - y_t\right\| + L \left\|x_t(v) - y_t\right\| + \langle \nabla \ell_t(x_t(v), v), x_t(v) - x^* \rangle) \times \mathbb{I}\{u \in S_t\} \mathbb{I}\{v \in S_t\},$$

which can be rewritten as

$$\begin{aligned}
\mathrm{Reg}_T^{\mathrm{net}} &\leq \sum_{t=1}^{T} \sum_{u\in\mathcal{V}} \frac{\mathbb{I}\{u \in S_t\}}{|S_t|} L \left\|x_t(u) - y_t\right\| + \sum_{t=1}^{T} \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|} L \left\|x_t(v) - y_t\right\| \\
&\quad + \sum_{t=1}^{T} \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|} \langle \nabla \ell_t(x_t(v), v), x_t(v) - x^* \rangle \\
&= 2 \sum_{t=1}^{T} \sum_{u\in\mathcal{V}} \frac{\mathbb{I}\{u \in S_t\}}{|S_t|} L \left\|x_t(u) - y_t\right\| + \sum_{t=1}^{T} \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|} \langle \nabla \ell_t(x_t(v), v), x_t(v) - y_t + y_t - x^* \rangle \\
&\leq 2 \sum_{t=1}^{T} \sum_{u\in\mathcal{V}} \frac{\mathbb{I}\{u \in S_t\}}{|S_t|} L \left\|x_t(u) - y_t\right\| + \sum_{t=1}^{T} \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|} L \left\|x_t(v) - y_t\right\| \\
&\quad + \sum_{t=1}^{T} \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|} \langle \nabla \ell_t(x_t(v), v), y_t - x^* \rangle ,
\end{aligned}$$

again by the Lipschitzness of the losses. Finally,

$$\mathrm{Reg}_T^{\mathrm{net}} \leq \underbrace{3 \sum_{t=1}^{T} \sum_{u\in\mathcal{V}} \frac{\mathbb{I}\{u \in S_t\}}{|S_t|} L \left\|x_t(u) - y_t\right\|}_{(A)} + \underbrace{\sum_{t=1}^{T} \sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|} \langle \nabla \ell_t(x_t(v), v), y_t - x^* \rangle}_{(B)} \tag{C.3}$$

concluding the proof. □

Lemma C.3. *Assuming that for $k = 1 \dots T$, W_k are doubly stochastic matrices and i.i.d., we have, $\forall v \in \mathcal{V}$, $\forall s, t \in [T]$ such that $t > s$,*

$$\begin{aligned}
\mathbb{E}\left[\left(W_t \cdots W_{s+1} e_v - \frac{\mathbf{1}}{N}\right)^T \left(W_t \cdots W_{s+1} e_v - \frac{\mathbf{1}}{N}\right)\right] &\leq e_v^T e_v \left\|\mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right\|_{\mathrm{op}}^{t-s} \\
&\leq \lambda_2(\mathbb{E}[W_1 W_1^\top])^{t-s} . \qquad \text{(C.4)}
\end{aligned}$$

This can be derived exactly as in the proof of (Lei et al., 2020, Lemma 2). For completeness, we provide a quick justification.

Proof. Let $\widetilde{W}_k = W_k - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$ and assume

$$\mathbb{E}\left[\left\|W_{k-1} \cdots W_{s+1} e_v - \frac{\mathbf{1}}{N}\right\|_2^2\right] \leq e_v^T e_v \left\|\mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right\|_{\mathrm{op}}^{k-s-1} \qquad \text{for some } k-1 > s.$$

Let $\mathcal{F}_{k-1}$ be the σ-algebra generated by all random events up to time $k-1$. We have that

$$\mathbb{E}\left[\left\|W_k^\top \cdots W_{s+1}^\top e_v - \frac{\mathbf{1}}{N}\right\|_2^2\right] = \mathbb{E}\left[e_v^T \widetilde{W}_{s+1}^\top \cdots \widetilde{W}_{k-1}^\top \widetilde{W}_k^\top \widetilde{W}_k \widetilde{W}_{k-1} \cdots \widetilde{W}_{s+1} e_v\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[e_v^T \widetilde{W}_{s+1}^\top \cdots \widetilde{W}_{k-1}^\top \mathbb{E}[\widetilde{W}_k^\top \widetilde{W}_k \mid \mathcal{F}_{k-1}] \widetilde{W}_{k-1} \cdots \widetilde{W}_{s+1} e_v\right] \\
&= \mathbb{E}\left[e_v^T \widetilde{W}_{s+1}^\top \cdots \widetilde{W}_{k-1}^\top \mathbb{E}[\widetilde{W}_1^\top \widetilde{W}_1] \widetilde{W}_{k-1} \cdots \widetilde{W}_{s+1} e_v\right] \\
&\qquad\qquad \text{(by independence of } W_k\text{)} \\
&\leq \left\| \mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top \right\|_{\mathrm{op}} e_v^T e_v \left\| \mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top \right\|_{\mathrm{op}}^{k-s-1} \\
&\leq \lambda_2(\mathbb{E}[W_1 W_1^\top]) e_v^T e_v \left\| \mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top \right\|_{\mathrm{op}}^{k-s-1}
\end{aligned}
$$

which by induction, suffices to prove Equation (C.4). □

C.4 Omitted details in Section 4.4

C.4.1 Regret upper bounds in Expectation

We start by bounding the network regret, and making a few comments, before getting to the slightly harder proof of Theorem C.1.

Theorem C.1. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta > 0$ and i.i.d gossip matrices W_t. Then, the expected network regret can be bounded by*

$$
\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{D^2 I_p}{\eta \bar{p}} + \frac{L^2 \eta}{\mu}(6 + 2I_p + 3\sqrt{\bar{p}N}\frac{\rho}{1-\rho})T\,, \tag{C.5}
$$

where $\rho = \sqrt{\lambda_2(\mathbb{E}[W_1 W_1^\top])}$, and $I_p = \bar{p}/p_{\min}$ is an imbalance factor. In the p-uniform case, we have

$$
\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{D^2}{p\eta} + \frac{L^2}{\mu}\eta(8 + 3\sqrt{pN}\frac{\rho}{1-\rho})T\ . \tag{C.6}
$$

If, in addition, $\eta = \frac{(D/L)\sqrt{\mu}}{2\sqrt{2}p^{3/4}N^{1/4}\sqrt{T}}$, then

$$
\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq 2\sqrt{2}\frac{DL}{\sqrt{\mu}}\frac{N^{1/4}}{p^{1/4}}\frac{1}{1-\rho}\sqrt{T}\,. \tag{C.7}
$$

Lower bound on activation probabilities. Our analysis assumes $\sum_v p_v \geq 1$ ensuring that the fraction of rounds with zero active agents is vanishingly small with high probability. If this assumption is dropped, the time horizon T in our bounds is replaced by the expected number $\widetilde{T} = \big(1 - \Pi_{v\in\mathcal{V}}(1-p_v)\big)T$ of time steps when there is at least one active agent (if no agents are active in a give step, then that step does not contribute to the regret). However, optimizing the learning rate with respect to $\widetilde{T}$ is problematic because this quantity depends on the activation probabilities. On the other hand, note that $\widetilde{T} \leq \big(1 - (1-p_{\max})^N\big)T \leq p_{\max}NT$. Hence, when $p_{\max}$ is known and smaller than $\frac{1}{N}$ (which, in turn, implies that $\sum_v p_v < 1$), we can tune the learning rate using $p_{\max}NT < T$.

Proof. In this proof we denote by $x^* = \arg\min_{x\in\mathcal{X}} \sum_{t=1}^T \ell_t^{\mathrm{net}}(S_t, x)$. The proof relies on the use of an omniscient agent knowing the gradients of all incurred losses up to time $t-1$.

Let us define the quantities $\bar{z}_t$ and $\bar{g}_t$

$$\bar{g}_t = \frac{1}{N}\sum_{v\in\mathcal{V}} g_t(x_t(v), v) = \sum_{v\in\mathcal{V}} \frac{\mathbb{1}(v \in S_t)}{N}\nabla\ell_t(x_t(v), v)$$
$$\bar{z}_t = \frac{1}{N}\sum_{v\in\mathcal{V}} z_t(v).$$

Then the decision of the omniscient agent is defined as

$$\bar{x}_t = \operatorname{argmin}_{x\in X}\left\{\langle \bar{z}_t, x\rangle + \frac{1}{\eta}\psi(x)\right\}.$$

Note that

$$\bar{z}_{t+1} = \bar{z}_t + \bar{g}_t. \tag{C.8}$$

The proof of the theorem relies on Lemma C.2, where y_t is set to $\bar{x}_t$.

$$\mathrm{Reg}_T^{\mathrm{net}} \le \underbrace{3\sum_{t=1}^{T}\sum_{u\in\mathcal{V}} \frac{\mathbb{I}\{u \in S_t\}}{|S_t|} L\,\|x_t(u) - \bar{x}_t\|}_{(A)} + \underbrace{\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|}\,\langle \nabla\ell_t(x_t(v), v), \bar{x}_t - x^*\rangle}_{(B)}. \tag{C.9}$$

General case. We start by analyzing the general case. Let us focus on Term (B) first.

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v \in S_t\}}{|S_t|}\,\langle \nabla\ell_t(x_t(v), v), \bar{x}_t - x^*\rangle\right] &= \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \mathbb{E}\Big(\frac{\mathbb{I}\{v \in S_t\}}{1 + \sum_{u\in\mathcal{V}\setminus v}\mathbb{I}\{u \in S_t\}}\Big)\mathbb{E}\left[\langle \nabla\ell_t(x_t(v), v), \bar{x}_t - x^*\rangle\right] \\
&\qquad \text{(by independence and a slight rewriting)} \\
&= \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} p_v c_v \mathbb{E}\left[\langle \nabla\ell_t(x_t(v), v), \bar{x}_t - x^*\rangle\right] \\
&\le \max_{v\in\mathcal{V}} c_v \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} p_v \mathbb{E}\left[\langle \nabla\ell_t(x_t(v), v), \bar{x}_t - x^*\rangle\right],
\end{aligned}$$

where $c_v \coloneqq \mathbb{E}\big(\frac{1}{1+\sum_{u\in\mathcal{V}\setminus v}\mathbb{I}\{u\in S_t\}}\big) \le 1$. Recall that $\widetilde{T} = \big(1 - \Pi_{v\in\mathcal{V}}(1-p_v)\big)T$ denotes the expected number of time steps when there is at least one active agent. In the p-uniform case for example, $c_v = \frac{\widetilde{T}}{TpN} = \frac{1-(1-p)^N}{Np} \le 1$. This holds because, on the one hand,

$$\sum_{v\in\mathcal{V}} \mathbb{E}\left[\frac{\mathbb{I}\{v \in S_t\}}{|S_t|}\right] = \sum_{v\in\mathcal{V}} pc_v = Npc_v$$

due to all c_v being equal. On the other hand,

$$\sum_{v\in\mathcal{V}} \mathbb{E}\left[\frac{\mathbb{I}\{v \in S_t\}}{|S_t|}\right] = \Pr(S_t \neq \emptyset) = \frac{\widetilde{T}}{T}. \tag{C.10}$$

In the non-uniform case, $c_v \le \frac{1}{Np_{\min}}$. Indeed, bounding $\mathbb{E}\big(\frac{1}{1+\sum_{u\in\mathcal{V}\setminus v}\mathbb{I}\{u\in S_t\}}\big)$ is essentially bounding $\mathbb{E}\big(\frac{1}{1+X_{v,t}}\big)$ where $X_{v,t}$ is a Poisson binomial variable with $N-1$ probability parameters : $\{p_u,\ \forall u \in$

$\mathcal{V} \setminus v\}$. Then stochastically, $X_{t,v}$ is lower bounded by Binomial$(N-1, p_{\min})$. Since $\frac{1}{1+s}$ is decreasing, this implies:

$$\mathbb{E}\left[\frac{1}{1+X_{t,v}}\right] \leq \mathbb{E}\left[\frac{1}{1+\mathrm{Bin}(N-1,p_{\min})}\right]$$

Yet, we know from our analysis of c_v in the p-uniform case that

$$\mathbb{E}\left[\frac{1}{1+\mathrm{Bin}(N-1,p_{\min})}\right] = \frac{1-(1-p_{\min})^N}{Np_{\min}} \leq \frac{1}{Np_{\min}}$$

Finally $c_v \leq \frac{1}{Np_{\min}}$

Since $\bar{x}_t$ are the predictions of FTRL on linear losses $\langle \bar{g}_t, \cdot \rangle$, we know from standard FTRL analysis (Orabona, 2025, Corollary 7.7),

$$\frac{1}{N}\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \langle \nabla \ell_t(x_t(v),v), \bar{x}_t - x^* \rangle \, \mathbb{I}\{v \in S_t\} \leq \frac{\psi(x^*)}{\eta} + \frac{L^2}{\mu}\sum_{t=1}^{T}\eta\frac{|S_t|^2}{N^2} \tag{C.11}$$

which by taking expectation and using the independence of S_t and $x_t(v)$ leads to

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} p_v \langle \nabla \ell_t(x_t(v),v), \bar{x}_t - x^* \rangle\right] \leq \frac{\psi(x^*)N}{\eta} + \frac{L^2}{\mu}\eta\sum_{t=1}^{T}\left(N\bar{p} + \bar{p}(1-\bar{p}) - \sigma_p^2\right),$$

where $\bar{p} = \frac{1}{N}\sum_{v\in\mathcal{V}} p_v$ and $\sigma_p^2 = \frac{1}{N}\sum_{v\in\mathcal{V}} p_v^2 - \bar{p}^2$, where $\sigma_p^2 = \frac{1}{N}\sum_{v\in\mathcal{V}} p_v^2 - \bar{p}^2$ is the variance.

This holds because $\mathbb{E}[|S_t|^2] = \mathbb{E}[|S_t|]^2 + \mathrm{Var}(|S_t|)$, and $|S_t|$ is the sum of independent Bernoulli of parameter p_v, so that $\mathrm{Var}(|S_t|) = \sum_{v\in\mathcal{V}} p_v(1-p_v)$, which can also be written as $N\bar{p}(1-\bar{p}) - N\sigma_p^2$. Hence

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}}{|S_t|} \langle \nabla \ell_t(x_t(v),v), \bar{x}_t - x^* \rangle\right] &\leq \max c_v \left(\frac{\psi(x^*)N}{\eta} + \frac{L^2}{\mu}\sum_{t=1}^{T}\eta(N\bar{p} + \bar{p}(1-\bar{p}) - \sigma_p^2)\right) \\
&\leq \frac{1}{Np_{\min}}\left(\frac{D^2N}{\eta} + \frac{L^2}{\mu}\sum_{t=1}^{T}\eta(N\bar{p} + \bar{p}(1-\bar{p}) - \sigma_p^2)\right) \\
&\leq \frac{D^2 I_p}{\eta\bar{p}} + \frac{L^2}{\mu}\sum_{t=1}^{T}\eta(N\bar{p} + \bar{p}(1-\bar{p}) - \sigma_p^2)/(Np_{\min}) \qquad \text{(C.12)} \\
&\leq \frac{D^2 I_p}{\eta\bar{p}} + 2\frac{L^2}{\mu}\eta I_p T\,. \qquad \text{(C.13)}
\end{aligned}$$

Regarding Term (A), since

$$\bar{x}_t = \mathrm{argmin}_{x\in X}\left\{\langle \bar{z}_t, x\rangle + \frac{1}{\eta}\psi(x)\right\},$$

and

$$x_t(v) = \mathrm{argmin}_{x\in X}\left\{\langle z_t(v), x\rangle + \frac{1}{\eta}\psi(x)\right\},$$

we have

$$\|x_t(v) - \bar{x}_t\| \leq \eta/\mu \, \|z_t(v) - \bar{z}_t\|_* \,, \tag{C.14}$$

thanks to the duality between strong convexity and smoothness (Orabona, 2025, Theorem 6.11). For any $t \in [T]$ and any $v \in [N]$, we have

$$Z_{t+1} = W_t Z_t + \Gamma_t = W_t W_{t-1} Z_{t-1} + W_t \Gamma_{t-1} + \Gamma_t = \sum_{s=1}^{t-1} W_t \cdots W_{s+1} \Gamma_s + \Gamma_t \,.$$

Simultaneously, we have

$$\bar{z}_{t+1} = \frac{1}{N} \sum_{s=1}^{t} \mathbf{1}^\top \Gamma_s,$$

so that

$$\begin{aligned} Z_{t+1} - \mathbf{1}\bar{z}_{t+1} &= \sum_{s=1}^{t} W_t \cdots W_{s+1} \Gamma_s + \Gamma_t - \frac{1}{N} \mathbf{1}\mathbf{1}^\top \Gamma_s \\ &= \sum_{s=1}^{t-1} \left[W_t \cdots W_{s+1} - \frac{1}{N} \mathbf{1}\mathbf{1}^\top \right] \Gamma_s + \Gamma_t - \frac{1}{N} \mathbf{1}\mathbf{1}^\top \Gamma_t \,. \end{aligned}$$

In turn,

$$\begin{aligned} z_{t+1}(v) - \bar{z}_{t+1} &= (Z_{t+1} - \mathbf{1}\bar{z}_{t+1})^T e_v \\ &= \sum_{s=1}^{t-1} \Gamma_s^T \left[W_t \cdots W_{s+1} - \frac{1}{N} \mathbf{1}\mathbf{1}^\top \right]^T e_v + \Gamma_t^\top (I - \frac{1}{N} \mathbf{1}\mathbf{1}^\top) e_v, \end{aligned}$$

so that we can compute :

$$\begin{aligned} \|z_{t+1}(v) - \bar{z}_{t+1}\|_* &= \left\| \sum_{s=0}^{t-1} \left(\sum_{u=1}^{N} \left([W_t \cdots W_{s+1}]_{u,v} - \frac{1}{N} \right) g_s(x_s(u), u) \right) + g_t(x_t(v), v) - \bar{g}_t \right\|_* \\ &\leq \sum_{s=0}^{t-1} \left\| \sum_{u=1}^{N} \left([W_t \cdots W_{s+1}]_{u,v} - \frac{1}{N} \right) g_s(x_s(u), u) \right\|_* + \|g_t(x_t(v), v) - \bar{g}_t\|_* \\ &\qquad \text{(triangular inequality)} \\ &\leq \sum_{s=0}^{t-1} \left(\sum_{u=1}^{N} \left| [W_t \cdots W_{s+1}]_{u,v} - \frac{1}{N} \right| \|g_s(x_s(u), u)\|_* \right) + \|g_t(x_t(v), v) - \bar{g}_t\|_* \\ &= \sum_{s=0}^{t-1} \left(\sum_{u \in S_t} \left| [W_t \cdots W_{s+1}]_{u,v} - \frac{1}{N} \right| \|g_s(x_s(u), u)\|_* \right) + \|g_t(x_t(v), v) - \bar{g}_t\|_* \\ &\leq \sum_{s=0}^{t-1} \left(\left(\sum_{u \in S_t} \left| [W_t \cdots W_{s+1}]_{u,v} - \frac{1}{N} \right|^2 \right)^{1/2} \left(\sum_{u \in S_t} \|g_s(x_s(u), u)\|_*^2 \right)^{1/2} \right) + \|g_t(x_t(v), v) - \bar{g}_t\|_* \end{aligned}$$

Since $\|g_s(x_s(u), u)\|_* \leq \mathbb{I}\{u \in S_t\} L$, we have $\left(\sum_{u \in S_t} \|g_s(x_s(u), u)\|_*^2 \right)^{1/2} \leq L\sqrt{|S_t|}$.

We also have $\left(\sum_{u \in S_t} \left| [W_t \cdots W_{s+1}]_{u,v} - \frac{1}{N} \right|^2 \right)^{1/2} = \left\| W_t \cdots W_{s+1} e_v - \frac{1}{N} \mathbf{1} \right\|_2$ so that in conclu-

sion,

$$\|z_{t+1}(v) - \bar{z}_{t+1}\|_* \leq \sum_{s=0}^{t-1} L\sqrt{|S_t|} \left\| W_t \cdots W_{s+1} e_v - \frac{1}{N}\mathbf{1} \right\|_2 + 2L\,. \tag{C.15}$$

By independence of S_t and $W_t \dots W_0$, we have

$$\mathbb{E}\left[\|z_{t+1}(v) - \bar{z}_{t+1}\|_*\right] \leq \sum_{s=0}^{t-1} L\mathbb{E}\left[\sqrt{|S_t|}\right] \mathbb{E}\left[\left\| W_t \cdots W_{s+1} e_v - \frac{1}{N}\mathbf{1} \right\|_2\right] + 2L\,.$$

To bound the left hand side of this expression, we need to bound $\mathbb{E}[\sqrt{|S_t|}]$. By Jensen's inequality,

$$\mathbb{E}[\sqrt{|S_t|}] \leq \sqrt{\mathbb{E}[|S_t|]} = \sqrt{\bar{p}N}\,.$$

Hence

$$\mathbb{E}\left[\|z_{t+1}(v) - \bar{z}_{t+1}\|_*\right] \leq \sum_{s=0}^{t-1} L\sqrt{\bar{p}N}\mathbb{E}\left[\left\| W_t \cdots W_{s+1} e_v - \frac{1}{N}\mathbf{1} \right\|_2\right] + 2L\,.$$

So we also have, thanks to Lemma C.3,

$$\begin{aligned}\mathbb{E}\left[\|z_{t+1}(v) - \bar{z}_{t+1}\|_*\right] &\leq \sum_{s=0}^{t-1} L\sqrt{\bar{p}N}\lambda_2(\mathbb{E}[W_1 W_1^\top])^{t-s} + 2L \\ &\leq L\sqrt{\bar{p}N}\,\frac{\rho}{1-\rho} + 2L\end{aligned}$$

Using Equation (C.14),

$$\mathbb{E}\left[\|x_t(u) - \bar{x}_t\|\right] \leq \frac{L\eta}{\mu}\left(\sqrt{\bar{p}N}\,\frac{\rho}{1-\rho} + 2\right) \tag{C.16}$$

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{u\in\mathcal{V}} \frac{1}{|\mathcal{S}_t|}\mathbb{I}\{u \in \mathcal{S}_t\} L\|x_t(u) - \bar{x}_t\|\right] &\leq \sum_{t=1}^{T}\sum_{u\in\mathcal{V}} L\mathbb{E}\left[\frac{1}{|\mathcal{S}_t|}\mathbb{I}\{u \in \mathcal{S}_t\}\right] \mathbb{E}\left[\|x_t(u) - \bar{x}_t\|\right] \\ &\qquad \text{(by independence)} \\ &\leq \eta\frac{L^2}{\mu}\left(2 + \sqrt{\bar{p}N}\frac{\rho}{1-\rho}\right)\widetilde{T}\,. \quad \text{(using the definition of } \widetilde{T})\end{aligned}$$

Combining this with Equation (C.13) and using Equation (C.9), we get

$$\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{D^2 I_p}{\eta\bar{p}} + \frac{L^2}{\mu}\eta(6 + \frac{(N+1)\bar{p}}{Np_{\min}}T/\widetilde{T} + 3\sqrt{\bar{p}N}\frac{\rho}{1-\rho})\widetilde{T}\,.$$

Consequently

$$\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{D^2 I_p}{\eta\bar{p}} + \frac{L^2}{\mu}\eta(6 + 2I_p T/\widetilde{T} + 3\sqrt{\bar{p}N}\frac{\rho}{1-\rho})\widetilde{T}\,.$$

We also have

$$\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{D^2 I_p}{\eta\bar{p}} + \frac{L^2}{\mu}\eta(6 + 2I_p + 3\sqrt{\bar{p}N}\frac{\rho}{1-\rho})T\,,$$

which is less tight in general but sufficient with the assumption that $\sum_{v\in\mathcal{V}} p_v \geq 1$.
This yields the first result of Theorem 4.1.

p-uniform case. Equation (4.4) follows from observing that $I_p = 1$ in the p-uniform case.

$$\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{D^2}{\eta p} + \frac{L^2}{\mu}\eta(8 + 3\sqrt{pN}\frac{\rho}{1-\rho})T\,.$$

Finally, Equation (4.5) follows from simple computations. □

Theorem 4.1. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta > 0$ and i.i.d gossip matrices W_t. Then, the expected individual regret for each $u \in \mathcal{V}$ can be bounded by*

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq \frac{D^2 I_p}{\eta\bar{p}} + \frac{L^2\eta}{\mu}(6 + 2I_p + 3p_u\sqrt{\bar{p}N}\frac{\rho}{1-\rho})T\,, \tag{4.3}$$

where $\rho = \sqrt{\lambda_2(\mathbb{E}[W_1 W_1^\top])}$, and $I_p = \bar{p}/p_{\min}$ is an imbalance factor. In the p-uniform case, we have, for all $u \in \mathcal{V}$,

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq \frac{D^2}{p\eta} + \frac{L^2}{\mu}\eta(8 + 3p\sqrt{pN}\frac{\rho}{1-\rho})T\ . \tag{4.4}$$

If, in addition, $\eta = \frac{(D/L)\sqrt{\mu}}{2\sqrt{2}N^{1/4}\sqrt{T}p^{5/4}}$, then for any $u \in \mathcal{V}$,

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq 2\sqrt{2}\frac{DL}{\sqrt{\mu}}N^{1/4}p^{1/4}\frac{1}{1-\rho}\sqrt{T}\,. \tag{4.5}$$

Proof. In this proof we denote by $x^* = \arg\min_{x\in\mathcal{X}} \sum_{t\leq T; u\in S_t} \ell_t^{\mathrm{net}}(S_t, x)$.

We will use Equation (C.2) with the same choice of $\bar{x}_t$ as in the proof of Theorem C.1. It is obvious that with this choice, we can bound Term (b) by Term (B).

$$\underbrace{\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{v\in S_t\}}{|S_t|}\left\langle \nabla\ell_t(x_t(v), v), \bar{x}_t - x^*\right\rangle}_{(b)} \leq \underbrace{\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{u, v\in S_t\}}{|S_t|}\left\langle \nabla\ell_t(x_t(v), v), \bar{x}_t - x^*\right\rangle}_{(B)}$$

So, thanks to the analysis in the proof of Theorem C.1,

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] \leq \frac{D^2 I_p}{\eta\bar{p}} + 2\frac{L^2}{\mu}\eta I_p T + \underbrace{2\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \frac{\mathbb{I}\{u, v\in S_t\}}{|S_t|} L\left\|x_t(v) - y_t\right\|}_{(a)} + \underbrace{\sum_{t=1}^{T}\mathbb{I}\{u\in S_t\} L\left\|x_t(u) - y_t\right\|}_{(c)}\,.$$

It remains to upper bound Terms (a) and (c). From Equation (C.16), we get

$$\mathbb{E}\left[\|x_t(u) - \bar{x}_t\|\right] \leq \frac{L\eta}{\mu}\left(\sqrt{\bar{p}N}\ \frac{\rho}{1-\rho} + 2\right)$$

Since

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{I}\{u\in S_t\}L\left\|x_t(u)-y_t\right\|\right]\leq\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{I}\{u\in S_t\}\right]L\mathbb{E}\left[\left\|x_t(u)-y_t\right\|\right]\leq Lp_uT\times\frac{L\eta}{\mu}\left(\sqrt{\bar{p}N}\ \frac{\rho}{1-\rho}+2\right)$$

□

by independence, we have

$$\mathbb{E}[(c)]\leq\eta\frac{L^2}{\mu}\left(2+\sqrt{\bar{p}N}\frac{\rho}{1-\rho}\right)p_uT\,.$$

It is also easy to prove that $\sum_{v\in\mathcal{V}}\frac{\mathbb{I}\{u,v\in S_t\}}{|S_t|}\leq\mathbb{I}\{u\in S_t\}$. Then

$$\mathbb{E}[(a)]=\mathbb{E}\left[2\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\frac{\mathbb{I}\{u,v\in S_t\}}{|S_t|}L\left\|x_t(v)-y_t\right\|\right]\leq2\mathbb{E}[(c)]\leq\eta\frac{L^2}{\mu}\left(2+\sqrt{\bar{p}N}\frac{\rho}{1-\rho}\right)p_uT$$

And

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big]\leq\frac{D^2I_p}{\eta\bar{p}}+\frac{L^2}{\mu}\eta(6p_u+2I_p+3p_u\sqrt{\bar{p}N}\frac{\rho}{1-\rho})T\,.$$

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big]\leq\frac{D^2I_p}{\eta\bar{p}}+\frac{L^2}{\mu}\eta(6+2I_p+3p_u\sqrt{\bar{p}N}\frac{\rho}{1-\rho})T\,.$$

C.4.2 High-probability Upper Bounds

Preliminary result The main difference between the proof of the bound in expectation and that of the high probability bound is the use of the following Lemma to bound the deviation between $W_t\cdots W_{s+1}$ and $\frac{1}{N}\mathbf{1}\mathbf{1}^\top$.

Lemma C.4. *Assuming that for $k=1\ldots T$, W_k are doubly stochastic matrices and i.i.d., we have, $\forall v\in\mathcal{V}$, $\forall s,t\in[T]$ such that $t>s$,*

$$\Pr\left(\left\|W_t\cdots W_{s+1}e_v-\frac{1}{N}\mathbf{1}\right\|_2\geq\epsilon\right)\leq\frac{\lambda_2(\mathbb{E}[W^2])^{t-s}}{\epsilon^2}.$$

When $t-s\geq\frac{3\log\epsilon^{-1}}{\log\lambda_2[W^2]^{-1}}=t^*$, we have

$$\Pr\left(\left\|W_t\cdots W_{s+1}e_v-\frac{1}{N}\mathbf{1}\right\|_2\geq\epsilon\right)\leq\epsilon$$

This lemma is from Boyd et al. (2006). We provide a proof for completeness.

Proof. By applying Markov's inequality, we have

$$\Pr\left(\left\|W_t\cdots W_{s+1}e_v-\frac{1}{N}\mathbf{1}\right\|_2\geq\epsilon\right)\leq\frac{\mathbb{E}\left(\left\|W_t\cdots W_{s+1}e_v-\frac{1}{N}\mathbf{1}\right\|_2^2\right)}{\epsilon^2}\,.$$

Denoting by $\widetilde{W}_k=W_k-\frac{1}{N}\mathbf{1}\mathbf{1}^\top$, we need to prove that $\mathbb{E}\left(\left\|W_t\cdots W_{s+1}e_v-\frac{1}{N}\mathbf{1}\right\|_2^2\right)$ is bounded by $\lambda_2(\mathbb{E}[W^2])^{t-s}$. This is done by using Lemma C.3. □

C.4.3 High-probability bound on the network regret

Like for the bounds in expectation, bounding the network regret is a bit easier than the individual regret. We start with the network regret before proceeding to the individual regret.

Theorem C.2. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta > 0$. Then, with probability $1-\delta$, the network regret is bounded by*

$$\mathrm{Reg}_T^{\mathrm{net}} \leq N\left(\frac{D^2}{\eta} + \frac{L^2}{\mu}\eta T\right) + 3\eta T N \frac{L^2}{\mu}\left(\frac{3}{1-\rho^2}\log\frac{NT^2}{\delta} + 3\right) .$$

There are notable differences between the bound in expectation provided by Theorem C.1 and this one. First, the high-probability bound has a $1/(1-\rho^2)$ factor instead of $\rho/(1-\rho)$, where the former is smaller than the latter when $\rho < (\sqrt{5}-1)/2$.

This difference in the dependence on ρ is caused by Markov's inequality, which is used here to bound the deviation probabilities between $\prod_s W_s$ and $\mathbf{1}\mathbf{1}^T/N$ in the gossiping analysis. Second, the dependence on N and p is worse. This is to be expected since the analysis involves high-probability upper and lower bounds on $|S_t|$, which unavoidably yield a dependence on N instead of p.

Proof. We have

$$\mathrm{Reg}_T^{\mathrm{net}} \leq \underbrace{3\sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\frac{\mathbb{I}\{u\in S_t\}}{|S_t|}L\left\|x_t(u)-\bar{x}_t\right\|}_{(A)} + \underbrace{\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\frac{\mathbb{I}\{v\in S_t\}}{|S_t|}\left\langle\nabla\ell_t(x_t(v),v),\bar{x}_t-x^*\right\rangle}_{(B)},$$

thanks to Lemma C.2. Let us start by bounding Term (B).

$$\begin{aligned}\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\frac{\mathbb{I}\{v\in S_t\}}{|S_t|}\left\langle\nabla\ell_t(x_t(v),v),\bar{x}_t-x^*\right\rangle &\leq N\left(\frac{1}{N}\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\mathbb{I}\{v\in S_t\}\left\langle\nabla\ell_t(x_t(v),v),\bar{x}_t-x^*\right\rangle\Pr(S_t\neq\emptyset)\right)\\ &\leq N\left(\frac{\psi(x^*)}{\eta}+\frac{L^2}{\mu}\sum_{t=1}^{T}\eta\right) \qquad \text{(C.17)}\end{aligned}$$

We then proceed by bounding Term (A), by observing

$$3\sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\frac{\mathbb{I}\{u\in S_t\}}{|S_t|}L\left\|x_t(u)-\bar{x}_t\right\| \leq 3\eta\sum_{t=1}^{T}\max_{u\in\mathcal{V}}L\left\|z_t(u)-\bar{z}_t\right\|$$

which is derived thanks to Equation (C.14).

Now, we focus on $\|z_t(v)-\bar{z}_t\|_*$. Starting from Equation (C.15) and applying Lemma C.4, we obtain

$$\begin{aligned}\left\|z_{t+1}(v)-\bar{z}_{t+1}\right\|_* &\leq \sum_{s=1}^{t-1}\sqrt{N}L\left\|W_t\cdots W_{s+1}e_v-\frac{1}{N}\mathbf{1}\right\|_2 + 2L\\ &\leq \sqrt{N}L(t-t^*)\epsilon + \sqrt{N}Lt^* + 2NL\\ &\leq \sqrt{N}LT\epsilon + \sqrt{N}Lt^* + 2NL \qquad \text{(C.18)}\end{aligned}$$

with probability at least $1-\epsilon T,\ \forall t \geq 1$. For $t=1$, we have $\|z_1(v)-\bar{z}_1\|=0$ Hence,

$$\mathrm{Reg}_T^{\mathrm{net}} \leq 3\eta\sqrt{N}\frac{L^2}{\mu}T(t^*+\epsilon T+2)+N\left(\frac{D^2}{\eta}+\frac{L^2}{\mu}\sum_{t=1}^{T}\eta\right) \leq N\frac{D^2}{\eta}+\eta N\frac{L^2}{\mu}T(3t^*+\epsilon T+3)$$

with probability at least $1-\epsilon NT^2$. Setting $\epsilon=\frac{\delta}{NT^2}$ and $t^*=\frac{3\log\left(\frac{NT^2}{\delta}\right)}{1-\rho^2}$, we have

$$\mathrm{Reg}_T^{\mathrm{net}} \leq N\frac{D^2}{\eta}+3\eta TN\frac{L^2}{\mu}\left(\frac{3\log\left(\frac{NT^2}{\delta}\right)}{1-\rho^2}+3+\frac{\delta}{NT}\right) \leq N\frac{D^2}{\eta}+3\eta TN\frac{L^2}{\mu}\left(\frac{3\log\left(\frac{NT^2}{\delta}\right)}{1-\rho^2}+4\right) \tag{C.19}$$

with probability at least $1-\delta$. □

C.4.4 High-probability bound on the individual regret

A similar bound as in Theorem C.2 holds for the individual regret.

Theorem C.3. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta>0$. Then, with probability $1-\delta$, the network regret is bounded by*

$$\mathrm{Reg}_T(u) \leq N\left(\frac{D^2}{\eta}+\frac{L^2}{\mu}\eta T\right)+3\eta TN\frac{L^2}{\mu}\left(\frac{3}{1-\rho^2}\log\left(\frac{NT^2}{\delta}\right)+4\right)\ .$$

The same remarks hold as for the network regret. In particular, the dependence on N and p is worse than in Theorem 4.1. This is not surprising, as the analysis involves high-probability upper and lower bounds on $|S_t|$, which unavoidably yield a dependence on N instead of p.

Proof. In this proof we denote by $x^*=\arg\min_{x\in\mathcal{X}}\sum_{t\leq T;u\in S_t}\ell_t^{\mathrm{net}}(S_t,x)$ We still use Equation (C.2) with the same choice of $\bar{x}_t$ as in the proof of Theorem C.1. With this choice, we can bound Term (b) by Term (B). Hence, thanks to Equation (C.17), we have

$$(b) \leq N\left(\frac{\psi(x^*)}{\eta}+\frac{L^2}{\mu}\sum_{t=1}^{T}\eta\right)$$

We can also bound Term (c) by $\sum_{t\in[T]}L\|x_t(u)-\bar{x}_t\|$ and Term (a) by $2L\sum_{t\in[T]}\max_{v\in\mathcal{V}}\|x_t(v)-\bar{x}_t\|$. Thanks to Equation (C.14),

$$\|x_t(v)-\bar{x}_t\| \leq \eta\,\|z_t(v)-\bar{z}_t\|_*\,,\ \forall v\in\mathcal{V}.$$

We also have

$$\|z_{t+1}(v)-\bar{z}_{t+1}\|_* \leq \sqrt{N}LT\epsilon+\sqrt{N}Lt^*+2NL$$

with probability at least $1-\epsilon T,\ \forall t\geq 1$, for each $v\in\mathcal{V}$, thanks to Equation (C.18). For $t=1$, we have $\|z_1(v)-\bar{z}_1\|=0$. Eventually

$$(a)+(c) \leq 3\sum_{t=1}^{T}\left(\sqrt{N}LT\epsilon+\sqrt{N}Lt^*+2NL\right)$$

with probability at least $1 - \epsilon NT^2$. Setting $\epsilon = \frac{\delta}{NT^2}$ and $t^* = \frac{3\log(\frac{NT^2}{\delta})}{1-\rho^2}$, we have

$$\mathrm{Reg}_T(u) \le N\frac{D^2}{\eta} + 3\eta TN\frac{L^2}{\mu}\left(\frac{3\log(\frac{NT^2}{\delta})}{1-\rho^2} + 3 + \frac{\delta}{NT}\right) \le N\frac{D^2}{\eta} + 3\eta TN\frac{L^2}{\mu}\left(\frac{3\log(\frac{NT^2}{\delta})}{1-\rho^2} + 4\right) \tag{C.20}$$

with probability at least $1 - \delta$. □

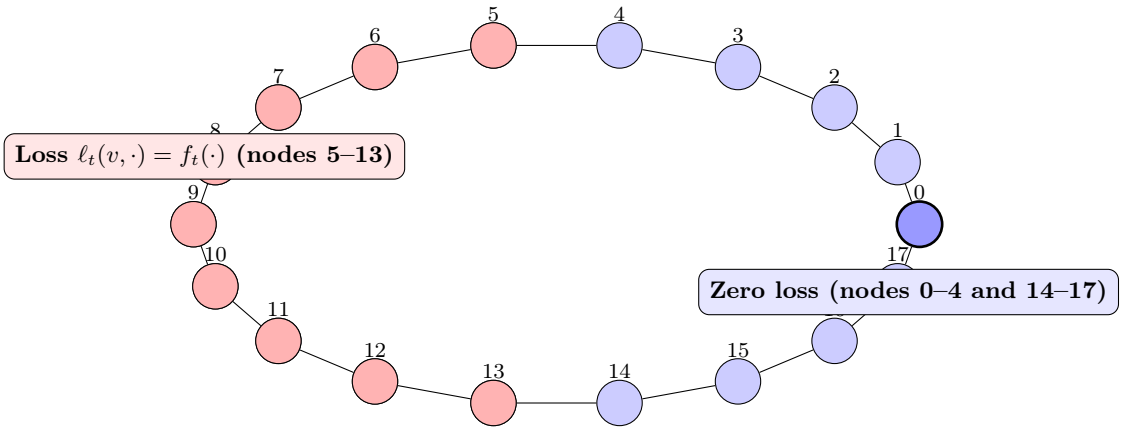
C.5 Omitted details in Section 4.5

In the following, we prove the lower bound presented in Section 4.5.

Theorem 4.2. *Let A be any algorithm for D-OCO on the decision set $\mathcal{X}$. Let $N = 2(M+1)$, where $M \ge 4$ is an even integer, and suppose $T \ge N^3$. Then, there exists a graph $\mathcal{G}$ with N nodes and a set of activation probabilities $\{p_v \mid v \in [N]\}$ with $p_{\min}N \ge 1$, and sequences of linear functions $\{\ell_t(v,\cdot)\}_{t=1}^T$, with each $\ell_t(v,\cdot)$ chosen adaptively based on $(S_k)_{k\le t}$, and satisfying $\|\nabla\ell_t(v,\cdot)\|_2 \le L$, such that the expected individual regret of A satisfies:*

$$\max_{u\in[N]} \mathbb{E}[\mathrm{Reg}_T(u)] \ge \frac{1}{2^7 p_{\min}} DL\,\kappa(\mathcal{G})^{\delta/2} N^{1/2-\delta}\sqrt{T}$$

for all $0 \le \delta \le \frac{1}{2}$, while the imbalance factor satisfies $I_p = \frac{2+p_{\min}(N-2)}{Np_{\min}} \le 3$.



Proof.

We let G denote a cycle graph with $N = 2(M+1)$ nodes where M is even, to simplify. Note that, on the N-cycle, the highest and smallest non-zero eigenvalues of the Laplacian are, respectively, $\lambda_1(\mathcal{G}) = 4$ and $\lambda_{N-1}(\mathcal{G}) = 2 - 2\cos(2\pi/N)$ (Spielman, 2019, Chapter 5.5). Using the inequality $1 - \cos(x) \ge x^2/5$, $\forall x \in [0,\pi]$ (recall that $N \ge 4$ implying $2\pi/N \le \pi$), we have $\lambda_{N-1}(\mathcal{G}) \ge \frac{8\pi^2}{5N^2}$ and so $\kappa(\mathcal{G}) \le \frac{20N^2}{8\pi^2} \le \frac{N^2}{2}$.

$$\kappa(\mathcal{G}) \le \frac{N^2}{2}. \tag{C.21}$$

We set $p_0 = p_{M+1} = 1$ and the remaining activation probabilities to $p_v = p, \forall v \notin \{0, M+1\}$.

Now suppose that for a subset of $M+1$ nodes around node 0, the local loss functions are identically zero at all times:

$$\ell_t(N - M/2,\cdot) = \cdots = f_t(M/2,\cdot) = 0 \quad \forall t \in [T].$$

The remaining nodes update their loss functions only after an entire path from their side of the graph to node 0 —either $(N - M/2 - 1, N - M/2, \dots, 0)$ or $(M/2 - 1, M/2 \dots 0)$— has been traversed since the last update. This implies that each edge along the path has been sequentially activated in the correct order, meaning both endpoints of each edge have been active.

Specifically, let d_k denote the time required, on the k-th occasion, to traverse an entire path from the half of the graph diametrically opposed to 0 to node 0. This time is given by $d_k = \min(X_{k,1}, X_{k,2})$ where each $X_{k,i} = M/2 + \widetilde{d}_k$ and $\widetilde{d}_k$ is the minimum of two independent negative binomial random variables with parameters $(M/2, 1/p^2)$.

Then, for each $k = 0, \dots$, the loss functions at nodes $\frac{M}{2} + 1$ through $\frac{3M}{2} + 1$ remain constant over the interval

$$\mathcal{I}_k \coloneqq \left[\min\left(\sum_{r=0}^{k} d_r, T \right), \dots, \min\left(\sum_{r=0}^{k+1} d_r, T \right) \right],$$

and are defined by

$$\ell_t\left(\frac{M}{2} + 1, \cdot\right) = \cdots = \ell_t\left(\frac{3M}{2} + 1, \cdot\right) = H_k(\cdot),$$

where $H_k(x) = \varepsilon_k L \langle w, x \rangle$, and ε_k are i.i.d. Rademacher random variables (i.e., taking values ± 1 with equal probability). The vector w is defined as

$$w = \frac{x_1 - x_2}{\|x_1 - x_2\|_2},$$

for some $x_1, x_2 \in \mathcal{X}$ such that $\|x_1 - x_2\|_2 = D$.

The global loss observed by agent 0 at time t, upon playing action x, is given by

$$\ell_t^{\text{net}}(x) = \frac{\sum_{v \in [M/2, 3M/2]} \mathbb{I}\{v \in S_t\}}{|S_t|} H_{k_t}(x),$$

where k_t denotes the index of the block such that $t \in \mathcal{I}_{k_t}$.

Due to the structure of the cycle, agent 0 cannot receive information on H_k until at least d_k time steps have passed. Thus, predictions $\{x_t(0) : \ t \in \mathcal{I}_k\}$ are made without access to H_k.

Defining $\mathcal{T}$ as the smallest integer k such that $\sum_{r=0}^{k} d_r \geq T$ and applying a variation of the standard lower bound from online learning (Orabona, 2025, Theorem 5.1), we obtain:

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} \ell_t^{\text{net}}\big(x_t(0)\big) - \min_{x \in \mathcal{X}} \sum_{t=1}^{T} \ell_t^{\text{net}}(x) \right] &\geq \mathbb{E}\left[\sum_{k=0}^{\mathcal{T}} \sum_{t \in \mathcal{I}_k} \frac{\sum_{v \in [M/2, 3M/2]} \mathbb{I}\{v \in S_t\}}{|S_t|} H_k(x_t(0)) \right. \\
&\qquad \left. - \min_{x \in \mathcal{X}} |\mathcal{I}_k| \sum_{k=0}^{\mathcal{T}} \frac{\sum_{v \in [M/2, 3M/2]} \mathbb{I}\{v \in S_t\}}{|S_t|} H_k(x) \right] \\
&\geq \mathbb{E}\left[- \min_{x \in \mathcal{X}} \sum_{k=0}^{\mathcal{T}} |\mathcal{I}_k| \frac{\sum_{v \in [M/2, 3M/2]} \mathbb{I}\{v \in S_t\}}{|S_t|} H_k(x) \right] \\
&\geq L\mathbb{E}\left[\max_{x \in \mathcal{X}} \sum_{k=0}^{\mathcal{T}-1} \frac{\sum_{v \in [M/2, 3M/2]} \mathbb{I}\{v \in S_t\}}{|S_t|} d_k \varepsilon_k \langle w, x \rangle \right] \\
&\geq L\mathbb{E}\left[\max_{x \in \{x_1, x_2\}} \sum_{k=0}^{\mathcal{T}-1} \frac{\sum_{v \in [M/2, 3M/2]} \mathbb{I}\{v \in S_t\}}{|S_t|} d_k \varepsilon_k \langle w, x \rangle \right]
\end{aligned}$$

$$\geq \frac{1}{4}LD\mathbb{E}\left[\left|\sum_{k=0}^{\mathcal{T}-1}\varepsilon_k d_k\right|\right]$$

(by independence and because $\mathbb{E}\left[\frac{\sum_{v\in[M/2,3M/2]}\mathbb{I}\left\{v\in S_t\right\}}{|S_t|}\right]\geq 1/4$)

$$= \frac{1}{4}LD\mathbb{E}_{d_1\ldots d_T}\left[\mathbb{E}_{\varepsilon_1\ldots\varepsilon_T}\left[\left|\sum_{k=0}^{\mathcal{T}-1}\varepsilon_k d_k\right|\right]\right]$$

$$\geq \frac{1}{4}LD\mathbb{E}_{d_1\ldots d_T}\left[\sqrt{\sum_{k=0}^{\mathcal{T}-1}d_k^2}\right] \quad \text{(Khintchine's inequality)}$$

$$\geq \frac{1}{4}LD\sqrt{\mathbb{E}\left[\sum_{k=0}^{\mathcal{T}-1}d_k^2\right]} \quad \text{(Jensen's inequality)} \tag{C.22}$$

By Wald's lemma, since $\mathcal{T}$ is a stopping time adapted to the sequence $d_1,\ldots,d_k$, we have:

$$,\mathbb{E}\left[\sum_{k=0}^{\mathcal{T}-1}d_k^2\right] = \mathbb{E}[\mathcal{T}-1]\cdot\mathbb{E}[d_1^2]. \tag{C.23}$$

To lower bound $\mathbb{E}[\mathcal{T}]$, let $\mu=\mathbb{E}[d_1]$. Note that:

$$\mathbb{P}\left(\sum_{k=1}^{\lfloor T/(2\mu)\rfloor}d_k\leq T\right) = 1-\mathbb{P}\left(\sum_{k=1}^{\lfloor T/(2\mu)\rfloor}d_k\geq T\right).$$

Using Markov's inequality,

$$\mathbb{P}\left(\sum_{k=1}^{\lfloor T/(2\mu)\rfloor}d_k\geq T\right)\leq\frac{\mathbb{E}[\sum d_k]}{T}\leq\frac{T/2}{T}=\frac{1}{2}.$$

Thus,

$$\mathbb{E}[\mathcal{T}]-1\geq\frac{T}{4\mu}-1. \tag{C.24}$$

Hence, using Equation (C.23)

$$\mathbb{E}\left[\sum_{k=0}^{\mathcal{T}}d_k^2\right]\geq\left(\frac{T}{4\mu}-1\right)\mathbb{E}[d_1^2]. \tag{C.25}$$

We now bound $\mathbb{E}[d_1]$ and $\mathbb{E}[d_1^2]$. Recall $d_1=\min(X_1,X_2)$ where

$$X_i=\frac{M}{2}+\widetilde{d}_i,\quad \widetilde{d}_i\sim\text{NegBin}\left(\frac{M}{2},\frac{1}{p^2}\right),$$

and $\widetilde{d}_1,\widetilde{d}_2$ are independent. Using order statistics, we have:

$$\mathbb{E}[X_1]-\sqrt{\text{Var}(\widetilde{d}_1)}\leq\mathbb{E}[d_1]\leq\mathbb{E}[X_1],$$

and

$$\mathbb{E}[d_1]^2 \leq \mathbb{E}[d_1^2].$$

Now,

$$\mathrm{Var}(\widetilde{d}_1) = \frac{M}{2} \cdot \frac{1-p^2}{p^4}, \quad \text{and} \quad \mathbb{E}[X_1] = \frac{M}{2} + \frac{M}{2} \cdot \frac{1-p^2}{p^2}.$$

Thus,

$$\frac{M}{2p^2} \leq \mathbb{E}[X_1] \leq \frac{M}{p^2}.$$

Then,

$$\mathbb{E}[d_1] \geq \mathbb{E}[X_1] - \sqrt{\mathrm{Var}(\widetilde{d}_1)} \geq \frac{M}{2} + \frac{M}{2} \cdot \frac{1-p^2}{p^2} - \frac{1}{p^2}\sqrt{\frac{M}{2}(1-p^2)}.$$

Using $\sqrt{M(1-p^2)} \leq \sqrt{M}$, we obtain:

$$\mathbb{E}[d_1] \geq \frac{M}{2p^2}\left(1 - \frac{1}{\sqrt{M}}\right) \geq \frac{M}{4p^2}. \tag{C.26}$$

Therefore,

$$\mathbb{E}[d_1]^2 \geq \frac{M^2}{16p^4}. \tag{C.27}$$

Also, since $\mu = \mathbb{E}[d_1] \leq \mathbb{E}[X_1] \leq \frac{M}{p^2}$, we have:

$$\frac{T}{4\mu} - 1 \geq \frac{Tp^2}{4M} - 1.$$

Assuming $T \geq N^3$, we get:

$$\frac{Tp^2}{4M} \geq 1 \quad \Rightarrow \quad \frac{T}{4\mu} - 1 \geq \frac{Tp^2}{8M}. \tag{C.28}$$

Combining Equation (C.28), Equation (C.27) and Equation (C.25), we get:

$$\mathbb{E}\left[\sum_{k=0}^{\mathcal{T}} d_k^2\right] \geq \frac{Tp^2}{8M} \cdot \frac{M^2}{16p^4} = \frac{MT}{2^7 p^2}.$$

Thus, thanks to Equation (C.22),

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_t\big(x_t(0)\big) - \min_{x\in\mathcal{X}} \sum_{t=1}^{T} \ell_t(x)\right] \geq \frac{1}{4}LD \cdot \frac{\sqrt{MT}}{16p}.$$

This implies there exists a realization of $\varepsilon_0, \ldots, \varepsilon_T$ for which:

$$\mathbb{E}[R_T(0)] \geq \frac{LD}{64p}\sqrt{MT}.$$

From the known inequality Equation (C.21):

$$\kappa(\mathcal{G}) \leq \frac{N^2}{2} \quad \Rightarrow \quad M \geq \frac{1}{4}N \geq \frac{\sqrt{2}}{4}\sqrt{\kappa(\mathcal{G})} \geq \frac{1}{2\sqrt{2}} \cdot \sqrt{\kappa(\mathcal{G})},$$

we obtain the final lower bound:

$$\mathbb{E}[R_T(0)] \geq \frac{1}{2^7 p} \cdot LD \cdot \sqrt{T} \cdot \kappa(\mathcal{G})^{\delta/2} N^{1/2-\delta}, \quad \forall\ 0 \leq \delta \leq \frac{1}{2}.$$

□

C.6 Omitted details in Section 4.6

We start by proving the general bounds on ρ of Theorem 4.3 and Corollary 4.1.

Theorem 4.3. *If W_t is set according to Equation* (4.8)*, then*

$$\rho^2 \leq 1 - b p_{\min}^2 \lambda_{N-1}(\mathcal{G}) \ . \tag{4.9}$$

Moreover, for $b = 1/\lambda_1(\mathcal{G})$ we have

$$\rho^2 \leq 1 - \frac{p_{\min}^2}{\kappa(\mathcal{G})} \ . \tag{4.10}$$

Proof. Recall $\rho^2 = \lambda_2(\mathbb{E}[W_1 W_1^\top])$. We have

$$\begin{aligned}
\lambda_2(\mathbb{E}[W_1 W_1^\top]) &\leq \lambda_2(\mathbb{E}[W_1]) && \text{(C.29)}\\
&\leq \lambda_2\Big(I - b\,\mathbb{E}\big[\mathrm{Lap}(\mathcal{G}_1)\big]\Big) \\
&\leq \lambda_2\big(I - b\,P\mathrm{Lap}(\mathcal{G})P\big) \\
&\leq \lambda_2\big(I - b\,p_{\min}^2\mathrm{Lap}(\mathcal{G})\big) && \text{(C.30)}\\
&\leq 1 - b\,p_{\min}^2\lambda_{N-1}(\mathcal{G}) \ , && \text{(C.31)}
\end{aligned}$$

where P is the diagonal matrix such that $P(v,v) = p_v$. Equation (C.29) holds because W_1 is symmetric and $W_1^2 \preceq W_1$. $P\mathrm{Lap}(\mathcal{G})P$ is also symmetric and, clearly, $P\mathrm{Lap}(\mathcal{G})P \succeq p_{\min}^2\mathrm{Lap}(\mathcal{G})$, implying Equation (C.30). Finally, Equation (C.31) holds because $\lambda_{N-1}(\mathcal{G})$ is the smallest non-zero eigenvalue of $\mathrm{Lap}(G)$. □

The following corollary follows by some easy computations.

Corollary 4.1. *If W_t is set according to Equation* (4.8) *and $b = 1/\lambda_1(\mathcal{G})$, then*

$$\frac{1}{1-\rho} \leq 2\frac{\kappa(\mathcal{G})}{p_{\min}^2} \, .$$

Proof. We know

$$\frac{1}{1-\rho} \leq \frac{1}{1-\sqrt{1-\frac{p_{\min}^2}{\kappa}}} \, ,$$

thanks to Theorem 4.3. The rest of the proof consists of proving that

$$\frac{1}{1-\sqrt{1-\frac{p_{\min}^2}{\kappa}}} \leq 2\frac{\kappa}{p_{\min}^2}.$$

In general, we prove that $\frac{1}{1-\sqrt{1-x}} \leq 2 \cdot \frac{1}{x}, \forall x \leq 1$.By concavity, we have: $\sqrt{1-x} \leq 1 - \frac{1}{2}x$. Thus

$$\frac{1}{1-\sqrt{1-x}} \leq \frac{1}{1-\left(1-\frac{1}{2}x\right)} = \frac{1}{\frac{1}{2}x} = \frac{2}{x}.$$

Hence

$$\frac{1}{1-\rho} \leq 2\frac{\kappa}{p_{\min}^2}\,.$$

□

C.6.1 Proof of Corollary 4.2

Corollary 4.2. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta > 0$. If the gossip matrix W_t is chosen as in Equation (4.8) with $b = 1/\lambda_1(\mathcal{G})$ and η tuned with respect to $\{p_v\}_{v\in\mathcal{V}}$ and N, the expected individual regret can be bounded by*

$$\max_{u\in\mathcal{V}} \mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(DLI_p \frac{\kappa(\mathcal{G})}{p_{\min}\bar{p}^{1/4}} N^{1/4}\sqrt{T/\mu}\right)\,, \tag{4.11}$$

in the general case and

$$\max_{u\in\mathcal{V}} \mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(DL \frac{\kappa(\mathcal{G})}{p^{3/4}} N^{1/4}\sqrt{T/\mu}\right) \tag{4.12}$$

in the p-uniform case, for all $p \leq 1$.

Proof. Thanks to Theorem 4.1 , we have

$$\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq \frac{ND^2}{\eta} + \frac{L^2}{\mu}\eta\left(\bar{p}(N+1) + 6 + 3\sqrt{\bar{p}N}\frac{\rho}{1-\rho}\right)T\,.$$

We note that $\bar{p}(N+1) \leq \bar{2}pN$ and

$$\frac{1}{1-\rho} \leq 2\frac{\kappa}{p_{\min}^2}\,,$$

thanks to Corollary 4.1.

Consequently, we can prove

$$\begin{aligned}
\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] &\leq \frac{D^2 I_p}{\eta\bar{p}} + \eta\left(6 + 2I_p + 8\sqrt{\bar{p}N}\frac{\rho}{1-\rho}\right)T \\
&\leq I_p\left(\frac{D^2}{\eta\bar{p}} + \eta\left(8 + 8\sqrt{\bar{p}N}\frac{\rho}{1-\rho}\right)\right)T \\
&\leq I_p\left(\frac{D^2}{\eta\bar{p}} + \eta\left(8\sqrt{\bar{p}N} + 8\sqrt{\bar{p}N}\frac{\rho}{1-\rho}\right)\right)T \\
&\leq I_p\left(\frac{D^2}{\eta\bar{p}} + \eta\left(16\sqrt{\bar{p}N}\frac{\kappa}{p_{\min}^2}\right)\right)T\,.
\end{aligned}$$

Setting $\eta = \frac{4D\sqrt{\mu}p_{\min}}{L\sqrt{T}\bar{p}^{3/4}N^{1/4}}$, which only requires knowing $p_{\min}$, $\bar{p}$ suffices to obtain the bound

$$\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{net}}\big] \leq 4DLI_p \frac{\kappa(\mathcal{G})}{p_{\min}\bar{p}^{1/4}} N^{1/4}\sqrt{\frac{T}{\mu}}\,. \tag{C.32}$$

The proof for the p-uniform case proceeds similarly □

Now, we are interested in results holding in the p-uniform case. We first prove the following equality that holds for arbitrary graphs, under the p-uniform assumption.

Theorem 4.4. *If W_t is set according to Equation (4.8), and $b = 1/\lambda_1(\mathcal{G})$, then in the p-uniform case we have*

$$\rho^2 = 1 - \frac{2p^2}{\kappa(\mathcal{G})}\left(1 - \frac{1-p}{\lambda_1(\mathcal{G})} - \frac{p}{2\kappa(\mathcal{G})}\right)\,.$$

Proof. Denoting by $L_1 = \mathrm{Lap}(\mathcal{G}_1)$, one has:

$$\mathbb{E}[W_1^2] = \mathbb{E}[I - 2bL_1 + b^2L_1^2] = I - 2b\mathbb{E}[L_1] + b^2\mathbb{E}(L_1^2)$$

$$\mathbb{E}[L_1] = p^2D_{\mathcal{G}} - p^2A_{\mathcal{G}} = p^2L_{\mathcal{G}}$$

We compute:

$$\mathbb{E}[L_1^2] = \mathbb{E}[D_1^2 - 2D_1A_1 + A_1^2]$$

where D_1 and A_1 denote the diagonal matrix of degrees and the adjacency matrix of G_1.

$$\begin{aligned}
\mathbb{E}[(D_1^2)_{ii}] &= \mathbb{E}\left[\left(\sum_{j\in\mathcal{N}_i}\mathbb{1}(j\in S_1)\right)^2\mathbb{1}(i\in S_1)\right] \\
&= p\mathbb{E}\left[\sum_{j\in\mathcal{N}_i}\mathbb{1}(j\in S_1)\sum_{k\in\mathcal{N}_i}\mathbb{1}(k\in S_1)\right] \\
&= p\mathbb{E}\left[\sum_{j\in\mathcal{N}_i}\mathbb{1}(j\in S_1)(\sum_{k\in\mathcal{N}_i,k\neq j}\mathbb{1}(k\in S_1)+\mathbb{1}(j\in S_1))\right] \\
&= p\left(p^2D_{\mathcal{G}}(D_{\mathcal{G}}-I) + D_{\mathcal{G}}p\right)_i\,,
\end{aligned}$$

where $D_{\mathcal{G}}$ denotes the matrix of degrees of the graph $\mathcal{G}$.

Finally,

$$\mathbb{E}[(D_1^2)] = p^2(pD_{\mathcal{G}}(D_{\mathcal{G}}-I) + D_{\mathcal{G}})\,.$$

where $A_{\mathcal{G}}$ denotes the adjacency matrix of the graph $\mathcal{G}$. Regarding A_1^2

$$\begin{aligned}
\mathbb{E}[(A_1^2)_{ij}] &= \mathbb{E}\left[\sum_{k\in\mathcal{N}_i,k\in\mathcal{N}_j}\mathbb{1}(k\in S_1)\mathbb{1}(i\in S_1)\mathbb{1}(j\in S_1)\right] \\
&= \begin{cases} p^3\left|\mathcal{N}_i\cap\mathcal{N}_j\right| = p^3(A_{\mathcal{G}}^2)_{ij} & \text{if } i\neq j \\ \mathbb{E}\left[\sum_{k\in\mathcal{N}_i}\mathbb{1}(k\in S_1)\mathbb{1}(i\in S_1)\right] = p^2(D_{\mathcal{G}})_i & \text{if } i=j \end{cases}
\end{aligned}$$

Hence, finally,

$$\mathbb{E}[A_1^2] = p^3 A_{\mathcal{G}}^2 - p^3 D_{\mathcal{G}} + p^2 D_{\mathcal{G}} \,.$$

Now, regarding $D_1 A_1$,

$$\begin{aligned}
\mathbb{E}[(D_1 A_1)_{ij}] &= \mathbb{E}\left[(D_1)_{ii}(A_1)_{ij}\right] \\
&= \mathbb{E}\left[\mathbb{1}(i \in S_1)\mathbb{1}(j \in S_1)\mathbb{1}((i,j) \in \mathcal{E}) \sum_{k \in \mathcal{N}_i} \mathbb{1}(k \in S_1)\right] \\
&= \begin{cases} p^3 (D_{\mathcal{G}} A_{\mathcal{G}})_{ij} - p^3 [A_{\mathcal{G}}]_{ij} + p^2 (A_{\mathcal{G}})_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}
\end{aligned}$$

Finally,

$$\mathbb{E}[D_1 A_1] = p^3 (D_{\mathcal{G}} - I) A_{\mathcal{G}} + p^2 A_{\mathcal{G}} \,.$$

Putting everything together,

$$\begin{aligned}
\mathbb{E}[L_1^2] &= p^2 (p D_{\mathcal{G}} (D_{\mathcal{G}} - I) + D_{\mathcal{G}}) + p^3 A_{\mathcal{G}}^2 - p^3 D_{\mathcal{G}} + p^2 D_{\mathcal{G}} - 2p^3 (D_{\mathcal{G}} - I) A_{\mathcal{G}} - 2p^2 A_{\mathcal{G}} \\
&= p^3 D_{\mathcal{G}}^2 - 2p^3 D_{\mathcal{G}} + 2p^2 D_{\mathcal{G}} - 2p^3 D_{\mathcal{G}} A_{\mathcal{G}} + 2p^3 A_{\mathcal{G}} - 2p^2 A_{\mathcal{G}} + p^3 A_{\mathcal{G}}^2 \,.
\end{aligned}$$

So in the end we have

$$\begin{aligned}
\mathbb{E}[W_1^2] &= I - 2bp^2 (D_{\mathcal{G}} - A_{\mathcal{G}}) + b^2 \left(p^3 (D_{\mathcal{G}}^2 - 2D_{\mathcal{G}} - 2D_{\mathcal{G}} A_{\mathcal{G}} + 2A_{\mathcal{G}}) + p^3 A_{\mathcal{G}}^2 + p^2 (2D_{\mathcal{G}} - 2A_{\mathcal{G}})\right) \\
&= I - 2bp^2 (D_{\mathcal{G}} - A_{\mathcal{G}}) + b^2 [p^3 ((D_{\mathcal{G}} - A_{\mathcal{G}})^2 + 2(A_{\mathcal{G}} - D_{\mathcal{G}})) + 2p^2 (D_{\mathcal{G}} - A_{\mathcal{G}}) \,.]
\end{aligned}$$

$$\mathbb{E}[W_1^2] = I - 2bp^2 L_{\mathcal{G}} + b^2 [p^3 (L_{\mathcal{G}}^2 - 2L_{\mathcal{G}}) + 2p^2 L_{\mathcal{G}}] \,. \tag{C.33}$$

So if x is an eigenvector of $L_{\mathcal{G}}$ with eigenvalue $\lambda_i(L_{\mathcal{G}})$, then

$$\begin{aligned}
\mathbb{E}[W_1^2]x &= x - 2bp^2 \lambda_i(L_{\mathcal{G}})x + b^2 [p^3 (\lambda_i(L_{\mathcal{G}})^2 - 2\lambda_i(L_{\mathcal{G}})) + 2p^2 \lambda_i(L_{\mathcal{G}})]x \\
&= \left(1 - 2bp^2 \lambda_i(L_{\mathcal{G}}) + b^2 [p^3 (\lambda_i(L_{\mathcal{G}})^2 - 2\lambda_i(L_{\mathcal{G}})) + 2p^2 \lambda_i(L_{\mathcal{G}})]\right) x \\
&= \left(1 + (-2bp^2 + 2b^2 p^2 - 2p^3 b^2) \lambda_i(L_{\mathcal{G}}) + b^2 p^3 (\lambda_i(L_{\mathcal{G}})^2\right) x
\end{aligned}$$

So that the eigenvalues of $\mathbb{E}[W_1^2]$ can easily be expressed as eigenvalues of $L_{\mathcal{G}}$. It is easy to check that $f_{b,p} : x \mapsto 1 + (-2bp^2 + 2b^2 p^2 - 2p^3 b^2)x + b^2 p^3 x^2$ is a quadratic function decreasing on $]-\infty, 1 + (1-b)/(bp)]$.

It is clear that

$$1 + (1-b)/(bp) = 1 + (1/b - 1)/p \geq 1/bp \geq 1/b \geq 2\Delta(\mathcal{G}) \geq \lambda_1(L_{\mathcal{G}}) \,.$$

So that $f_{b,p}$ is decreasing on an interval containing all eigenvalues of the Laplacian $L_{\mathcal{G}}$ and that

$$\lambda_2(\mathbb{E}[W_1^2]) = f_{b,p}(\lambda_{N-1}(L_{\mathcal{G}})) \,.$$

Hence

$$\lambda_2(\mathbb{E}[W_1^2]) = 1 + (-2bp^2 + 2b^2 p^2 - 2p^3 b^2) \lambda_{N-1}(L_{\mathcal{G}}) + b^2 p^3 (\lambda_{N-1}(L_{\mathcal{G}}))^2 \,. \tag{C.34}$$

Setting $b = 1/\lambda_1(\mathcal{G})$ and rewriting yields

$$\rho^2 = \lambda_2(\mathbb{E}[W_1^2]) = 1 - \frac{2p^2}{\kappa(\mathcal{G})}\left(1 - \frac{1-p}{\lambda_1(\mathcal{G})} - \frac{p}{2\kappa(\mathcal{G})}\right) .$$

□

C.6.2 Special cases

In the following, we prove inequalities on ρ in the p uniform case, when W_t is set as in Equation (4.8), in some specific network configurations.

Clique. When $\mathcal{G}$ is the clique, and $b = 1/(N)$, since $\lambda_{N-1}(\mathcal{G}) = N$

$$\lambda_2(\mathbb{E}[W_1^2]) = 1 - p^2 - \frac{1}{N}((N-2)p^3 + 2p^2)\,,$$

using Theorem 4.4.

Strongly Regular graphs. Strongly Regular graphs are such that

- they are k-regular, for some integer k
- there exists an integer m such that for every pair of vertices u and v that are neighbors in $\mathcal{G}$, there are m vertices that are neighbors of both u and v
- there exists an integer n such that for every pair of vertices u and v that are not neighbors in $\mathcal{G}$, there are n vertices that are neighbors of both u and v.

Such graphs' adjacency matrices have eigenvalues k with multiplicity 1 and r and s defined as follows:
$r = \frac{m-n+\sqrt{(m-n)^2+4(k-n)}}{2}$ and
$s = \frac{m-n-\sqrt{(m-n)^2+4(k-n)}}{2}$. (Spielman, 2019, Chapter 9.6) Hence, their Laplacian have eigenvalues 0 with multiplicity 1 and and $k-r$ and $k-s$. This yields $\lambda_1(\mathrm{Lap}(\mathcal{G})) = k-s$ and $\lambda_{N-1}(\mathcal{G}) = k-r$.

Using Theorem 4.3, we have that if $b = 1/\lambda_1(\mathrm{Lap}(\mathcal{G}))$, then

$$\rho^2 = \lambda_2(\mathbb{E}[W_1^2]) \leq 1 - p^2\frac{k-r}{k-s}$$

In particular, when $\mathcal{G}$ is the lattice graph, with $N = M^2$ vertices

$$\rho^2 = \lambda_2(\mathbb{E}[W_1^2]) \leq 1 - \frac{1}{2}p^2\,.$$

replacing k by $2M-2$, m by $M-2$ and n by 2.

Grid. Consider $\mathcal{G}$ a grid of dimension 2, with $N = M^2$. $\mathcal{G}$ is the product of two paths graphs of length M. Then if $\mu_1 \dots \mu_M$ are the eigenvalues of the path graph of length M, then all eigenvalues of $\mathrm{Lap}(G)$ can be rewritten as $\mu_i + \mu_j$ for some i and j—see (Barik et al., 2015, Theorem 3). Furthermore we know that $\mu_i = 2(1-\cos(\pi(M-i)/M))$—see, e.g., (Spielman, 2019, Theorem 6.6)—so that $\lambda_1(\mathrm{Lap}(\mathcal{G})) = 2\mu_1 = 4 - 4\cos(\pi(M-1)/M$ and $\lambda_{N-1}(\mathcal{G}) = \mu_N + \mu_{N-1} = 2 - 2\cos(\pi/M)$.

Hence setting $b = 1/\lambda_1(\mathrm{Lap}(\mathcal{G}))$

$$\rho^2 = \lambda_2(\mathbb{E}[W_1^2]) \leq 1 - p^2 \frac{2 - 2\cos(\pi/M)}{4 - 4\cos(\pi(M-1)/M)}$$

by using Theorem 4.3.

C.7 Omitted details in Section 4.7

Corollary 4.3. *Assume each agent runs an instance of* `Gossip-FTRL` *with learning rate $\eta > 0$. If the gossip matrix W_t is chosen as in Equation (4.8) with $b = 1/\lambda_1(\mathcal{G})$, then*

$$\rho^2 = 1 - \frac{2p^2 q}{\kappa(\mathcal{G})}\left(1 - \frac{1 - pq}{\lambda_1(\mathcal{G})} - \frac{pq}{2\kappa(\mathcal{G})}\right) .$$

By tuning η with respect to p and N,the expected individual regret of each $u \in \mathcal{V}$ on $\mathcal{G}_1, \mathcal{G}_2, \ldots$ drawn i.i.d. from $\mathcal{G}(\mathcal{G}, p, q)$ can be bounded by

$$\mathbb{E}\big[\mathrm{Reg}_T(u)\big] = \mathcal{O}\left(\frac{\kappa(\mathcal{G})}{q}\frac{N^{1/4}}{p^{3/4}}\sqrt{T}\right) . \tag{4.13}$$

Proof. As bound Equation (4.4) in Theorem 4.1 applies, we just have to compute the spectral gap.

Regarding the expression of $\rho^2 = \lambda_2(\mathbb{E}[W_1^2])$, we take the same steps as for the proof of Theorem 4.4.

We observe that

$$\mathbb{E}[W_1^2] = \mathbb{E}[I - 2bL_1 + b^2 L_1^2] = I - 2b\mathbb{E}[L_1] + b^2\mathbb{E}(L_1^2) .$$

We compute

$$\mathbb{E}[L_1] = p^2 D_{\mathcal{G}} - p^2 A_{\mathcal{G}} = p^2 L_{\mathcal{G}}$$

and

$$\mathbb{E}[L_1^2] = \mathbb{E}[D_1^2 - 2D_1 A_1 + A_1^2] .$$

Mutatis mutandis in the computations of the proof of Theorem 4.4, we get the following equalities for the first term,

$$\begin{aligned}
\mathbb{E}[(D_1^2)_{ii}] &= \mathbb{E}\left[\left(\sum_{j\in\mathcal{N}_i} \mathbb{1}((i,j)\in\mathcal{E}_1)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{j\in\mathcal{N}_i} \mathbb{1}((i,j)\in\mathcal{E}_1)\Big(\sum_{k\in\mathcal{N}_i, k\neq j} \mathbb{1}((i,k)\in\mathcal{E}_1) + \mathbb{1}((i,j)\in\mathcal{E}_1)\Big)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{j\in\mathcal{N}_i} \mathbb{1}((i,j)\in\mathcal{E}_1)\Big(\sum_{k\in\mathcal{N}_i, k\neq j} \mathbb{E}\left[\mathbb{1}((i,k)\in\mathcal{E}_1) + \mathbb{1}((i,j)\in\mathcal{E}_1)\Big|(i,k)\in S_1\right]\Big)\Big|(i,j)\in S_1\right]\right] \\
&= p^2 q\left(pqD_{\mathcal{G}}(D_{\mathcal{G}} - I) + D_{\mathcal{G}}\right)_i ,
\end{aligned}$$

the second term,

$$\mathbb{E}[(A_1^2)_{ij}] = \mathbb{E}\left[\sum_{k\in\mathcal{N}_i, k\in\mathcal{N}_j} \mathbb{1}((k,i)\in\mathcal{E}_1)\mathbb{1}((k,j)\in\mathcal{E}_1)\right]$$
$$= \begin{cases} p^3\,|\mathcal{N}_i\cap\mathcal{N}_j| = p^3q^2(A_{\mathcal{G}}^2)_{ij} & \text{if } i\neq j \\ p^2\sum_{k\in\mathcal{N}_i}\mathbb{E}\left[\mathbb{1}((i,k)\in\mathcal{E}_1)\middle|\mathbb{1}((i,k)\in S_1)\right] = p^2q(D_{\mathcal{G}})_i & \text{if } i=j \end{cases}$$

and the third term.

$$\mathbb{E}[(D_1A_1)_{ij}] = \mathbb{E}\left[(D_1)_{ii}(A_1)_{ij}\right]$$
$$= \mathbb{E}\left[\mathbb{1}((i,j)\in\mathcal{E}_t)\sum_{k\in\mathcal{N}_i}\mathbb{1}((i,k)\in\mathcal{E}_1)\right]$$
$$= \begin{cases} p^3q^2(D_{\mathcal{G}}A_{\mathcal{G}})_{ij} - p^3q^2[A_{\mathcal{G}}]_{ij} + p^2q(A_{\mathcal{G}})_{ij} & \text{if } i\neq j \\ 0 & \text{if } i=j \end{cases}$$

Finally, by adding these three inequalities and rearranging,

$$\mathbb{E}[W_1^2] = I - 2bp^2qL_{\mathcal{G}} + b^2\left(p^3q^2(L_{\mathcal{G}}^2 - 2L_{\mathcal{G}}) + 2p^2qL_{\mathcal{G}}\right). \tag{C.35}$$

It is easy to check that $f_{b,p,q} : x \mapsto 1 + (-2bp^2q + 2b^2p^2q - 2p^3b^2q^2)x + b^2p^3q^2x^2$ is a quadratic function decreasing on $(-\infty, 1 + (1-b)/(bpq)]$.

Again,

$$1 + (1-b)/(bpq) = 1 + (1/b - 1)/pq \geq 1/bpq \geq 1/b \geq 2\Delta(\mathcal{G}) \geq \lambda_1(L_{\mathcal{G}})\,.$$

so that $f_{b,p,q}$ is decreasing on an interval containing all eigenvalues of the Laplacian $L_{\mathcal{G}}$ and that

$$\rho^2 = \lambda_2(\mathbb{E}[W_1^2]) = f_{b,p,q}(\lambda_{N-1}(L_{\mathcal{G}})) = 1 - \frac{2p^2q}{\kappa(\mathcal{G})}\left(1 - \frac{1-pq}{\lambda_1(\mathcal{G})} - \frac{pq}{2\kappa(\mathcal{G})}\right)$$

by a s simple rewriting, concluding the proof. □

Appendix D

Proof Details for Chapter 5

D.1 Preliminary Results

In this section, we show several auxiliary lemmas that will be helpful.

D.1.1 General properties of FTRL

The following FTRL stability lemma bounds the distance between two FTRL iterates with different linear losses and possibly different regularizers. It also shows a simplified upper bound in the case when two decisions are made by using FTRL with the same regularizer.

Lemma D.1 (Lemma A.2 of Qiu et al. (2025a)). *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed and non-empty. Let $A_1, A_2 \succeq 0$ be two positive semidefinite matrices, $b_1, b_2 \in \mathbb{R}^n$, and $c_1, c_2 \in \mathbb{R}$. Define $\psi_1(x) = x^\top A_1 x + b_1^\top x + c_1$ and $\psi_2(x) = x^\top A_2 x + b_2^\top x + c_2$. Suppose that $z_1 \in \arg\min_{x\in\mathcal{X}} \{\langle w_1, x\rangle + \psi_1(x)\}$ and $z_2 \in \arg\min_{x\in\mathcal{X}} \{\langle w_2, x\rangle + \psi_2(x)\}$. Then, we have*

$$\|z_1 - z_2\|_{A_1}^2 + \|z_1 - z_2\|_{A_2}^2 \leq \langle w_1 - w_2, z_2 - z_1\rangle + (\psi_1(z_2) - \psi_2(z_2)) - (\psi_1(z_1) - \psi_2(z_1)) \ .$$

Furthermore, if $\psi_1(x) = \psi_2(x) = x^\top A x + b^\top x + c$ with positive definite $A \succ 0$, we have

$$\|z_1 - z_2\|_A \leq \frac{1}{2}\|w_1 - w_2\|_{A^{-1}},$$

where $\|x\|_A = \sqrt{x^\top A x}$ denotes the Mahalanobis norm induced by a positive semi-definite matrix A.

D.1.2 Basic analysis facts

Lemma D.2 (Lemma 4.13 in Orabona (2025)). *Let $a_0 \geq 0$ and let $f : [0, +\infty) \to [0, +\infty)$ be a non-increasing function. Then*

$$\sum_{t=1}^{T} a_t f\left(a_0 + \sum_{i=1}^{t} a_i\right) \leq \int_{a_0}^{\sum_{t=0}^{T} a_t} f(x)\mathrm{d}x.$$

D.1.3 Facts on the delay

The following lemma illustrates the relationship between the cumulative number of missing observations at the end of each block and total delay, which will be useful in later analysis.

Lemma D.3. *For any $u \in \mathcal{V}$ and any fixed integer $B > 0$ with T/B an integer,*

$$B \sum_{s=1}^{T/B} |m_{sB+1}(u)| \leq \sum_{s=1}^{T} d_s(u) + BT.$$

Consequently, we also have for all $s \in [T/B]$.

$$BM_s \leq \frac{1}{N} \sum_{s=1}^{T} \sum_{v \in \mathcal{V}} d_s(u) + BT = d_{\text{tot}} + BT,$$

where $M_s \triangleq \frac{1}{N} \sum_{u \in \mathcal{V}} |m_{sB+1}(u)|$ for all $s \in [T/B]$.

Proof. Note that each gradient $g_t(u)$ that is delayed by $d_t(u)$ remains unobserved for $d_t(u)$ rounds, and therefore contributes to $|m_{kB+1}(u)|$ for exactly $\lceil d_t(u)/B \rceil$ consecutive blocks. Summing over all $t \in [T]$, we obtain that

$$B \sum_{k=1}^{T/B} |m_{kB+1}(u)| = B \sum_{t=1}^{T} \left\lceil \frac{d_t(u)}{B} \right\rceil \leq \sum_{t=1}^{T} (d_t(u) + B) = \sum_{t=1}^{T} d_t(u) + BT.$$

This proves the first inequality. To obtain the bound on M_s, since $M_s = \frac{1}{N} \sum_{u \in \mathcal{V}} |m_{sB+1}(u)|$, summing both sides over $s = 1, \ldots, T/B$ and applying the bound above lead to

$$B \sum_{s=1}^{T/B} M_s = \frac{B}{N} \sum_{s=1}^{T/B} \sum_{u \in \mathcal{V}} |m_{sB+1}(u)| \leq \frac{1}{N} \sum_{u \in \mathcal{V}} \left(\sum_{t=1}^{T} d_t(u) + BT \right) = d_{\text{tot}} + BT.$$

□

D.2 Omitted Details in Section 5.3

D.2.1 Non-Adaptive Algorithm with Known Total Delay

In this section, we show the omitted details in Section "Non-Adaptive Algorithm with Known Total Delay". For completeness, we first restate the theorem and then present its proof. After establishing the main result, we proceed to prove several auxiliary lemmas that will be used in the algorithm's regret analysis.

Theorem 5.1. *Assume each agent $u \in \mathcal{V}$ runs an instance of Algorithm 5.1 with a valid communication matrix W, parameters θ and B defined in Equation* (5.6)*, and a fixed learning rate*

$$\eta_s(u) = \eta = \frac{D}{L\sqrt{d_{\text{tot}} + \sqrt{N}BT}}. \tag{5.7}$$

Then, under Assumption 5.1 and Assumption 5.2, the regret is bounded as

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

Furthermore, when $d_t(u) = d(u)$ for all $t \in [T]$, we have

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

with $\eta = \frac{D}{L\sqrt{d_{\text{tot}}+BT}}$.

Proof. We start the proof with some notations. We define

$$\bar{z}_{s-1} \triangleq \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}} y_l(v). \tag{D.1}$$

Direct calculation shows that $\bar{z}_{s-1}$ equals to the cumulative received gradients till block $s-1$ averaged over all agents:

$$\begin{aligned}\bar{z}_{s-1} &= \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{lB+1}(v)\setminus o_{(l-1)B+1}(v)} g_\tau(v) && \text{(Definition of } y_l(v)\text{)}\\ &= \frac{1}{N}\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{(s-1)B+1}(v)} g_\tau(v),\end{aligned}$$

where the last inequality is due to $o_1(v) = \emptyset$ for any $v \in \mathcal{V}$. Then for all $v \in \mathcal{V}$, define

$$\bar{x}_s(v) \triangleq \arg\min_{x\in\mathcal{X}} \langle \bar{z}_{s-1}, x\rangle + \frac{1}{\eta_s(v)}\|x\|_2^2. \tag{D.2}$$

In this case, since $\eta_s(v) = \eta$ for all $s \in [T/B]$ and $v \in \mathcal{V}$, we have $\bar{x}_s(u) = \bar{x}_s(v)$ for all $u, v \in \mathcal{V}$ and

we let $\bar{x}_s$ denote this value. We also define

$$\widetilde{z}_{s-1} = \frac{1}{N}\sum_{l=1}^{s-1}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}} g_\tau(v)$$

to be the cumulative gradients till block $s-1$ averaged over all agents assuming no delay, where $\mathcal{T}_l = \{(l-1)B+1,\dots,lB\}$. We also define

$$F_s(x) \triangleq \langle \widetilde{z}_{s-1}, x\rangle + \frac{1}{\eta}\|x\|_2^2,$$

and let $\widetilde{x}_s \triangleq \arg\min_{x\in\mathcal{X}} F_s(x)$ be the minimizer of $F_s(x)$.

With all the above notations, we apply the regret decomposition proven in Lemma D.4 and obtain that:

$$\begin{aligned}
\mathrm{Reg}_T(u) &\le \underbrace{\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s(u) - x^*\rangle}_{\spadesuit} \\
&\quad + \underbrace{2BL\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}(\|\bar{x}_s(u)-\bar{x}_s(v)\|_2 + \|x_s(v)-\bar{x}_s(v)\|_2) + NBL\sum_{s=1}^{T/B}\|x_s(u)-\bar{x}_s(u)\|_2}_{\clubsuit} \\
&= \underbrace{\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s - x^*\rangle}_{\spadesuit} + \underbrace{2BL\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}\|x_s(v)-\bar{x}_s\|_2 + NBL\sum_{s=1}^{T/B}\|x_s(u)-\bar{x}_s\|_2}_{\clubsuit},
\end{aligned}$$

where the last equality uses the fact that $\bar{x}_s(u) = \bar{x}_s(v) = \bar{x}_s$ for all $u, v \in \mathcal{V}$.

We start analyzing Term ♠ by decomposing it as follows:

$$\begin{aligned}
\frac{1}{N}\spadesuit &= \frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s(u) - \widetilde{x}_s + \widetilde{x}_s - x^*\rangle \\
&= \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \widetilde{x}_s - x^*\rangle}_{\text{full-info}_T} + \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s(u) - \widetilde{x}_s\rangle}_{\text{drift}_T},
\end{aligned} \tag{D.3}$$

where full-info$_T$ corresponds to the regret assuming there is no delay and drift$_T$ corresponds to the regret induced by delayed feedback.

To analyze full-info$_T$, since

$$\widetilde{x}_s = \arg\min\{\langle \frac{1}{N}\sum_{v\in\mathcal{V}}\sum_{\tau\in\mathcal{T}_s} g_\tau(v), \cdot\rangle + \frac{\|x\|_2^2}{\eta}\},$$

invoking Assumption 5.1, Assumption 5.2, and applying Corollary 7.7 in Orabona (2025) yields the

following bound

$$\text{full-info}_T \leq \frac{D^2}{\eta} + \frac{\eta B L^2 T}{2}. \tag{D.4}$$

Now we turn to the analysis of drift_T in Term ♠. Specifically,

$$\begin{aligned}
\text{drift}_T &= \frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s - \widetilde{x}_s\rangle \\
&\leq BL\sum_{s=1}^{T/B}\|\bar{x}_s - \widetilde{x}_s\|_2 && \text{(Cauchy–Schwarz inequality and Assumption 5.2)} \\
&= BL\sum_{s=2}^{T/B}\|\bar{x}_s - \widetilde{x}_s\|_2 && (\bar{x}_1 = \widetilde{x}_1 = \mathbf{0}) \\
&\leq \frac{\eta BL}{2}\sum_{s=2}^{T/B}\|\bar{z}_{s-1} - \widetilde{z}_{s-1}\|_2 && \text{(according to Lemma D.1)} \\
&\leq \frac{\eta BL}{2}\sum_{s=2}^{T/B}\left\|\frac{1}{N}\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{(s-1)B+1}(v)} g_\tau(v) - \frac{1}{N}\sum_{l=1}^{s-1}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}} g_\tau(v)\right\|_2 \\
&&& \text{(by definition of } \bar{z}_{t-1} \text{ and } \widetilde{z}_{t-1}) \\
&= \frac{\eta BL}{2}\sum_{s=2}^{T/B}\left\|-\frac{1}{N}\sum_{v\in\mathcal{V}}\sum_{\tau\in m_{(s-1)B+1}(v)} g_\tau(v)\right\|_2 \\
&&& (\mathcal{T}_s = \{(s-1)B+1,\ldots,sB\} \text{ and } m_t(v) = [t-1]\backslash o_t(v)) \\
&\leq \frac{\eta BL^2}{2}\sum_{s=1}^{T/B}\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|\right) && \text{(D.5)}
\end{aligned}$$

where the last inequality is by the Assumption 5.2. Combining Equation (D.3), Equation (D.4), and Equation (D.5), we obtain

$$\frac{1}{N}\spadesuit \leq \frac{D^2}{\eta} + \frac{\eta BL^2}{2}\sum_{s=1}^{T/B}\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)| + B\right). \tag{D.6}$$

Now we start analyzing Term ♣. For notational convenience, we use $z_s(u)$ to denote $z_s^B(u)$ for all $u \in \mathcal{V}$. From Lemma D.6, we know that $\forall w \in V$ and $\forall s \in [1, T/B]$,

$$\|z_s(w) - \bar{z}_s\|_2 \leq \frac{2}{\sqrt{N}}\sum_{l=1}^{s-1} b^{(s-l-1)B}\left(\sqrt{\sum_{v\in\mathcal{V}}\|y_l(v)\|_2^2}\right). \tag{D.7}$$

Note that $x_1(u) = \bar{x}_1 = \mathbf{0}$. Combining Lemma D.1 with Equation (D.7), we derive the following bound on the cumulative deviation between $x_s(w)$ and $\bar{x}_s$ for any $w \in V$:

$$\sum_{s=1}^{T/B}\|x_s(w) - \bar{x}_s\|_2 = \sum_{s=2}^{T/B}\|x_s(w) - \bar{x}_s\|_2 \tag{D.8}$$

$$
\begin{aligned}
&\leq \sum_{s=1}^{T/B-1} \eta \|z_s(w) - \bar{z}_s\|_2 && \text{(according to Lemma D.1)} \\
&= \frac{2\eta}{\sqrt{N}} \sum_{s=1}^{T/B-1} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) && \text{(Equation (D.7))} \\
&= \frac{2\eta}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \cdot \sum_{s=l+1}^{T/B} b^{(s-l-1)B} \right) \\
& && \text{(swapping the order of summation)} \\
&\leq \frac{2\eta}{\sqrt{N}} \frac{1}{1-b^B} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) && \text{(D.9)} \\
&\leq \frac{2\eta}{\sqrt{N}} \frac{1}{1-\frac{1}{\sqrt{14N}}} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) \\
& && \text{(since } b^B \leq \tfrac{1}{\sqrt{14N}} \text{ shown in Equation (D.23))} \\
&\leq \frac{3\eta}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right), && \text{(D.10)}
\end{aligned}
$$

where the last inequality follows from $N \geq 1$. Furthermore, according to Lemma D.10, we have

$$
\sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) \leq NTL. \tag{D.11}
$$

Combining Equation (D.10) and Equation (D.11),

$$
\sum_{s=2}^{T/B} \|x_s(w) - \bar{x}_s\|_2 \leq 3\eta\sqrt{N}TL
$$

for all $w \in V$. Hence we obtain

$$
\clubsuit \leq 18B\eta N\sqrt{N}TL^2 \tag{D.12}
$$

according to the definition of $\clubsuit$. Furthermore, when $d_t(u) = d(u)$ for all $t \in [T]$, due to the definition of $y_l(v)$, each agent $v \in \mathcal{V}$ can receive at most B gradient in any block $s \in [T/B]$. Hence, we obtain

$$
\begin{aligned}
\sum_{s=2}^{T/B} \|x_s(v) - \bar{x}_s\|_2 &\leq \frac{3}{\sqrt{N}}\eta \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) && \text{(Equation (D.10))} \\
&\leq 3\eta TL.
\end{aligned}
$$

Hence, this gives us an improved upper bound of $\clubsuit$:

$$
\clubsuit \leq 18B\eta NTL^2. \tag{D.13}
$$

Finally, combining Equation (D.12) with Equation (D.6), Equation (D.12) and Lemma D.4, we can bound the overall regret as follows:

$$\begin{aligned}\mathrm{Reg}_T(u) &\le \frac{D^2N}{\eta} + \frac{\eta BL^2N}{2}\sum_{s=1}^{T/B}\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)| + B\right) + 18B\eta N\sqrt{N}TL^2 \\ &\le \frac{D^2N}{\eta} + \frac{L^2N}{2}\eta d_{\mathrm{tot}} + \frac{L^2NB}{2}\eta T + 18B\eta N\sqrt{N}TL^2 && \text{(Lemma D.3)} \\ &\le \frac{D^2N}{\eta} + \frac{L^2N}{2}\eta d_{\mathrm{tot}} + 19B\eta N\sqrt{N}TL^2. && \text{(D.14)}\end{aligned}$$

Picking η to be $\frac{D}{L\sqrt{d_{\mathrm{tot}}+\sqrt{N}BT}}$ leads to

$$\mathrm{Reg}_T(u) = \mathcal{O}\left(DLN\sqrt{d_{\mathrm{tot}} + \frac{\ln(N)\sqrt{N}}{\sqrt{1-\sigma_2(W)}}T}\right) = \widetilde{\mathcal{O}}\left(DLN\left(\sqrt{d_{\mathrm{tot}}} + \frac{N^{1/4}}{(1-\sigma_2(W))^{1/4}}\sqrt{T}\right)\right). \tag{D.15}$$

When $d_t(u) = d(u)$ for all $t \in [T]$ for all $u \in \mathcal{V}$, we combine Equation (D.13) with Equation (D.6), Equation (D.13) and Lemma D.4 we obtain

$$\mathrm{Reg}_T(u) \le \frac{D^2N}{\eta} + \frac{L^2N}{2}\eta d_{\mathrm{tot}} + 19BL^2N\eta T.$$

Picking η to be $\frac{D}{L\sqrt{d_{\mathrm{tot}}+BT}}$ gives us

$$\mathrm{Reg}_T(u) = \mathcal{O}\left(DLN\sqrt{d_{\mathrm{tot}} + BT}\right) = \mathcal{O}\left(DLN\sqrt{d_{\mathrm{tot}} + \frac{\ln(N)}{\sqrt{1-\sigma_2(W)}}T}\right). \tag{D.16}$$

□

We now turn to proving the auxiliary lemmas invoked in the proof of the main theorem. The following lemma introduces the decomposition of the regret for AD-FTRL-DF.

Lemma D.4. *For any sequences $\{\bar{x}_s(v)\}_{s\in[T/B],v\in\mathcal{V}}$, $\bar{x}_s(v) \in \mathcal{X}$, the regret of Algorithm 5.1 can be bounded as*

$$\begin{aligned}\mathrm{Reg}_T(u) \le &\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s(u) - x^*\rangle \\ &+ 2BL\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}(\|\bar{x}_s(u) - \bar{x}_s(v)\|_2 + \|x_s(v) - \bar{x}_s(v)\|_2) + NBL\sum_{s=1}^{T/B}\|x_s(u) - \bar{x}_s(u)\|_2,\end{aligned}$$

where $\mathcal{T}_s \triangleq \{(s-1)B+1, \dots sB\}$ and $x^* = \arg\min_{x\in\mathcal{X}}\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\ell_t(v,x)$.

Proof. By definition of $\mathrm{Reg}_T(u)$, we know that

$$\mathrm{Reg}_T(u) = \sum_{t=1}^{T}\sum_{v\in\mathcal{V}}(\ell_t(v, x_t(u)) - \ell_t(v, x^*))$$

$$
\begin{aligned}
&= \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\ell_t(v, x_t(v)) - \ell_t(v, x^*)\right) + \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\ell_t(v, x_t(u)) - \ell_t(v, x_t(v))\right) \\
&\leq \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), x_t(v) - x^*\rangle\right) + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \|x_t(u) - x_t(v)\|_2 \\
&\qquad\qquad \text{(Assumption 5.2 and the convexity of } \ell_t\text{)} \\
&= \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), x_t(v) + \bar{x}_t(v) - \bar{x}_t(v) + \bar{x}_t(u) - \bar{x}_t(u) - x^*\rangle\right) + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \|x_t(u) - x_t(v)\|_2 \\
&= \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), \bar{x}_t(u) - x^*\rangle\right) + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\|\bar{x}_t(v) - \bar{x}_t(u)\|_2 + \|x_t(v) - \bar{x}_t(v)\|_2\right) \\
&\quad + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \|x_t(u) - x_t(v)\|_2 \qquad \text{(Assumption 5.2)} \\
&\leq \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), \bar{x}_t(u) - x^*\rangle\right) + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\|\bar{x}_t(v) - \bar{x}_t(u)\|_2 + \|x_t(v) - \bar{x}_t(v)\|_2\right) \\
&\qquad + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\|x_t(u) - \bar{x}_t(u)\|_2 + \|\bar{x}_t(u) - \bar{x}_t(v)\|_2 + \|x_t(v) - \bar{x}_t(v)\|_2\right) \\
&\qquad\qquad \text{(triangle inequality)} \\
&= \sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} \langle g_t(v), \bar{x}_s(u) - x^*\rangle \\
&\quad + 2BL\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}} \left(\|\bar{x}_s(u) - \bar{x}_s(v)\|_2 + \|x_s(v) - \bar{x}_s(v)\|_2\right) + NBL\sum_{s=1}^{T/B} \|x_s(u) - \bar{x}_s(u)\|_2,
\end{aligned}
\tag{D.17}
$$

where the last equality is due to the fact that the algorithm uses the same decision over all time steps in the same block. □

D.2.2 Properties induced by the gossiping mechanism

The following two lemmas characterize the properties induced by the accelerated gossiping mechanism used in Algorithm 5.1.

Lemma D.5. *For any $n \geq 0$, any $u \in \mathcal{V}$ and any $s \in [T/B - 1]$, we define*

$$
y_s^n(u) = y_s(u) \tag{D.18}
$$

if $n = 0$ or $n = -1$ and

$$
y_s^{n+1}(u) = (1+\theta)\sum_{v\in\mathcal{V}} W(u,v) y_s^n(u) - \theta y_s^{n-1}(u) \tag{D.19}
$$

otherwise. For any $k \geq 0$, any $u \in \mathcal{V}$ and any $s \in [T/B - 1]$, Algorithm 5.1 ensures

$$z_s^k(u) = \sum_{l=1}^{s-1} y_l^{(s-l-1)B+k}(v), \forall k = 1, \ldots, B. \tag{D.20}$$

Proof. The proof is taken from Lemma 2 in Wan et al. (2024b). We provide it here for completeness. We introduce a new notation $z_s(u)$ to denote $z_s^B(u)$. We use a double induction method. Recall that

$$y_s^0(u) = y_s^{-1}(u) = y_s(u). \tag{D.21}$$

It is easy to verify by induction on k that Equation (D.20) holds for $s = 2$ due to $z_2^0(u) = z_2^{-1}(u) = y_1(u)$ (initialization) and by using Equation (D.19) for the induction . Then, we assume that Equation (D.20) holds for some $s > 2$, and prove it also holds for $s + 1$. From the update of Algorithm 5.1, we have

$$\begin{aligned} z_{s+1}^0(u) &= z_s(u) + y_s(u) \\ &= z_s^B(u) + y_s^0(u) \\ &= \sum_{l=1}^{s} y_l^{(s-l)B}(u) \end{aligned}$$

and

$$\begin{aligned} z_{s+1}^{-1}(u) &= z_s^{B-1}(u) + y_s(u) \\ &= z_s^{B-1}(u) + y_s^{-1}(u) \\ &= \sum_{l=1}^{s} y_l^{(s-l)B-1}(u). \end{aligned}$$

By induction, suppose that $z_{s+1}^k(u)$ and $z_{s+1}^{k-1}(u)$ satisfy Equation (D.20). By the update of Algorithm 5.1, we have

$$\begin{aligned} z_{s+1}^k(u) &= (1+\theta) \sum_{v \in \mathcal{V}} W(u,v) z_{s+1}^{k-1}(v) - \theta z_{s+1}^{k-2}(u) \\ &= (1+\theta) \sum_{v \in \mathcal{V}} W(u,v) \sum_{l=1}^{s} y_l^{(s-l)B+k-1}(u) - \theta \sum_{l=1}^{s} y_l^{(s-l)B+k-2}(u) \\ &= \sum_{l=1}^{s} \left((1+\theta) \sum_{v \in \mathcal{V}} W(u,v) \sum_{l=1}^{s} y_l^{(s-l)B+k-1}(u) - \theta \sum_{l=1}^{s} y_l^{(s-l)B+k-2}(u) \right) \\ &= \sum_{l=1}^{s} y_s^{(s-l)B+k}(u), \end{aligned}$$

which suffices to complete the induction for block $s + 1$. □

The following lemma bounds the deviations between $z_s(u)$ and $\bar{z}_s$, for all agent $u \in \mathcal{V}$.

Lemma D.6. *Algorithm 5.1 guarantees that for any $u \in \mathcal{V}$, for any $s \in [1, T/B]$,*

$$\|z_s(u) - \bar{z}_s\|_2 \leq \frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right), \tag{D.22}$$

where $b = \left(1 - (1 - 1/\sqrt{2})\sqrt{1 - \sigma_2(W)}\right)$ *and* $B = \left\lceil \frac{\sqrt{2}\ln(\sqrt{14N})}{(\sqrt{2}-1)\sqrt{1-\sigma_2(W)}} \right\rceil$.

Proof. According to Equation 22 in Wan et al. (2024b), we know

$$b^B \leq \frac{1}{\sqrt{14N}}. \tag{D.23}$$

Then, with the same notation as in Lemma D.5,

$$\begin{aligned}
\|z_s(u) - \bar{z}_s\|_2 &= \left\| \sum_{l=1}^{s-1} y_l^{(s-l-1)B}(u) - \frac{1}{N} \sum_{l=1}^{s-1} \sum_{v \in \mathcal{V}} y_l(v) \right\|_2 && \text{(from Lemma D.5)} \\
&\leq \sum_{l=1}^{s-1} \left\| y_l^{(s-l-1)B}(u) - \frac{1}{N} \sum_{v \in \mathcal{V}} y_l^0(v) \right\|_2 && \text{(from the triangle inequality)} \\
&\leq \sum_{l=1}^{s-1} \left\| Y_l^{(s-l-1)B} - \bar{Y}_l \right\|_F \\
&\leq \sum_{l=1}^{s-1} \sqrt{14} b^{(s-l)B} \left\| Y_l^0 - \bar{Y}_l \right\|_F && \text{(from Proposition 5.1)} \\
&\leq \sum_{l=1}^{s-1} \sqrt{14} b^{(s-l)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \left\| y_l(v) - \frac{1}{N} \sum_{v \in \mathcal{V}} y_l(v) \right\|_2^2} \right) \\
&\leq \sum_{l=1}^{s-1} \sqrt{14} b^{(s-l)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|^2} + \sqrt{N \left\| \frac{1}{N} \sum_{v \in \mathcal{V}} y_l(v) \right\|_2^2} \right) && \text{(triangle inequality)} \\
&\leq \sum_{l=1}^{s-1} 2\sqrt{14} b^{(s-l)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) && \text{(D.24)} \\
&\leq \frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right), && \text{(D.25)}
\end{aligned}$$

where Y_s^n are defined as

$$Y_s^n = [y_s^{(n)}(0), y_s^{(n)}(1) \dots y_s^{(n)}(N)] \in \mathbb{R}^{N \times 1}$$

and in the third inequality, we apply Proposition 5.1 with $X^k = Y_s^k$. □

Similarly, we can show the following two lemmas for the accelerated gossiping mechanism in Algorithm 5.2 by replacing $y_s^n(u)$ with $q_s^n(u)$, $y_s(u)$ with $q_s(u)$, and $z_s^k(u)$ with $\zeta_s^k(u)$, noting that the gossip mechanisms for z in Algorithm 1 and for ζ in Algorithm 2 are identical. The proof for Lemma D.7 is omitted as they follow exactly the same steps as the one in Lemma D.5.

Lemma D.7. *For any $n \geq 0$, any $u \in \mathcal{V}$ and any $s \in [T/B - 1]$, we define*

$$q_s^n(u) = q_s(u). \tag{D.26}$$

if $n = 0$ or $n = -1$ and

$$q_s^{n+1}(u) = (1+\theta)\sum_{v\in\mathcal{V}} W(u,v)q_s^n(u) - \theta q_s^{n-1}(u). \tag{D.27}$$

otherwise. For any $k \geq 0$, any $u \in \mathcal{V}$ and any $s \in [T/B - 1]$, Algorithm 5.2 ensures

$$\zeta_s^k(u) = \sum_{l=1}^{s-1} q_l^{(s-l-1)B+k}(v), \forall k = 1, \ldots, B. \tag{D.28}$$

We introduce new notations $\widehat{M}_s(u)$ to denote $\zeta_s^B(u)$ and $M_s \triangleq \frac{1}{N}\sum_{k=1}^{s}\sum_{v\in[N]}|m_{kB+1,v}|$ to be the cumulative missing observations averaged over all agents till block s. Then, we can bound the deviations between $\widehat{M}_s(u)$ and $M_s(u)$ for all agents $u \in \mathcal{V}$ as follows. The proof follows a similar analysis to Lemma D.6.

Lemma D.8. *Algorithm 5.2 guarantees that for any $u \in \mathcal{V}$, for any $s \in [1, T/B]$,*

$$\left|\widehat{M}_s(u) - M_s\right| \leq \frac{2}{\sqrt{N}}\sum_{l=1}^{s-1} b^{(s-l-1)B}\left(\sqrt{\sum_{v\in\mathcal{V}}|m_{lB+1}(v)|^2}\right), \tag{D.29}$$

and consequently

$$\left|\widehat{M}_s(u) - M_s\right| \leq 3sB, \tag{D.30}$$

where $b = \left(1 - (1 - 1/\sqrt{2})\sqrt{1-\sigma_2(W)}\right)$ and $B = \left\lceil \frac{\sqrt{2}\ln(\sqrt{14N})}{(\sqrt{2}-1)\sqrt{1-\sigma_2(W)}} \right\rceil$.

Proof. From Equation 22 from Wan et al. (2024b), we obtain

$$b^B \leq \frac{1}{\sqrt{14N}}. \tag{D.31}$$

With the same notation as in Lemma D.7,

$$\begin{aligned}
\|\zeta_s(u) - M_s\| &= \left|\sum_{l=1}^{s-1} q_l^{(s-l-1)B}(u) - \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}} q_l(v)\right| && \text{(from Lemma D.7)}\\
&\leq \sum_{l=1}^{s-1}\left|q_l^{(s-l-1)B}(u) - \frac{1}{N}\sum_{v\in\mathcal{V}} q_l^0(v)\right| && \text{(from the triangle inequality)}\\
&\leq \sum_{l=1}^{s-1}\left\|Q_l^{(s-l-1)B} - \bar{Q}_l\right\|_F &&\\
&\leq \sum_{l=1}^{s-1}\sqrt{14}b^{(s-l)B}\left\|Q_l^0 - \bar{Q}_l\right\|_F && \text{(from Proposition 5.1)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{l=1}^{s-1} \sqrt{14} b^{(s-l)B} \left(\sqrt{\left| \sum_{v\in\mathcal{V}} q_l(v) - \frac{1}{N} \sum_{v\in\mathcal{V}} q_l(v) \right|^2} \right) \\
&\leq \sum_{l=1}^{s-1} \sqrt{14} b^{(s-l)B} \left(\sqrt{\sum_{v\in\mathcal{V}} |q_l(v)|^2} + \sqrt{N \left| \frac{1}{N} \sum_{v\in\mathcal{V}} q_l(v) \right|^2} \right) \qquad \text{(triangle inequality)} \\
&\leq \sum_{l=1}^{s-1} 2\sqrt{14} b^{(s-l)B} \left(\sqrt{\sum_{v\in\mathcal{V}} |q_l(v)|^2} \right) \\
&\leq \frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v\in\mathcal{V}} |q_l(v)|^2} \right) \qquad \text{(from Equation (D.31))} \\
&\leq \frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|} \right), \qquad \text{(D.32)}
\end{aligned}
$$

where Q_s^n are defined as

$$Q_s^n = [q_s^{(n)}(0), q_s^{(n)}(1) \dots q_s^{(n)}(N)] \in \mathbb{R}^{N\times 1}$$

and Proposition 5.1 is used with $X^k = Q_s^k$. Observing that $\zeta_s(u) = \widehat{M}_s(u)$ directly yields Equation (D.29).

It also holds that

$$|\zeta_s(u) - M_s| \leq \frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v\in\mathcal{V}} |Bs|^2} \right) \leq 2Bs \sum_{l=1}^{s-1} b^{(s-l-1)B} \leq 2Bs \frac{1}{1-b^B} \leq \frac{2}{1-\frac{1}{\sqrt{14N}}} Bs \leq 3Bs$$

thanks to Equation (D.31), which along with $\zeta_s(u) = \widehat{M}_s(u)$ directly yields the first inequality of Lemma D.8.

□

Similarly, we can establish the following lemma characterising the properties induced by the accelerated gossiping mechanism in Algorithm 5.3, by replacing $y_s(u)$ by $y_s^+(u)$ and by observing that the accelerated gossip mechanisms for z in Algorithm 5.1 and in Algorithm 5.3 are identical. The proof for Lemma D.9 is omitted for conciseness since it directly follows the proof of Lemma D.9.

Lemma D.9. *Algorithm 5.3 guarantees that for any $u \in \mathcal{V}$, for any $s \in [1, T/B]$,*

$$\|z_s(u) - \bar{z}_s\|_2 \leq \frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v\in\mathcal{V}} \|y_l^+(v)\|_2^2} \right), \qquad \text{(D.33)}$$

where $b = \left(1 - (1-1/\sqrt{2})\sqrt{1-\sigma_2(W)}\right)$ *and* $B = \left\lceil \frac{\sqrt{2}\ln(\sqrt{14N})}{(\sqrt{2}-1)\sqrt{1-\sigma_2(W)}} \right\rceil$.

The following lemma, used in the proof of Algorithm 5.1, provides a uniform upper bound on the square root of the cumulative squared norms of received gradient sums across all agents and blocks.

Lemma D.10. *It holds that*

$$\sum_{l=1}^{T/B-1}\left(\sqrt{\sum_{v\in\mathcal{V}}\|y_l(v)\|_2^2}\right)\leq NTL. \tag{D.34}$$

Proof. We have

$$\begin{aligned}
\sum_{s=1}^{T/B-1}\left(\sqrt{\sum_{v\in\mathcal{V}}\|y_l(v)\|_2^2}\right) &= \sum_{s=1}^{T/B-1}\left(\sqrt{\sum_{v\in\mathcal{V}}\left\|\sum_{\tau\in o_{sB+1}(v)\setminus o_{(s-1)B+1}(v)} g_\tau(v)\right\|_2^2}\right)\\
&\leq L\sum_{s=1}^{T/B}\left(\sqrt{\sum_{v\in\mathcal{V}}\left(|o_{sB+1}(v)|-|o_{(s-1)B+1(v)}|\right)^2}\right) && \text{(Assumption 5.2)}\\
&\leq L\sum_{s=1}^{T/B}\left(\sum_{v\in\mathcal{V}}\left(|o_{sB+1}(v)|-|o_{(s-1)B+1}(v)|\right)\right) && (\|\cdot\|_2\leq\|\cdot\|_1)\\
&\leq NTL,
\end{aligned}$$

where the last inequality holds because

$$\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{sB+1}(v)\setminus o_{(s-1)B+1}(v)}1=\sum_{v\in\mathcal{V}}\sum_{s=1}^{T/B}\sum_{\tau\in o_{sB+1}(v)\setminus o_{(s-1)B+1}(v)}1=\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{TB+1}(v)}1\leq NT.$$

□

D.2.3 Adaptive Algorithm with Unknown Total Delay

In this section, we show omitted details in Section "Non-Adaptive Algorithm with Known Total Delay". For completeness, we first restate the theorem and then present its proof.

Theorem 5.2. *Assuming each agent $u\in[N]$ runs an instance of Algorithm 5.2 with a valid communication matrix W and parameters θ and B defined in Equation* (5.6) *together with an instance of Algorithm 5.1 parametrized by the same W, θ and B and using $\eta_s(u)$ computed by Algorithm 5.2. Then, under Assumption 5.1 and Assumption 5.2, the regret is bounded as*

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

Proof. We define $\bar{z}_{s-1}$ and $\widetilde{z}_{s-1}$ as in Equation (D.1) and Appendix D.2.1, respectively:

$$\bar{z}_{s-1}\triangleq\frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}y_l(v), \tag{D.35}$$

$$\widetilde{z}_{s-1}\triangleq\frac{1}{N}\sum_{l=1}^{s-1}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}}g_\tau(v). \tag{D.36}$$

We also define the following:

$$\begin{aligned}\bar{x}_s(u) &\triangleq \arg\min_{x\in\mathcal{X}} \langle \bar{z}_{s-1}, x\rangle + \frac{1}{\eta_{s-1}(u)}\|x\|_2^2, \\ F_s(u,x) &\triangleq \langle \widetilde{z}_{s-1}, x\rangle + \frac{1}{\eta_{s-1}(u)}\|x\|_2^2, \\ \widetilde{x}_s(u) &\triangleq \arg\min_{x\in\mathcal{X}} F_s(u,x).\end{aligned} \tag{D.37}$$

Recall that in Algorithm 5.1, we have

$$x_1 = \mathbf{0} = \arg\min_{x\in\mathcal{X}} \frac{1}{\eta_0(u)}\|x\|_2^2,$$

where $\eta_0(u) = \eta_1(u) = \frac{D}{L\sqrt{\sqrt{N}BT+3B^2}}$.

Applying the regret decomposition proven in Lemma D.4 with the decision sequence $\{\bar{x}_s(u)\}_{s\in[T/B],u\in\mathcal{V}}$ defined in Equation (D.37), we know that

$$\begin{aligned}\mathrm{Reg}_T(u) \le& \underbrace{\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} \langle g_t(v), \bar{x}_s(u) - x^*\rangle}_{\spadesuit} \\ &+ \underbrace{2BL\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}} (\|\bar{x}_s(u) - \bar{x}_s(v)\|_2 + \|x_s(v) - \bar{x}_s(v)\|_2) + NBL\sum_{s=1}^{T/B}\|x_s(u) - \bar{x}_s(u)\|_2}_{\clubsuit}.\end{aligned} \tag{D.38}$$

We start by analyzing Term ♠. Similar to the non-adaptive learning rate analysis, we further decompose ♠ as follows:

$$\begin{aligned}&\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} \langle g_t(v), \bar{x}_s(u) - \widetilde{x}_s(u) + \widetilde{x}_s(u) - x^*\rangle \\ &= \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} \langle g_t(v), \widetilde{x}_s(u) - x^*\rangle}_{\text{full-info}_T} + \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} \langle g_t(v), \bar{x}_s(u) - \widetilde{x}_s(u)\rangle}_{\text{drift}_T}.\end{aligned} \tag{D.39}$$

For notational convenience, we define

$$\ell_s(x) \triangleq \left\langle \frac{1}{N}\sum_{\tau\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} g_t(v), x\right\rangle,$$

for all $s \in [T/B]$. Regarding full-info$_T$, by using Lemma 7.1 in Orabona (2025), we obtain

$$\text{full-info}_T = \frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} \langle g_t(v), \widetilde{x}_s(u) - x^*\rangle$$

$$= \sum_{s=1}^{T/B} \left\langle \frac{1}{N} \sum_{t\in\mathcal{T}_s} \sum_{v\in\mathcal{V}} g_t(v), \widetilde{x}_s(u) - x^* \right\rangle$$

$$\leq \frac{1}{\eta_{T/B}(u)}\|x^*\|_2^2 - \min_{x\in\mathcal{X}} \frac{1}{\eta_0(u)}\|x\|_2^2 + \sum_{s=1}^{T/B} \left[F_s(u, \widetilde{x}_s(u)) - F_{s+1}(u, \widetilde{x}_{s+1}(u)) + \ell_s(\widetilde{x}_s(u))\right] \tag{D.40}$$

$$+ F_{T/B+1}\left(u, \widetilde{x}_{T/B+1}\right) - F_{T/B+1}(u, x^*)$$

$$\leq \frac{D^2}{\eta_{T/B}(u)} + \sum_{s=1}^{T/B} \left[F_s(u, \widetilde{x}_s(u)) - F_{s+1}(u, \widetilde{x}_{s+1}(u)) + \ell_s(\widetilde{x}_s(u))\right] \tag{D.41}$$

where the last inequality holds because $F_{T/B+1}\left(u, \widetilde{x}_{T/B+1}\right) - F_{T/B+1}(u, x^*)$ is a negative term by definition of $\widetilde{x}_{T/B+1}(u)$, Assumption 5.1 and together with non-negativity of $\min_{x\in\mathcal{X}} \frac{1}{\eta_0(u)}\|x\|_2^2$.

To analyze the term $\sum_{s=1}^{T/B} \left[F_s(u, \widetilde{x}_s(u)) - F_{s+1}(u, \widetilde{x}_{s+1}(u)) + \ell_s(\widetilde{x}_s(u))\right]$, we proceed as follows:

$$\sum_{s=1}^{T/B} \left[F_s(u, \widetilde{x}_s(u)) - F_{s+1}(u, \widetilde{x}_{s+1}(u)) + \ell_s(\widetilde{x}_s(u))\right]$$

$$\leq \sum_{s=1}^{T/B} \left[\left\langle \nabla\ell_s\left(\widetilde{x}_s(u)\right), \widetilde{x}_s(u) - \widetilde{x}_{s+1}(u)\right\rangle - \frac{\lambda_{s-1}}{2}\left\|\widetilde{x}_s(u) - \widetilde{x}_{s+1}(u)\right\|_2^2 + \frac{1}{\eta_{s-1}(u)}\left\|\widetilde{x}_s(u)\right\|^2 - \frac{1}{\eta_s(u)}\left\|\widetilde{x}_{s+1}(u)\right\|_2^2\right] \tag{D.42}$$

$$\leq \sum_{s=1}^{T/B} \left[\left\|\nabla\ell_s\left(\widetilde{x}_s(u)\right)\right\|_2 \left\|\widetilde{x}_s(u) - \widetilde{x}_{s+1}(u)\right\|_2 - \frac{\lambda_{s-1}}{2}\left\|\widetilde{x}_s(u) - \widetilde{x}_{s+1}(u)\right\|^2 + \frac{1}{\eta_{s-1}(u)}\left\|\widetilde{x}_{s+1}(u)\right\|^2 - \frac{1}{\eta_s(u)}\left\|\widetilde{x}_s(u)\right\|^2\right]$$

(Cauchy–Schwarz inequality)

$$\leq \sum_{s=1}^{T/B} \left[\frac{1}{\lambda_{s-1}}\left\|\nabla\ell_s\left(\widetilde{x}_s(u)\right)\right\|_2^2 - \frac{\lambda_{s-1}}{4}\left\|\widetilde{x}_{s-1}(u) - \widetilde{x}_s(u)\right\|^2 + \frac{1}{\eta_{s-1}(u)}\left\|\widetilde{x}_{s+1}(u)\right\|^2 - \frac{1}{\eta_s(u)}\left\|\widetilde{x}_{s+1}(u)\right\|^2\right]$$

$(ab \leq \frac{a^2}{\lambda_{s-1}} + \frac{\lambda_{s-1}}{4}b^2)$

$$\leq \sum_{s=1}^{T/B} \left[\frac{1}{\lambda_{s-1}}\left\|\nabla\ell_s\left(\widetilde{x}_s(u)\right)\right\|_2^2 + \frac{1}{\eta_{s-1}(u)}\left\|\widetilde{x}_{s+1}(u)\right\|^2 - \frac{1}{\eta_s(u)}\left\|\widetilde{x}_{s+1}(u)\right\|^2\right]$$

$$\leq \sum_{s=1}^{T/B} \left[\frac{B^2L^2}{2}\eta_{s-1}(u) + \frac{1}{\eta_{s-1}(u)}\left\|\widetilde{x}_{s+1}(u)\right\|_2^2 - \frac{1}{\eta_s(u)}\left\|\widetilde{x}_{s+1}(u)\right\|_2^2\right], \tag{D.43}$$

where the first inequality is because $\frac{1}{\eta_s(u)}\|x\|_2^2$ is λ_s-strongly convex convexity and $\lambda_s = 2/\eta_s(u)$, and the last inequality is because of Assumption 5.2.

Combining Equation (D.41) and Equation (D.43), we obtain

$$\frac{1}{N}\,\text{full-info}_T \leq \frac{D^2}{\eta_{T/B}(u)} + \frac{B^2L^2}{2}\sum_{s=1}^{T/B}\eta_{s-1}(u) + \sum_{s=1}^{T/B}\left(\frac{1}{\eta_{s-1}(u)} - \frac{1}{\eta_s(u)}\right)\left\|\widetilde{x}_{s+1}(u)\right\|_2^2$$

$$= \frac{D^2}{\eta_{T/B}(u)} + \frac{B^2L^2}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u) + \sum_{s=1}^{T/B}\left(\frac{1}{\eta_{s-1}(u)} - \frac{1}{\eta_s(u)}\right)\left\|\widetilde{x}_{s+1}(u)\right\|_2^2 + \frac{B^2L^2}{2}\eta_0 \tag{D.44}$$

We now analyze the drift term drift_T. By definition, we have:

$$\begin{aligned}
\frac{1}{N}\text{drift}_T &= \frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s(u) - \widetilde{x}_s(u)\rangle \\
&\le BL\sum_{s=1}^{T/B}\|\bar{x}_s(u) - \widetilde{x}_s(u)\|_2 && \text{(Cauchy–Schwarz and Assumption 5.2)} \\
&= BL\sum_{s=2}^{T/B}\|\bar{x}_s(u) - \widetilde{x}_s(u)\|_2 && (\bar{x}_1(u) = \widetilde{x}_1(u) = \mathbf{0}) \\
&\le \frac{BL}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\,\|\bar{z}_{s-1} - \widetilde{z}_{s-1}\|_2 && \text{(Lemma D.1)} \\
&\le \frac{BL}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\left\|\frac{1}{N}\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{(s-1)B+1}(v)} g_\tau(v) - \frac{1}{N}\sum_{l=1}^{s-1}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}} g_\tau(v)\right\|_2 \\
& && \text{(Definition of } \bar{z}_{s-1} \text{ and } \widetilde{z}_{s-1}) \\
&= \frac{BL}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\left\|-\frac{1}{N}\sum_{v\in\mathcal{V}}\sum_{\tau\in m_{(s-1)B+1}(v)} g_\tau(v)\right\|_2 \\
& && (\mathcal{T}_s = \{(s-1)B+1,\dots,sB\},\ m_t(v) = [t-1]\setminus o_t(v)) \\
&\le \frac{BL^2}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|\right), && \text{(D.45)}
\end{aligned}$$

Combining Equation (D.39), Equation (D.44), and Equation (D.45), we obtain:

$$\begin{aligned}
\frac{1}{N}\spadesuit &\le \frac{D^2}{\eta_{T/B}(u)} + \frac{BL^2}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)| + B\right) + \frac{B^2L^2}{2}\eta_0 \\
&\quad + \sum_{s=1}^{T/B}\left(\frac{1}{\eta_{s-1}(u)} - \frac{1}{\eta_s(u)}\right)\|\widetilde{x}_{s+1}(u)\|_2^2 && \text{(D.46)} \\
&\le \frac{D^2}{\eta_{T/B}(u)} + \frac{BL^2}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)| + B\right) + \frac{1}{2}BDL \\
&\quad + \sum_{s=1}^{T/B}\left(\frac{1}{\eta_{s-1}(u)} - \frac{1}{\eta_s(u)}\right)\|\widetilde{x}_{s+1}(u)\|_2^2 && \text{(D.47)}
\end{aligned}$$

Recall that M_s is defined as $M_s \triangleq \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}|m_{lB+1}(v)|$, and $\widehat{M}_s(u) \triangleq \zeta_s^B(u)$.

Now we focus on the second term in Equation (D.47). Using the bound $\left|\widehat{M}_s(u) - M_s\right| \le 3sB$ from Lemma D.8, we obtain

$$\eta_{s+1}(u) = \frac{D}{L\sqrt{B\sqrt{N}T + B\widehat{M}_s(u) + 3sB^2}} \le \frac{D}{L\sqrt{\sqrt{N}BT + BM_s}} \le \frac{D}{L\sqrt{BT + BM_s}} \le \frac{D}{L\sqrt{BT}}, \forall u \in \mathcal{V}. \tag{D.48}$$

We start with the first term in Equation (D.47). We have

$$\begin{aligned}
\frac{D^2}{\eta_{T/B}(u)} &= DL\sqrt{\sqrt{N}BT + B\widehat{M}_{T/B-1}(u) + 3B(T-1)} && \text{(the definition of } \eta_s(u)\text{)} \\
&\leq DL\sqrt{\sqrt{N}BT + BM_{T/B-1} + 6B(T-1)} \\
&\qquad\qquad \text{(using the bound } \left|\widehat{M}_s(u) - M_s\right| \leq 3sB \text{ from Lemma D.8)} \\
&\leq DL\sqrt{\sqrt{N}BT + BM_{T/B} + 6BT} && (M_s \text{ is non-decreasing}) \\
&\leq DL\sqrt{\sqrt{N}BT + d_{\text{tot}} + 7BT} && \text{(Lemma D.3)} \\
&\leq DL\sqrt{8\sqrt{N}BT + d_{\text{tot}}}, && \text{(D.49)}
\end{aligned}$$

Focus on the second term in Equation (D.47), we thus have

$$\begin{aligned}
&\frac{BL^2}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)| + B\right) \\
&\leq \frac{BL^2}{2}\sum_{s=2}^{T/B}\eta_{s-1}(u)\left(\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|\right) + \frac{1}{2}DL\sqrt{BT} && \text{(Equation (D.48))} \\
&\leq DL\left(B\sum_{s=3}^{T/B}\frac{\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|}{\sqrt{BT + BM_{s-2}}}\right) + \frac{1}{2}DL\sqrt{BT} + \frac{1}{2}BDL \\
&\qquad\qquad (|m_{B+1}(v)| \leq B \text{ and } \eta_1(u) = \tfrac{D}{L\sqrt{\sqrt{N}BT+3B^2}}) \\
&= 2DL\left(B\sum_{s=3}^{T/B}\frac{\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|}{\sqrt{4BT + 4BM_{s-2}}}\right) + \frac{1}{2}DL\sqrt{BT} + \frac{1}{2}BDL \\
&= 2DL\left(B\sum_{s=3}^{T/B}\frac{\frac{1}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|}{\sqrt{BM_s}}\right) + \frac{1}{2}DL\sqrt{BT} + \frac{1}{2}BDL && (|M_s| \leq |M_{s-1}| + T) \\
&= 2DL\left(\sum_{s=3}^{T/B}\frac{\frac{B}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|}{\sqrt{\frac{B}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}|m_{lB+1}|}}\right) + \frac{1}{2}DL\sqrt{BT} + \frac{1}{2}BDL && \text{(definition of } |M_s|\text{)} \\
&\leq 4DL\sqrt{BM_{T/B}} + \frac{1}{2}DL\sqrt{BT} + \frac{1}{2}BDL && \text{(Lemma D.2)} \\
&\leq 4DL\sqrt{BT + d_{\text{tot}}} + \frac{1}{2}DL\sqrt{BT} + \frac{1}{2}BDL, && \text{(D.50)}
\end{aligned}$$

where the last inequality holds because of Lemma D.3.

Let us now analyze the third term of Equation (D.47).

$$\begin{aligned}
&\sum_{s=1}^{T/B}\left(\frac{1}{\eta_{s-1}(u)} - \frac{1}{\eta_s(u)}\right)\|\widetilde{x}_{s+1}(u)\|_2^2 \\
&\quad= \sum_{s=2}^{T/B}\left(\frac{1}{\eta_{s-1}(u)} - \frac{1}{\eta_s(u)}\right)\|\widetilde{x}_{s+1}(u)\|_2^2 && (\eta_0(u) = \eta_1(u) = \tfrac{D}{L\sqrt{\sqrt{N}BT+3B^2}})
\end{aligned}$$

$$
\begin{aligned}
&\leq D^2 \sum_{s=1}^{T/B-1} \left| \frac{1}{\eta_s(u)} - \frac{1}{\eta_{s+1}(u)} \right| && \text{(Assumption 5.1)} \\
&\leq D^2 \sum_{s=1}^{T/B} \left| \frac{1}{\eta_s(u)} - \frac{1}{\eta_{s+1}(u)} \right| \\
&= D^2 \sum_{s=1}^{T/B} \left| \frac{1/\eta_{s+1}(u)^2 - 1/\eta_s(u)^2}{1/\eta_s(u) + 1/\eta_{s+1}(u)} \right| \\
&\leq DL \sum_{s=1}^{T/B} \frac{\left(B \left| \widehat{M}_s(u) - \widehat{M}_{s-1}(u) \right| + 3B^2 \right)}{\sqrt{B\sqrt{N}T + B\widehat{M}_s(u) + 3sB^2} + \sqrt{B\sqrt{N}T + B\widehat{M}_{s-1}(u) + 3(s-1)B^2}} \\
& && \text{(plugging the definition of the learning rate)} \\
&\leq DL \sum_{s=1}^{T/B} \frac{\left(B \left| \widehat{M}_s(u) - \widehat{M}_{s-1}(u) \right| + 3B^2 \right)}{\sqrt{B\sqrt{N}T + BM_s}} && \text{(using Equation (D.30) in Lemma D.8)} \\
&\leq DL \sum_{s=1}^{T/B} \frac{\left(B \left| \widehat{M}_s(u) - \widehat{M}_{s-1}(u) \right| + 3B^2 \right)}{\sqrt{BT + BM_s}}. && \text{(D.51)}
\end{aligned}
$$

Now decomposing the numerator and using the triangle inequality,

$$
\begin{aligned}
&\sum_{s=1}^{T/B} \left(\frac{1}{\eta_{s-1}(u)} - \frac{1}{\eta_s(u)} \right) \|\widetilde{x}_{s+1}(u)\|_2^2 \\
&\leq DL \sum_{s=1}^{T/B} \frac{B \left(|\widehat{M}_s(u) - M_s| + |M_s - M_{s-1}| + |\widehat{M}_{s-1}(u) - M_{s-1}| \right) + 3B^2}{\sqrt{BT + BM_s}} \\
&\leq DL \sum_{s=1}^{T/B} \frac{B \left(|\widehat{M}_s(u) - M_s| + |M_s - M_{s-1}| + |\widehat{M}_{s-1}(u) - M_{s-1}| \right)}{\sqrt{BT + BM_s}} + 3DL\sqrt{BT} \\
&\leq DL \sum_{s=1}^{T/B} \frac{B|M_s - M_{s-1}| + \frac{2B}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} |m_{lB+1}(v)|^2} \right)}{\sqrt{BT + BM_s}} \\
&\quad + \frac{2B}{\sqrt{N}} \sum_{l=1}^{s-2} b^{(s-l-2)B} \left(\sqrt{\sum_{v \in \mathcal{V}} |m_{lB+1}(v)|^2} \right) \Big/ \sqrt{BT + BM_s} + 3DL\sqrt{BT} \\
& \qquad \text{(from Equation (D.29) in Lemma D.8)} \\
&\leq DL \sum_{s=1}^{T/B} \frac{\frac{B}{N} \sum_{v \in \mathcal{V}} |m_{(s-1)B+1}(v)| + \frac{2B}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} |m_{lB+1}(v)|^2} \right)}{\sqrt{BT + BM_s}} \\
&\quad + \frac{2B}{\sqrt{N}} \sum_{l=1}^{s-2} b^{(s-l-2)B} \left(\sqrt{\sum_{v \in \mathcal{V}} |m_{lB+1}(v)|^2} \right) \Big/ \sqrt{BT + BM_s} + 3DL\sqrt{BT} \quad \text{(definition of } M_s\text{)} \\
&= DL \sum_{s=1}^{T/B} \frac{\frac{2B}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} |m_{lB+1}(v)|^2} \right)}{\sqrt{BT + BM_s}} \\
&\quad + DL \sum_{s=1}^{T/B} \frac{\frac{2B}{\sqrt{N}} \sum_{l=1}^{s-2} b^{(s-l-2)B} \left(\sqrt{\sum_{v \in \mathcal{V}} |m_{lB+1}(v)|^2} \right)}{\sqrt{BT + BM_s}} \qquad \text{(a rearranging of terms)}
\end{aligned}
$$

$$+\sum_{s=1}^{T/B}\frac{\frac{B}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|}{\sqrt{BT+BM_s}}+3DL\sqrt{BT} \tag{D.52}$$

Let us consider the first two summation terms in Equation (D.52). We have

$$\begin{aligned}
&\frac{\sum_{s=1}^{T/B}\frac{2B}{\sqrt{N}}\sum_{l=1}^{s-1}b^{(s-l-1)B}\left(\sqrt{\sum_{v\in\mathcal{V}}|m_{lB+1}(v)|^2}\right)}{\sqrt{BT+BM_s}}\\
&=\frac{2B}{\sqrt{N}}\sum_{l=1}^{T/B-1}\sum_{s=1+l}^{T/B}\frac{b^{(s-l-1)B}\left(\sum_{v\in\mathcal{V}}|m_{lB+1}(v)|\right)}{\sqrt{BT+BM_s}} && \text{(swapping the order of summation)}\\
&\le\frac{1}{1-b^B}\frac{2B}{\sqrt{N}}\sum_{l=1}^{T/B-1}\left(\sum_{v\in\mathcal{V}}\frac{|m_{lB+1}(v)|}{\sqrt{BT+BM_{l+1}}}\right) && (M_l \text{ is non-decreasing})\\
&\le\frac{8B}{\sqrt{N}}\sum_{l=1}^{T/B-1}\left(\sum_{v\in\mathcal{V}}\frac{|m_{lB+1}(v)|}{\sqrt{BT+BM_{l+1}}}\right) && (\text{ from } \tfrac{1}{1-b^B}\le\tfrac{1}{1-1/(14\sqrt{N})}\le 4)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\frac{\sum_{s=1}^{T/B}\frac{2B}{\sqrt{N}}\sum_{l=1}^{s-2}b^{(s-l-2)B}\left(\sqrt{\sum_{v\in\mathcal{V}}|m_{lB+1}(v)|^2}\right)}{\sqrt{BT+BM_s}}\\
&\le\frac{8B}{\sqrt{N}}\sum_{l=1}^{T/B-2}\left(\sum_{v\in\mathcal{V}}\frac{|m_{lB+1}(v)|}{\sqrt{BT+BM_{l+2}}}\right)\le\frac{8B}{\sqrt{N}}\sum_{l=1}^{T/B-1}\left(\sum_{v\in\mathcal{V}}\frac{|m_{lB+1}(v)|}{\sqrt{BT+BM_{l+1}}}\right).
\end{aligned}$$

Plugging above two inequalities back into Equation (D.52),

$$\begin{aligned}
&\sum_{s=1}^{T/B}\left(\frac{1}{\eta_{s-1}(u)}-\frac{1}{\eta_s(u)}\right)\|\widetilde{x}_{s+1}(u)\|_2^2\\
&\le DL\left(16\sqrt{N}\sum_{s=1}^{T/B-1}\frac{\frac{B}{N}\sum_{v\in\mathcal{V}}|m_{sB+1}(v)|}{\sqrt{BT+M_{s+1}}}+\sum_{s=1}^{T/B}\frac{\frac{B}{N}\sum_{v\in\mathcal{V}}|m_{(s-1)B+1}(v)|}{\sqrt{BT+BM_s}}+3\sqrt{BT}\right)\\
&\le DL\left(17\sqrt{N}\sum_{s=1}^{T/B}\frac{\frac{B}{N}\sum_{v\in\mathcal{V}}|m_{sB+1}(v)|}{\sqrt{B/N\sum_{l=1}^{s}\sum_{v\in\mathcal{V}}|m_{lb+1,v}|}}+3\sqrt{BT}\right) && \text{(definition of } M_s)\\
&\le DL\left(17\sqrt{N}\sqrt{\frac{B}{N}\sum_{l=1}^{T/B}\left(\sum_{v\in\mathcal{V}}|m_{lB+1}(v)|\right)}+3\sqrt{BT}\right) && \text{(Lemma D.2)}\\
&\le DL\left(17\sqrt{N}\sqrt{d_{\text{tot}}+BT}+3\sqrt{BT}\right). && \text{(Lemma D.3)}
\end{aligned}$$

The above inequality, together with Equation (D.49),Equation (D.50) and Equation (D.47), yields

$$\begin{aligned}
\spadesuit&\le N\left(DL\sqrt{8\sqrt{N}BT+d_{\text{tot}}}+4DL\sqrt{BT+d_{\text{tot}}}+\frac{1}{2}DL\sqrt{BT}+BDL+DL\left(17\sqrt{N}\sqrt{d_{\text{tot}}+BT}+3\sqrt{BT}\right)\right)\\
&\le 33N\sqrt{N}DL\sqrt{BT}+33N\sqrt{N}DL\sqrt{d_{\text{tot}}}+NBDL\\
&=\mathcal{O}(N^{1.5}DL\sqrt{BT+d_{\text{tot}}}+NBDL) \qquad (D.53)
\end{aligned}$$

Let us now turn to the analysis of Term ♣. From Lemma D.6, we have for all $w \in [N]$,

$$\|z_s(w) - \bar{z}_s\|_2 \leq \frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right). \tag{D.54}$$

Define $\bar{\eta}$ such that $\eta_s(v) \leq \bar{\eta} \triangleq \frac{D}{L\sqrt{B\sqrt{NT}}}$ for all $v \in \mathcal{V}$. Note that $x_1(v) = \bar{x}_1 = \mathbf{0}$. Using Lemma D.1, we know that for any $w \in V$,

$$\begin{aligned}
\sum_{s=1}^{T/B} \|x_s(w) - \bar{x}_s(w)\|_2 &= \sum_{s=2}^{T/B} \|x_s(w) - \bar{x}_s(w)\|_2 \\
&\leq \sum_{s=1}^{T/B-1} \eta_s(w) \|z_s(w) - \bar{z}_s\|_2 && \text{(according to Lemma D.1)} \\
&= \frac{2}{\sqrt{N}} \sum_{s=1}^{T/B-1} \eta_s(w) \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) && \text{(Equation (D.7))} \\
&\leq \frac{2\bar{\eta}}{\sqrt{N}} \sum_{s=1}^{T/B-1} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) \\
&= \frac{2\bar{\eta}}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \sum_{s=l+1}^{T/B-1} \cdot b^{(s-l-1)B} \right) \\
& && \text{(Swapping the order of summation)} \\
&\leq \frac{2\bar{\eta}}{\sqrt{N}} \frac{1}{1 - \frac{1}{\sqrt{14N}}} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) && \text{(Equation (D.31))} \\
&\leq \frac{3\bar{\eta}}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right). && \text{(D.55)}
\end{aligned}$$

Moreover, according to Lemma D.10, we have

$$\sum_{s=1}^{T/B-1} \left(\sqrt{\sum_{v \in \mathcal{V}} \|y_l(v)\|_2^2} \right) \leq NTL.$$

Therefore, combining the above two inequalities, we know that

$$\sum_{s=2}^{T/B} \|x_s(w) - \bar{x}_s(w)\|_2 \leq 3\bar{\eta}\sqrt{N}TL,$$

leading to a bound on term ♣:

$$\clubsuit \leq 18BN\bar{\eta}\sqrt{N}TL^2 + 2BL \sum_{v \in \mathcal{V}} \sum_{s=1}^{T/B} \|\bar{x}_s(u) - \bar{x}_s(v)\|_2. \tag{D.56}$$

Now we bound the second term in Equation (D.56). By using Lemma D.1, we obtain

$$\begin{aligned}
&\frac{1}{\eta_s(u)}\|\bar{x}_{s+1}(u)-\bar{x}_{s+1}(v)\|_2^2+\frac{1}{\eta_s(v)}\|\bar{x}_{s+1}(u)-\bar{x}_{s+1}(v)\|_2^2\\
&\leq \frac{1}{\eta_s(u)}\|\bar{x}_{s+1}(v)\|_2^2-\frac{1}{\eta_s(v)}\|\bar{x}_{s+1}(v)\|_2^2+\frac{1}{\eta_{s+1}(v)}\|\bar{x}_{s+1}(u)\|_2^2-\frac{1}{\eta_s(u)}\|\bar{x}_{s+1}(u)\|_2^2
\end{aligned}$$

Rearranging the last inequality, we have

$$\begin{aligned}
&\frac{\eta_s(u)+\eta_s(v)}{\eta_s(v)\eta_s(u)}\|\bar{x}_{s+1}(u)-\bar{x}_{s+1}(v)\|_2^2\\
&\leq \frac{\eta_s(u)-\eta_s(v)}{\eta_s(v)\eta_s(u)}(\|\bar{x}_{s+1}(u)\|_2^2-\|\bar{x}_{s+1}(v)\|_2^2)\\
&\leq \left|\frac{\eta_s(u)-\eta_s(v)}{\eta_s(v)\eta_s(u)}\right|\|\bar{x}_{s+1}(u)-\bar{x}_{s+1}(v)\|_2\cdot\|\bar{x}_{s+1}(u)+\bar{x}_{s+1}(v)\|_2.
\end{aligned}$$

Since $\|\bar{x}_{s-1}(u)-\bar{x}_{s-1}(v)\|_2\geq 0$ and Assumption 5.1, we have

$$\|\bar{x}_{s+1}(u)-\bar{x}_{s+1}(v)\|_2\leq 2D\left|\frac{\eta_s(u)-\eta_s(v)}{\eta_s(u)+\eta_s(v)}\right|. \tag{D.57}$$

By pure algebraic computations, we have

$$\left|\frac{\eta_s(u)-\eta_s(v)}{\eta_s(u)+\eta_s(v)}\right|=\left|\frac{\frac{\eta_s(v)-\eta_s(u)}{\eta_s(v)\eta_s(u)}}{\frac{\eta_s(u)+\eta_s(v)}{\eta_s(v)\eta_s(u)}}\right|=\left|\frac{\eta_s(u)^{-1}-\eta_s(v)^{-1}}{\eta_s(u)^{-1}+\eta_s(v)^{-1}}\right|=\frac{\left|(\eta_s(u)^{-1})^2-(\eta_s(v)^{-1})^2\right|}{(\eta_s(u)^{-1}+\eta_s(v)^{-1})^2}\leq\frac{\left|(\eta_s(u)^{-1})^2-(\eta_s(v)^{-1})^2\right|}{(\eta_s(u)^{-2}+\eta_s(v)^{-2})}.$$

Combining this with Equation (D.57),

$$\begin{aligned}
\sum_{s=1}^{T/B}\|\bar{x}_s(u)-\bar{x}_s(v)\|_2&=\sum_{s=2}^{T/B}\|\bar{x}_s(u)-\bar{x}_s(v)\|_2 && (\bar{x}_s(u)=\bar{x}_s(v)=\mathbf{0})\\
&\leq 2D\sum_{s=1}^{T/B-1}\frac{\left|\eta_s(u)^{-2}-\eta_s(v)^{-2}\right|}{(\eta_s(u)^{-2}+\eta_s(v)^{-2})}\\
&\leq 2D\sum_{s=2}^{T/B-1}\frac{\left|\eta_s(u)^{-2}-\eta_s(v)^{-2}\right|}{(\eta_s(u)^{-2}+\eta_s(v)^{-2})} && (\eta_1(u)=\eta_1(v)=\tfrac{D}{L\sqrt{\sqrt{N}BT+3B^2}})\\
&\leq 2D\sum_{s=1}^{T/B}\frac{\left|\left(\frac{D}{L\sqrt{TB\sqrt{N}+B\widehat{M}_s(u)+3sB^2}}\right)^{-2}-\left(\frac{D}{L\sqrt{TB\sqrt{N}+B\widehat{M}_s(v)+3sB^2}}\right)^{-2}\right|}{\left(\frac{D}{L\sqrt{TB\sqrt{N}+B\widehat{M}_s(u)+3sB^2}}\right)^{-2}+\left(\frac{D}{L\sqrt{TB\sqrt{N}+B\widehat{M}_s(v)+3sB^2}}\right)^{-2}}\\
&=2DB\sum_{s=1}^{T/B}\frac{\left|\widehat{M}_s(u)-\widehat{M}_s(v)\right|}{2\sqrt{N}BT+6sB^2+B\widehat{M}_s(u)+B\widehat{M}_s(v)} && \text{(re-arranging)}\\
&\leq DB\sum_{s=1}^{T/B}\frac{\left|\widehat{M}_s(u)-\widehat{M}_s(v)\right|}{\sqrt{N}BT+BM_s} && (\text{ using Equation (D.30) in Lemma D.8})
\end{aligned}$$

$$\leq DB \sum_{s=1}^{T/B} \frac{\frac{2}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|^2}\right)}{\sqrt{N}BT + BM_s}.$$

(using Equation (D.29) in Lemma D.8)

Now notice that we have

$$\begin{aligned}
&\frac{\sum_{s=1}^{T/B} \frac{1}{\sqrt{N}} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|^2}\right)}{BM_s + \sqrt{N}BT} \\
&\leq \sum_{s=1}^{T/B} \frac{1}{\sqrt{N}} \sum_{l=1}^{s-1} \frac{b^{(s-l-1)B} \left(\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|\right)}{BM_s + \sqrt{N}BT} && \text{(D.58)} \\
&\leq \frac{1}{\sqrt{N}} \sum_{l=1}^{T/B-1} \sum_{s=1+l}^{T/B} \frac{b^{(s-l-1)B} \left(\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|\right)}{BM_s + \sqrt{N}BT} && \text{(swapping sums)} \\
&\leq \frac{1}{1-b^B} \frac{1}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\sum_{v\in\mathcal{V}} \frac{|m_{lB+1}(v)|}{BM_{l+1} + \sqrt{N}BT}\right) && (M_l \text{ is non-decreasing}) \\
&\leq \frac{1}{1-\frac{1}{\sqrt{14N}}} \frac{1}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\sum_{v\in\mathcal{V}} \frac{|m_{lB+1}(v)|}{BM_{l+1} + \sqrt{N}BT}\right) && (\text{since } b^B \leq \tfrac{1}{\sqrt{14N}} \text{ shown in Equation (D.23)}) \\
&\leq \frac{4\sqrt{N}}{B} \sum_{l=1}^{T/B-1} \left(\frac{\frac{B}{N}\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|}{BM_{l+1} + \sqrt{N}BT}\right). && (\text{ from } \tfrac{1}{1-b^B} \leq \tfrac{1}{1-1/(14\sqrt{N})} \leq 4)
\end{aligned}$$

To analyze the term $\sum_{l=1}^{T/B-1} \left(\frac{\frac{B}{N}\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|}{BM_{l+1}+\sqrt{N}BT}\right)$, we use Lemma D.2 and Lemma D.3:

$$\begin{aligned}
\sum_{l=1}^{T/B-1} \left(\frac{\frac{B}{N}\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|}{BM_{l+1} + BT}\right) &= \sum_{l=1}^{T/B-1} \left(\frac{\frac{B}{N}\sum_{v\in\mathcal{V}} |m_{lB+1}(v)|}{\frac{B}{N}\sum_{\tau=1}^{l}\sum_{v\in\mathcal{V}} |m_{\tau B+1}(v)| + BT}\right) \\
&\qquad \text{(by definition of } M_{l+1}) \\
&\leq \ln\left(BT + \sum_{l=1}^{T/B-1} \frac{B}{N} \sum_{v\in\mathcal{V}} |m_{lB+1}(v)|\right) - \ln(BT) \\
&\qquad \text{(using Lemma D.2)} \\
&\leq \ln\left(BT + \sum_{l=1}^{T/B-1} \frac{B}{N} \sum_{v\in\mathcal{V}} |m_{lB+1}(v)|\right) \\
&\leq \ln\left(2BT + d_{\text{tot}}\right). \qquad \text{(using Lemma D.3)}
\end{aligned}$$

Combining the above two bounds, we can obtain the bound for $\sum_{s=1}^{T/B} \|\bar{x}_s(u) - \bar{x}_s(v)\|_2$:

$$\begin{aligned}
\sum_{s=1}^{T/B} \|\bar{x}_s(u) - \bar{x}_s(v)\|_2 &\leq 8DB\frac{\sqrt{N}}{B} \ln(2BT + d_{\text{tot}}) \\
&= 8D\sqrt{N} \ln(2BT + d_{\text{tot}}) \\
&\leq 8D\sqrt{N} \ln(2BT + T^2),
\end{aligned}$$

where the last inequality uses $d_{\text{tot}} \leq T^2$.

Plugging the above inequality into Equation (D.56), we have

$$\clubsuit \leq 18BL^2N\sqrt{N}\bar{\eta}T + 16BDLN\sqrt{N}\ln(2BT+T^2). \tag{D.59}$$

Recall that $\bar{\eta} = \frac{D}{L\sqrt{BN\sqrt{N}T}}$. The above upper bound further implies the following bound on $\clubsuit$.

$$\clubsuit \leq 18DLN^{5/4}\sqrt{BT} + 16BDLN^{3/2}\ln(2BT+T^2) = \mathcal{O}(DLN^{5/4}\sqrt{BT} + BDLN^{3/2}\ln(BT+T^2)). \tag{D.60}$$

Finally, combining the above inequality, Equation (D.38), and Equation (D.53), we obtain

$$\begin{aligned}\mathrm{Reg}_T(u) &\leq \clubsuit + \spadesuit \\ &\leq \mathcal{O}(N^{1.5}DL\sqrt{BT+d_{\mathrm{tot}}} + NBDL) + \mathcal{O}(DLN^{5/4}\sqrt{BT} + BDLN^{3/2}\ln(BT+T^2)) \\ &\leq \mathcal{O}(N^{1.5}DL\sqrt{BT+d_{\mathrm{tot}}} + NBDL + BDLN^{3/2}\ln(BT+T^2)).\end{aligned}$$

Plugging in the form of $B = \Theta\left(\frac{\ln N}{\sqrt{1-\sigma_2(W)}}\right)$, we obtain our final bound:

$$\mathrm{Reg}_T(u) = \widetilde{\mathcal{O}}\left(DL\left(\frac{N^{3/2}}{(1-\sigma_2(W))^{1/4}}\sqrt{T} + N^{3/2}\sqrt{d_{\mathrm{tot}}}\right)\right).$$

□

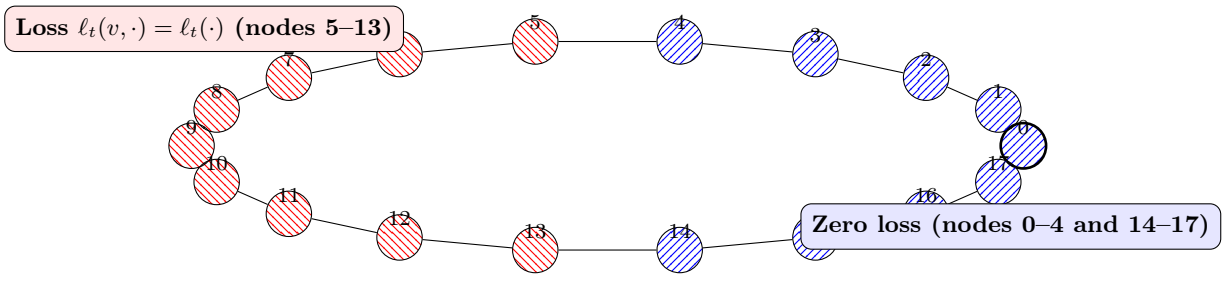
D.2.4 Lower Bound for the general convex case

In this section, we show omitted details for the lower bound for the general convex case. For completeness, we first restate the theorem and then present its proof.

Theorem 5.3. *Let d be the constant feedback delay suffered by all agents $u \in [N]$ in the network. Then, there exists a graph $\mathcal{G} = (\{0,1,\ldots,N\},\mathcal{E})$, with $N = 2M+1$ where M is an even integer, and a sequence of L-Lipschitz loss functions $\{\ell_1(0,\cdot),\ldots,\ell_1(N,\cdot)\},\ldots,\{\ell_T(0,\cdot),\ldots,\ell_T(N,\cdot)\}$ such that any algorithm has to suffer regret at least:*

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

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



Proof. We consider the setting where all delays are fixed and equal to d, and we let $\mathcal{G}$ denote a cycle graph with $N = 2(M+1)$ nodes where M is even, to simplify.

For the N-cycle graph, the smallest nonzero and largest eigenvalues of the Laplacian are given by $\sigma_{N-1}(\mathrm{Lap}(\mathcal{G})) = 2 - 2\cos(2\pi/N)$ and $\sigma_1(\mathrm{Lap}(\mathcal{G})) = 4$, respectively (Spielman, 2019, Chapter 5.5). Applying the inequality $1 - \cos(x) \geq x^2/5$ for all $x \in [0, \pi]$ (which holds since $N \geq 4 \Rightarrow 2\pi/N \leq \pi$), we obtain

$$\sigma_{N-1}(\mathrm{Lap}(\mathcal{G})) \geq \frac{8\pi^2}{5N^2}.$$

Consequently, we derive the bound

$$\frac{1}{1 - \sigma_2(W)} \leq \frac{N^2}{2} \tag{D.61}$$

since $\sigma_2(W) = \sigma_2(I - \frac{1}{\sigma_1(\mathrm{Lap}(\mathcal{G}))}\mathrm{Lap}(\mathcal{G})) = \sigma_2(I - \frac{1}{4}\mathrm{Lap}(\mathcal{G}))$ is the second highest eigenvalue of $I - \frac{1}{4}\mathrm{Lap}(\mathcal{G})$. The eigenvalues of $I - \frac{1}{4}\mathrm{Lap}(\mathcal{G})$ can be expressed as $1 - \lambda/4$, where λ is an eigenvalue of $\mathrm{Lap}(\mathcal{G})$, so $\sigma_2(W) = 1 - \frac{1}{4}\sigma_{N-1}(\mathrm{Lap}(\mathcal{G}))$. Hence $\frac{1}{1-\sigma_2(W)} \leq \frac{5N^2 \cdot 4}{8\pi^2} \leq \frac{N^2}{2}$.

Now suppose that for a subset of $M + 1$ nodes, the local loss functions are identically zero at all times:

$$\ell_t(N - M/2, \cdot) = \cdots = \ell_t(M/2, \cdot) = 0 \quad \forall t \in [T].$$

The remaining nodes update their loss functions every $M + d$ rounds. Specifically, for each $k = 0, \ldots, \lceil T/(M + d) \rceil - 1$,

$$\ell_t(M/2 + 1, \cdot) = \cdots = \ell_t(3M/2 + 1, \cdot) = \ell_k(\cdot) \quad \text{for } t \in [(M + d)k + 1, (M + d)(k + 1)],$$

where $\ell_k(x) = \varepsilon_k L\langle w, x\rangle$, with ε_k being i.i.d. Rademacher random variables (± 1 with probability $1/2$). The vector w is defined as $w = (x_1 - x_2)/\|x_1 - x_2\|_2$, for $x_1, x_2 \in \mathcal{X}$ such that $\|x_1 - x_2\|_2 = D$.

The resulting global loss at time t, observed by agent 0 when it plays x, is:

$$\ell_t(x) = M\ell_{\lceil t/(M+d) \rceil}(x).$$

Due to the structure of the cycle, agent 0 cannot receive information about any node in $\{M/2, \ldots, 3M/2\}$ until at least $M/2+d$ time steps have passed. Thus, predictions $x_{kt+1}(0), \ldots, x_{kt+M+d}(0)$ are made without access to ℓ_k.

Applying the standard lower bound from online learning (Orabona, 2025, Theorem 5.1), we obtain:

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T} \ell_t\big(x_t(0)\big) - \min_{x \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(x)\right] &= (M+1)\mathbb{E}\left[\sum_{k=0}^{\lceil T/(M+d) \rceil - 1} \sum_{t=k(M+d)+1}^{(k+1)(M+d)} \ell_k(x_t(0)) \right. \\
&\qquad \left. - \min_{x \in \mathcal{X}} (M+d) \sum_{k=0}^{\lceil T/(M+d) \rceil - 1} \ell_k(x)\right] \\
&= (M+1)(M+d)\mathbb{E}\left[-\min_{x \in \mathcal{X}} \sum_{k=0}^{\lceil T/(M+d) \rceil - 1} \ell_k(x)\right] \\
&= (M+1)(M+d)L\mathbb{E}\left[\max_{x \in \mathcal{X}} \sum_{k=0}^{\lceil T/(M+d) \rceil - 1} \varepsilon_k \langle w, x\rangle\right]
\end{aligned}$$

$$
\begin{aligned}
&\geq M(M+d)L\mathbb{E}\left[\max_{x\in\{x_1,x_2\}}\sum_{k=0}^{\lceil T/(M+d)\rceil-1}\varepsilon_k\langle w,x\rangle\right]\\
&= M(M+d)LD\mathbb{E}\left[\left|\sum_{k=0}^{\lceil T/(M+d)\rceil-1}\varepsilon_k\right|\right]\\
&\geq M(M+d)LD\sqrt{\frac{T}{M+d}} \quad \text{(Khintchine inequality)}\\
&= MLD\sqrt{(M+d)T}.
\end{aligned}
$$

Thus, there exists a realization of $\varepsilon_0,\ldots,\varepsilon_{\lceil T/(M+d)\rceil-1}$ for which:

$$\mathrm{Reg}_T \geq MLD\sqrt{(M+d)T}.$$

Now, from Equation (D.61), we know:

$$\frac{1}{1-\sigma_2(W)} \leq \frac{N^2}{2}, \text{ so } M \geq \frac{1}{4}N \geq \frac{\frac{1}{4}\sqrt{2}}{\sqrt{1-\sigma_2(W)}}\,.$$

which implies the lower bound:

$$\mathrm{Reg}_T \geq \frac{N}{4}LD\sqrt{T}\sqrt{\frac{1}{2}\sqrt{\frac{1}{1-\sigma_2(W)}}+d}.$$

□

D.3 Omitted Details in Section 5.4

In this section, we include the omitted details in Section "DOCO with Strongly-Convex Loss Functions". For completeness, we first restate the theorem and then present its proof.

Theorem 5.4. *Assume each agent $u \in \mathcal{V}$ runs an instance of AD-FTRL-DF-SC with a valid communication matrix W and parameters θ and B defined in Equation (5.6). Then, under Assumption 5.1, 5.2 and Assumption 5.3, the global regret is bounded as*

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

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 $\bar{d} = \frac{1}{N}\sum_{v\in\mathcal{V}} d(v)$ and the global regret is bounded as

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

Proof. We start the proof with some notations. With a slight abuse of notation, we also define $\bar{z}_{s-1}$ as the the cumulative received augmented gradients till block $s-1$ averaged over all agents in the strongly convex case:

$$\bar{z}_{s-1} = \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}} y_l^+(v). \tag{D.62}$$

Direct calculation shows that

$$\begin{aligned}
\bar{z}_{s-1} &= \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}\left(\sum_{\tau\in o_{lB+1}(v)\setminus o_{(l-1)B+1}(v)} g_\tau(v) - \alpha B x_l(v)\right) && \text{(Definition of } y_l^+(v)\text{)} \\
&= \frac{1}{N}\sum_{v\in\mathcal{V}}\sum_{\tau\in o_{(s-1)B+1}(v)} g_\tau(v) - \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}\alpha B x_l(v),
\end{aligned}$$

where the last inequality is due to $o_1(v) = \emptyset$ for any $v \in \mathcal{V}$. Again with an abuse of notation, similar to the case where the loss functions are convex in general, we define $\bar{x}_s$, $\widetilde{z}_s$, and $\widetilde{x}_s$ in the following. Specifically, we define $\bar{x}_s$, which is the FTRL strategy at block s assuming the agent has the received gradient information among all agent:

$$\bar{x}_s = \arg\min_{x\in\mathcal{X}}\left\{\langle \bar{z}_{s-1}, x\rangle + \frac{\alpha(s-1)B}{2}\|x\|_2^2\right\}. \tag{D.63}$$

We also define $\widetilde{z}_{s-1}$ as follows

$$\widetilde{z}_{s-1} = \frac{1}{N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}\left(\sum_{\tau\in\mathcal{T}_l} g_\tau(v) - \alpha B x_l(v)\right),$$

where $\mathcal{T}_l = \{(l-1)B+1, \dots, lB\}$, and define $\widetilde{x}_s$ to be the FTRL strategy with respect to $\widetilde{z}_{s-1}$:

$$\begin{aligned}\widetilde{x}_s &= \arg\min_{x\in\mathcal{X}} \left\{ \langle \widetilde{z}_{s-1}, x\rangle + \frac{\alpha(s-1)B}{2}\|x\|_2^2 \right\} \\ &= \arg\min_{x\in\mathcal{X}} \left\{ \left\langle \frac{1}{N}\sum_{l=1}^{s-1}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}} g_\tau(v), x \right\rangle + \frac{\alpha B}{2N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}\|x - x_l(v)\|_2^2 \right\}.\end{aligned}$$

Finally, we define

$$G_s(x) \triangleq \left\langle \frac{1}{N}\sum_{l=1}^{s-1}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}} g_\tau(v), x \right\rangle + \psi_s(x) \tag{D.64}$$

where $\psi_s(x)$ is defined as

$$\psi_s(x) \triangleq \frac{\alpha B}{2N}\sum_{l=1}^{s-1}\sum_{v\in\mathcal{V}}\|x - x_l(v)\|_2^2.$$

Next, we apply a regret decomposition that almost mirrors the one in Lemma D.4 except that we use the property that all loss functions are now α-strongly convex.

$$\begin{aligned}\mathrm{Reg}_T(u) &= \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} (\ell_t(v, x_t(u)) - \ell_t(v, x^*)) \\ &= \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} (\ell_t(v, x_t(v)) - \ell_t(v, x^*)) + \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} (\ell_t(v, x_t(u)) - \ell_t(v, x_t(v))) \\ &\le \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), x_t(v) - x^*\rangle - \frac{\alpha}{2}\|x_t(v) - x^*\|_2^2 \right) + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\|x_t(u) - x_t(v)\|_2 \\ &\qquad \text{(Assumption 5.2 and Assumption 5.3)} \\ &\le \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), x_t(v) + \bar{x}_t - \bar{x}_t - x^*\rangle - \frac{\alpha}{2}\|x_t(v) - x^*\|_2^2 \right) \\ &\qquad + L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} (\|x_t(u) - \bar{x}_t\|_2 + \|x_t(v) - \bar{x}_t\|_2) \qquad \text{(Triangular inequality)} \\ &\le \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), \bar{x}_t - x^*\rangle - \frac{\alpha}{2}\|x_t(v) - x^*\|_2^2 \right) \\ &\qquad + 2L\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\|x_t(v) - \bar{x}_t\|_2 + NL\sum_{t=1}^{T}\|x_t(u) - \bar{x}_t\|_2 \qquad \text{(Assumption 5.2)} \\ &= \underbrace{\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} \left(\langle g_t(v), \bar{x}_s - x^*\rangle - \frac{\alpha}{2}\|x_s(v) - x^*\|_2^2 \right)}_{\spadesuit} \\ &\qquad + \underbrace{2BL\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}\|x_s(v) - \bar{x}_s\|_2 + NBL\sum_{s=1}^{T/B}\|x_s(u) - \bar{x}_s\|_2}_{\clubsuit}\end{aligned} \tag{D.65}$$

where the last equality holds because the algorithm uses the same decision over all time steps in the same block, and the block length is B.

We first analyze the term ♠ by decomposing it as follows:

$$\begin{aligned}
\frac{1}{N}\spadesuit &= \frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\Big(\langle g_t(v), \bar{x}_s - x^*\rangle - \frac{\alpha}{2}\|x_s(v) - x^*\|_2^2\Big) \\
&= \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \widetilde{x}_s - x^*\rangle}_{\text{full-info}_T} + \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s - \widetilde{x}_s\rangle}_{\text{drift}_T} \\
&\quad - \frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\frac{\alpha}{2}\|x_s(v) - x^*\|_2^2 \\
&= \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \widetilde{x}_s - x^*\rangle}_{\text{full-info}_T} + \underbrace{\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \bar{x}_s - \widetilde{x}_s\rangle}_{\text{drift}_T} \\
&\quad - \frac{\alpha B}{2N}\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}\|x_s(v) - x^*\|_2^2,
\end{aligned} \tag{D.66}$$

where the last equality is because $|\mathcal{T}_s| = B$. First, we analyze full-info$_T$ by using Lemma 7.1 in Orabona (2025):

$$\begin{aligned}
\text{full-info}_T &= \frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v), \widetilde{x}_s - x^*\rangle \\
&= \sum_{s=1}^{T/B}\left\langle \frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} g_t(v), \widetilde{x}_s - x^*\right\rangle \\
&= \psi_{T/B+1}(x^*) - \min_{x\in\mathcal{X}}\psi_1(x) + \sum_{s=1}^{T/B}\left[G_s(\widetilde{x}_s) - G_{s+1}(\widetilde{x}_{s+1}) + \left\langle \frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} g_t(v), \widetilde{x}_s\right\rangle\right] \\
&\quad + G_{T/B+1}(\widetilde{x}_{T/B+1}) - G_{T/B+1}(x^*) \qquad \text{(Lemma 7.1 in Orabona (2025))} \\
&\le \psi_{T/B+1}(x^*) + \sum_{s=1}^{T/B}\left[\left(G_s(\widetilde{x}_s) + \left\langle \frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} g_t(v), \widetilde{x}_s\right\rangle\right) - \left(G_s(\widetilde{x}_{s+1}) + \left\langle \frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} g_t(v), \widetilde{x}_{s+1}\right\rangle\right)\right] \\
&\quad + \sum_{s=1}^{T/B}(\psi_s(\widetilde{x}_{s+1}) - \psi_{s+1}(\widetilde{x}_{s+1})) \\
&\le \psi_{T/B+1}(x^*) + \sum_{s=1}^{T/B}\left[\left(G_s(\widetilde{x}_s) + \left\langle \frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} g_t(v), \widetilde{x}_s\right\rangle\right) - \left(G_s(\widetilde{x}_{s+1}) + \left\langle \frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}} g_t(v), \widetilde{x}_{s+1}\right\rangle\right)\right],
\end{aligned} \tag{D.67}$$

where the first inequality holds because $G_{T/B+1}(\widetilde{x}_{T/B+1}) \le G_{T/B+1}(x^*)$ by optimality of $\widetilde{x}_{T/B+1}$

and together with non-negativity of ψ_1, and the second inequality holds because

$$\sum_{s=1}^{T/B}\left(\psi_s(\widetilde{x}_{s+1})-\psi_{s+1}(\widetilde{x}_{s+1})\right)=-\frac{\alpha B}{2N}\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}\|\widetilde{x}_{s+1}-x_s(v)\|_2^2\le 0.$$

As for the difference between $\left(G_s(\widetilde{x}_s)+\left\langle\frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}g_t(v),\widetilde{x}_s\right\rangle\right)$ and $\left(G_s(\widetilde{x}_{s+1})+\left\langle\frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}g_t(v),\widetilde{x}_s\right.\right.$ direct calculation shows that

$$\begin{aligned}
&\left(G_s(\widetilde{x}_s)+\left\langle\frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}g_t(v),\widetilde{x}_s\right\rangle\right)-\left(G_s(\widetilde{x}_{s+1})+\left\langle\frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}g_t(v),\widetilde{x}_{s+1}\right\rangle\right)\\
&\le\left\langle\frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}g_t(v),\widetilde{x}_s-\widetilde{x}_{s+1}\right\rangle\qquad\qquad(\text{since } G_s(\widetilde{x}_s)=\min_{x\in\mathcal{X}}G_s(x)\le G_s(\widetilde{x}_{s+1}))\\
&\le BL\|\widetilde{x}_s-\widetilde{x}_{s+1}\|_2.
\end{aligned}$$

To bound $\|\widetilde{x}_s-\widetilde{x}_{s+1}\|_2$, applying Lemma D.1 with $w_1=\frac{1}{N}\sum_{l=1}^{s-1}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}}g_\tau(v)$ and $w_2=\frac{1}{N}\sum_{l=1}^{s}\sum_{\tau\in\mathcal{T}_l}\sum_{v\in\mathcal{V}}g_\tau(v)$, $\psi_1=\psi_s$, $\psi_2=\psi_{s+1}$ shows that

$$\begin{aligned}
\frac{\alpha B(2s-1)}{2}\|\widetilde{x}_{s+1}-\widetilde{x}_s\|_2^2&\le\left\langle\frac{1}{N}\sum_{\tau\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}g_\tau(v),\widetilde{x}_s-\widetilde{x}_{s+1}\right\rangle-\frac{\alpha B}{2N}\sum_{v\in\mathcal{V}}\|\widetilde{x}_{s+1}-x_s(v)\|_2^2+\frac{\alpha B}{2N}\sum_{v\in\mathcal{V}}\|\widetilde{x}_s-x_s(v)\|_2^2\\
&\le BL\|\widetilde{x}_s-\widetilde{x}_{s+1}\|_2+\frac{\alpha B}{2N}\|\widetilde{x}_s-\widetilde{x}_{s+1}\|_2\cdot\|\widetilde{x}_s+\widetilde{x}_{s+1}-x_s(v)\|_2\\
&\le BL\|\widetilde{x}_s-\widetilde{x}_{s+1}\|_2+\alpha BD\|\widetilde{x}_s-\widetilde{x}_{s+1}\|_2.
\end{aligned}$$

Rearranging the terms leads to

$$\|\widetilde{x}_s-\widetilde{x}_{s+1}\|_2\le\frac{2}{\alpha(2s-1)}L+\frac{2D}{2s-1}.$$

Plugging the above into Equation (D.67), we obtain

$$\text{full-info}_T\le\psi_{T/B+1}(x^*)+\frac{2BL(L+\alpha D)}{\alpha}\sum_{s=1}^{T/B}\frac{1}{2s-1}\le\psi_{T/B+1}(x^*)+\frac{2BL(L+\alpha D)}{\alpha}\ln(2T/B)$$

Combining with the negative term in Equation (D.67), we know that

$$\text{full-info}_T-\frac{\alpha B}{2N}\sum_{s=1}^{T/B}\sum_{v\in\mathcal{V}}\|x_s(v)-x^*\|_2^2\le\frac{2BL(L+\alpha D)}{\alpha}\ln(2T/B).\tag{D.68}$$

Now we turn to analyze drift_T in the term ♠.

$$\begin{aligned}
\frac{1}{N}\text{drift}_T&=\frac{1}{N}\sum_{s=1}^{T/B}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}\langle g_t(v),\bar{x}_s-\widetilde{x}_s\rangle\\
&=\sum_{s=1}^{T/B}\left\langle\frac{1}{N}\sum_{t\in\mathcal{T}_s}\sum_{v\in\mathcal{V}}g_t(v),\bar{x}_s-\widetilde{x}_s\right\rangle
\end{aligned}$$

$$
\begin{aligned}
&\leq BL \sum_{s=1}^{T/B} \|\bar{x}_s - \widetilde{x}_s\|_2 && \text{(Cauchy–Schwarz inequality, Assumption 5.2 and } |\mathcal{T}_s| = B) \\
&= BL \sum_{s=2}^{T/B} \|\bar{x}_s - \widetilde{x}_s\|_2 && (\bar{x}_1 = \widetilde{x}_1 = \mathbf{0}) \\
&\leq BL \sum_{s=2}^{T/B} \frac{1}{(s-1)B\alpha} \|\bar{z}_{s-1} - \widetilde{z}_{t-1}\|_2 && \text{(Lemma D.1)} \\
&\leq \frac{L}{\alpha} \sum_{s=2}^{T/B} \frac{1}{s-1} \left\| \frac{1}{N} \sum_{v \in \mathcal{V}} \sum_{\tau \in o_{(s-1)B+1}(v)} g_\tau(v) - \frac{1}{N} \sum_{l=1}^{s-1} \sum_{\tau \in \mathcal{T}_l} \sum_{v \in \mathcal{V}} g_\tau(v) \right\|_2 \\
& && \text{(definition of } \bar{z}_{t-1} \text{ and } \widetilde{z}_{t-1}) \\
&= \frac{L}{\alpha} \sum_{s=2}^{T/B} \frac{1}{s-1} \left\| -\frac{1}{N} \sum_{v \in \mathcal{V}} \sum_{\tau \in m_{(s-1)B+1}(v)} g_\tau(v) \right\|_2 \\
& && (\mathcal{T}_s = \{(s-1)B+1, \ldots, sB\} \text{ and } m_t(v) = [t-1] \backslash o_t(v)) \\
&\leq \frac{L^2}{\alpha} \sum_{s=2}^{T/B} \frac{1}{s-1} \left(\frac{1}{N} \sum_{v \in \mathcal{V}} |m_{(s-1)B+1}(v)| \right) && \text{(Assumption 5.2)} \\
&\leq \frac{\delta_{\max} L^2}{\alpha} \sum_{s=2}^{T/B} \frac{1}{s-1} && (\delta_{\max} = \max_{t \in [T]} \tfrac{1}{N} \textstyle\sum_{u \in \mathcal{V}} |m_t(u)|) \\
&\leq \frac{\delta_{\max} L^2}{\alpha} \left(\ln(T/B) + 1 \right), && \text{(D.69)}
\end{aligned}
$$

where the second inequality applies Lemma D.1 using the definition of $\bar{x}_s$ and $\widetilde{x}_s$, and the last inequality is due to $\sum_{s=2}^{T/B} \frac{1}{s-1} \leq \ln(T/B) + 1$. When $d_t(u) = d(u)$ for all $u \in [N]$, we can further upper bound $\frac{1}{N} \sum_{v \in \mathcal{V}} |m_{(s-1)B+1}(v)|$ by $\frac{1}{N} \sum_{u \in \mathcal{V}} d(u) \triangleq \bar{d}$, leading to

$$
\frac{1}{N} \operatorname{drift}_T \leq \frac{\bar{d} L^2}{\alpha} (\ln(T/B) + 1). \tag{D.70}
$$

Combining Equation (D.66), Equation (D.68) and Equation (D.69), we have

$$
\frac{1}{N} \spadesuit \leq \frac{\delta_{max} L^2}{\alpha} \left(\ln \left(T/B \right) + 1 \right) + \frac{2BL(L + \alpha D)}{\alpha} \ln(2T/B). \tag{D.71}
$$

We now turn to the analysis of the term ♣. By definition, $x_1(v) = \bar{x}_1 = 0$, which implies $\|x_1(v) - \bar{x}_1\|_2 = 0$. To bound $\|x_{s+1}(u) - \bar{x}_{s+1}\|_2$ for any $s \geq 1$ and $u \in \mathcal{V}$, we proceed as follows.

$$
\begin{aligned}
&\sum_{s=1}^{T/B-1} \|x_{s+1}(u) - \bar{x}_{s+1}\|_2 && \text{(D.72)} \\
&\leq \sum_{s=1}^{T/B-1} \frac{1}{\alpha s B} \|z_s(u) - \bar{z}_s\|_2 && \text{(Lemma D.1)} \\
&= \frac{2}{\sqrt{N}} \sum_{s=1}^{T/B-1} \frac{1}{\alpha s B} \sum_{l=1}^{s-1} b^{(s-l-1)B} \left(\sqrt{\sum_{v \in \mathcal{V}} \left\| y_l^+(v) \right\|_2^2} \right) && \text{(Lemma D.9)}
\end{aligned}
$$

$$\leq \frac{2}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\sqrt{\sum_{v\in\mathcal{V}} \left\|y_l^+(v)\right\|_2^2} \sum_{s=l+1}^{T/B} \frac{1}{\alpha s B} b^{(s-l-1)B} \right) \qquad \text{(swap the summation order)}$$

$$\leq \frac{2}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\frac{1}{\alpha(l+1)B} \sqrt{\sum_{v\in\mathcal{V}} \left\|y_l^+(v)\right\|_2^2} \sum_{s=l+1}^{T/B} b^{(s-l-1)B} \right)$$

$$\leq \frac{2}{\sqrt{N}} \frac{1}{1-\frac{1}{\sqrt{14N}}} \sum_{l=1}^{T/B-1} \left(\frac{1}{\alpha(l+1)B} \sqrt{\sum_{v\in\mathcal{V}} \left\|y_l^+(v)\right\|_2^2} \right) \qquad \text{(Equation (D.31) and Geometric sum)}$$

$$\leq \frac{3}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\frac{1}{\alpha(l+1)B} \sqrt{\sum_{v\in\mathcal{V}} \left\|y_l^+(v)\right\|_2^2} \right), \tag{D.73}$$

where the last inequality follows from $N \geq 1$. Plugging in the definition of $y_l^+(v)$ in Equation (D.73), we obtain that

$$\sum_{s=1}^{T/B-1} \|x_{s+1}(u) - \bar{x}_{s+1}\|_2 \tag{D.74}$$

$$\leq \frac{3}{\sqrt{N}} \sum_{s=1}^{T/B-1} \left(\frac{1}{\alpha(s+1)B} \sqrt{\sum_{v\in\mathcal{V}} \left\|y_s^+(v)\right\|_2^2} \right) \tag{D.75}$$

$$= \frac{3}{\sqrt{N}} \sum_{s=1}^{T/B-1} \left(\frac{1}{\alpha(s+1)B} \sqrt{\sum_{v\in\mathcal{V}} \left\| \sum_{\tau\in o_{sB+1}(v)\setminus o_{(s-1)B+1}(v)} g_\tau(v) - \alpha B x_s(u) \right\|_2^2} \right)$$

$$\leq \frac{3}{\sqrt{N}} \sum_{s=1}^{T/B-1} \left(\frac{L}{\alpha(s+1)B} \sqrt{\sum_{v\in\mathcal{V}} \left(|o_{sB+1}(v)| - |o_{(s-1)B+1}(v)|\right)^2} \right) + \frac{3}{\sqrt{N}} \sum_{s=1}^{T/B-1} \left(\frac{D}{(s+1)B\alpha} \sqrt{N\alpha^2 B^2} \right)$$

(Triangular inequality)

$$\leq \frac{3}{\sqrt{N}} \sum_{s=1}^{T/B-1} \left(\frac{L}{(s+1)B\alpha} \sqrt{\sum_{v\in\mathcal{V}} \left(|o_{sB+1}(v)| - |o_{(s-1)B+1}(v)|\right)^2} \right) + 3D\left(\ln\left(T/B\right) + 1\right), \tag{D.76}$$

where the last inequality is due to $\sum_{z=1}^{T/B-1} \frac{1}{s+1} \leq \ln(T/B) + 1$. by definition of $o_t(v)$, we observe that

$$\begin{aligned} |o_{sB+1}(v)| - |o_{(s-1)B+1}(v)| &= |sB| - |m_{sB+1}(v)| - \left(|(s-1)B| - |m_{(s-1)B+1}(v)|\right) \\ &= B + |m_{(s-1)B+1}(v)| - |m_{sB+1}(v)|. \end{aligned}$$

Hence, we obtain

$$\begin{aligned} \sqrt{\sum_{v\in\mathcal{V}} \left(|o_{sB+1}(v)| - |o_{(s-1)B+1}(v)|\right)^2} &\leq \sqrt{\sum_{v\in\mathcal{V}} \left(B + |m_{(s-1)B+1}(v)| - |m_{sB+1}(v)|\right)^2} \\ &\leq \sqrt{\sum_{v\in\mathcal{V}} B^2} + \sqrt{\sum_{v\in\mathcal{V}} \left(|m_{(s-1)B+1}(v)| - |m_{sB+1}(v)|\right)^2} \end{aligned}$$

(triangle inequality)

$$\leq B\sqrt{N} + \sum_{v\in\mathcal{V}} |m_{(s-1)B+1}(v)| + \sum_{v\in\mathcal{V}} |m_{sB+1}(v)|$$
$$\leq B\sqrt{N} + 2N\delta_{\max}, \tag{D.77}$$

where the last inequality is due to the definition of $\delta_{\max}$. Combining Equation (D.76) and Equation (D.77) and using $\sum_{z=1}^{T/B-1} \frac{1}{s} \leq \ln(T/B)+1$, we obtain

$$\sum_{s=1}^{T/B-1} \|x_{s+1}(u) - \bar{x}_{s+1}\|_2 \leq \frac{3(\alpha D + L)}{\alpha} (\ln(T/B)+1) + \frac{6\sqrt{N}\delta_{\max} L}{\alpha B} (\ln(T/B)+1). \tag{D.78}$$

Furthermore, when $d_t(u) = d(u)$ for all $t \in [T]$, due to the definition of $y_l^+(v)$, each agent $v \in \mathcal{V}$ can receive at most B gradients and actions in any block $s \in T/B$. Hence, we obtain

$$\sum_{s=2}^{T/B} \|x_s(v) - \bar{x}_s\|_2 \leq \frac{3}{\sqrt{N}} \sum_{l=1}^{T/B-1} \left(\frac{1}{(l+1)sB} \sqrt{\sum_{v\in\mathcal{V}} \left\|y_l^+(v)\right\|_2^2} \right) \qquad \text{(Equation (D.73))}$$
$$\leq \frac{3(\alpha D + L)}{\alpha} (\ln(T/B)+1). \tag{D.79}$$

Finally, we obtain

$$\begin{aligned}
\mathrm{Reg}_T(u) &= \sum_{s=1}^{T/B} \sum_{t\in\mathcal{T}_s} \sum_{v\in\mathcal{V}} \left(\langle g_t(v), \bar{x}_s - x^* \rangle - \frac{\alpha}{2} \|x_s(v) - x^*\|_2^2 \right) \\
&\quad + 2BL \sum_{s=1}^{T/B} \sum_{v\in\mathcal{V}} \|x_s(v) - \bar{x}_s\|_2 + NBL \sum_{s=1}^{T/B} \|x_s(u) - \bar{x}_s\|_2 \qquad \text{(Equation (D.65))} \\
&\leq \frac{N\delta_{max} L^2}{\alpha} (\ln(T/B)+1) + \frac{2NBL(L+\alpha D)}{\alpha} \ln(2T/B) \\
&\quad + 2BL \sum_{s=1}^{T/B} \sum_{v\in\mathcal{V}} \|x_s(v) - \bar{x}_s\|_2 + NBL \sum_{s=1}^{T/B} \|x_s(u) - \bar{x}_s\|_2 \qquad \text{(Equation (D.71))} \\
&\leq \frac{N\delta_{max} L^2}{\alpha} (\ln(T/B)+1) + \frac{2NBL(L+\alpha D)}{\alpha} \ln(2T/B) \\
&\quad + \frac{9BN(\alpha DL + L^2)}{\alpha} (\ln(T/B)+1) + \frac{18N\sqrt{N}\delta_{\max} L^2}{\alpha} (\ln(T/B)+1) \\
&\qquad\qquad \text{(Equation (D.78))} \\
&= \mathcal{O}\left(\frac{N(\alpha DL + L^2)}{\alpha} \left(\sqrt{N}\delta_{\max} + \frac{\ln(N)}{\sqrt{1-\sigma_2(W)}} \right) \ln(T) \right).
\end{aligned}$$

When $d_t(u) = d(u)$ for all $t \in [T]$, we define $\bar{d} = \frac{1}{N}\sum_{v\in\mathcal{V}} d(v)$, then we have $\bar{d} = \delta_{\max}$. Combining Equation (D.65), Equation (D.68), Equation (D.70), and Equation (D.79), we obtain

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

□

D.3.1 Lower bound for the strongly convex case

In this section, we provide the proof for the lower bound for the strongly convex case.

Theorem D.1. *Let d be a constant feedback delay experienced by each agent in the network, and let A be any algorithm for DOCO over the domain $\mathcal{X}$. Then, there exists a graph $\mathcal{G} = ([0, N-1], \mathcal{E})$, with $N = 2M+1$ where M is an even integer and $16(N+d)+1 \leq T$, and a sequence of αD-Lipschitz and α-strongly convex loss functions assigned to the agents, denoted by*

$$\{\ell_1(0,\cdot),\dots,\ell_1(N,\cdot)\},\dots,\{\ell_t(0,\cdot),\dots,\ell_t(N,\cdot)\},$$

such that the regret of algorithm A satisfies the lower bound:

$$\mathrm{Reg}_T = \Omega\left(\alpha N D^2\left(\frac{1}{\sqrt{1-\sigma_2(W)}}+d\right)\ln\left(\frac{T}{\frac{1}{\sqrt{1-\sigma_2(W)}}+d}\right)\right),$$

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

Proof. This proof is an adaptation of that of Wan et al. (2024c, Theorem 4). Specifically, we consider the setting where all delays are fixed and equal to d, and we let $\mathcal{G}$ denote a cycle graph with $N = 2(M+1)$ nodes where M is even, to simplify, and $\mathcal{V} = \{0,1,2,\dots,N-1\}$.

For any DOCO algorithm A, we denote the sequence of decisions made by **agent 0** as $x_0(0),\dots,x_T(0)$. We divide the total T rounds into the following $Z+1$ blocks:

$$[c_0+1, c_1], [c_1+1, c_2], \dots, [c_Z+1, c_{Z+1}] \tag{D.80}$$

where $Z = \lfloor (T-1)/\tau \rfloor$, $\tau = M/2+d$, $c_{Z+1} = T$, and $c_i = i\tau$ for $i = 0,\dots,Z$.

At each round t, we set:

$$\ell_t(u,x) = \frac{\alpha}{2}\|x\|_2^2 \quad \text{for } u \in \{0,\dots,M/2\}\cup\{N-M/2,\dots,N-1\},$$

which is α-strongly convex and satisfies Assumption 5.2 with $L = \alpha D$ over the set $\mathcal{X} = [0, D/\sqrt{n}]^n$.

Let $\mathcal{B}_p$ denote the Bernoulli distribution with success probability p and $\mathcal{B}_p^n$ the distribution of vectors whose coordinates are equal to each other and the value is drawn from $\mathcal{B}_p$. For any $i \in \{0,\dots,Z\}$ and $t \in [c_i+1, c_{i+1}]$, define:

$$\ell_t(u,x) = \ell_i(x) = \frac{\alpha}{2}\left\|x - \frac{D\mathbf{w}_i}{\sqrt{n}}\right\|_2^2, \quad \text{for } u \in \{M/2+1,\dots,3M/2+1\},$$

where $\mathbf{w}_i \in \{\mathbf{0},\mathbf{1}\}$ is sampled from $\mathcal{B}_p^n$, meaning that with probability p, $\mathbf{w}_i = \mathbf{1}$; otherwise, $\mathbf{w}_i = \mathbf{0}$.

Then, the global loss function at time t is:

$$\begin{aligned}\ell_t(x) &= \sum_{u=0}^{N-1}\ell_t(u,x)\\ &= \frac{\alpha(M+1)}{2}\left\|x - \frac{D\mathbf{w}_i}{\sqrt{n}}\right\|_2^2 + \frac{\alpha(M+1)}{2}\|x\|_2^2\end{aligned}$$

$$= \frac{\alpha N}{2}\|x\|_2^2 - \frac{\alpha(M+1)D}{\sqrt{n}}\langle x, \mathbf{w}_i\rangle + \frac{\alpha(M+1)D^2}{2n}\|\mathbf{w}_i\|_2^2.$$

Taking expectation, we obtain that

$$\begin{aligned}\mathbb{E}_{\mathbf{w}_i}[\ell_t(x)] &= \frac{\alpha N}{2}\|x\|_2^2 + \frac{\alpha(M+1)D}{\sqrt{n}}\langle x, \mathbf{p}\rangle + \frac{\alpha(M+1)D^2}{2n}\langle \mathbf{1}, \mathbf{p}\rangle \\ &= \frac{\alpha N}{2}\left\|x - \frac{(M+1)D\mathbf{p}}{N\sqrt{n}}\right\|_2^2 + \frac{\alpha(M+1)D^2}{2n}\left\langle \mathbf{1} - \frac{(M+1)\mathbf{p}}{N}, \mathbf{p}\right\rangle,\end{aligned}$$

where $\mathbf{p} = p \cdot \mathbf{1}$. Let $F(x) \triangleq \mathbb{E}_{\mathbf{w}_i}[\ell_t(x)]$. Then, direct calculation shows that the minimizer of $F(x)$ has the following form:

$$x^* = \frac{(M+1)D \cdot \mathbf{p}}{N\sqrt{n}}.$$

and for any $x \in \mathcal{X}$, we have

$$F(x) - F(x^*) = \frac{\alpha N}{2}\left\|x - \frac{(M+1)D\mathbf{p}}{N\sqrt{n}}\right\|_2^2 \geq 0. \tag{D.81}$$

Moreover, according to Jensen's inequality, we have

$$\mathbb{E}_{\mathbf{w}_0,\ldots,\mathbf{w}_Z}\left[\min_{x\in\mathcal{X}}\sum_{i=0}^{Z}\sum_{t=c_i+1}^{c_{i+1}}\ell_t(x)\right] \leq \sum_{i=0}^{Z}\sum_{t=c_i+1}^{c_{i+1}}F(x^*). \tag{D.82}$$

Because of the feedback delay d and the delay $M/2+1$ induced by communication in the graph, the decisions $x_{c_i+1}(0), \ldots, x_{c_{i+1}}(0)$ are independent of $\mathbf{w}_i$. Thus:

$$\begin{aligned}\mathbb{E}_{\mathbf{w}_0,\ldots,\mathbf{w}_Z}[\mathrm{Reg}_T(0)] &= \mathbb{E}\left[\sum_{i=0}^{Z}\sum_{t=c_i+1}^{c_{i+1}}\ell_t(x_t(0)) - \min_{x\in\mathcal{X}}\sum_{i=0}^{Z}\sum_{t=c_i+1}^{c_{i+1}}\ell_t(x)\right] \\ &= \sum_{i=0}^{Z}\sum_{t=c_i+1}^{c_{i+1}}\left(\mathbb{E}[F(x_t(0))] - F(x^*)\right) \\ &\geq \sum_{i=0}^{Z}\sum_{t=c_i+1}^{c_{i+1}}\mathbb{E}[F(x_t(0)) - F(x^*)].\end{aligned} \tag{D.83}$$

To achieve a lower bound on (D.83) , we assume without loss of generality that the DOCO algorithm is deterministic.* Recall that given $i \in \{0, 1, 2, \ldots, Z\}$, for each round $t \in [c_i + 1, c_{i+1}]$, all local functions $\{\ell_t(0, x), \ell_t(1, x), \ldots, \ell_t(N, x)\}$ are jointly dependent on the same random vector $\mathbf{w}_i \in \{\mathbf{0}, \mathbf{1}\}$ sampled from the Bernoulli distribution $\mathcal{B}_p^n$. Consequently, the decision $x_t(0)$ made by agent 0 at time $t \in [c_i + 1, c_{i+1}]$ can be expressed as a deterministic function of a sequence $X \in \{\mathbf{0}, \mathbf{1}\}^i$, where X is sampled from $(\mathcal{B}_p^n)^i$, where $(\mathcal{B}_p^n)^i$ represents the joint probability law of i independent draws from $\mathcal{B}_p^n$ (used to sample the $(\mathbf{w}_j)_{j\leq i}$). That is, $x_t(0) = \mathcal{A}_t(X)$ for some mapping $\mathcal{A}_t : \{0, 1\}^i \to \mathcal{X}$. Similarly, if we replace p with another value p', the corresponding random vectors $\mathbf{w}'_0, \ldots, \mathbf{w}'_Z$ yield a new decision sequence $x'_1(0), \ldots, x'_T(0)$. For each $t \in [c_i + 1, c_{i+1}]$, the decision

*This reduction is also used in Wan et al. (2024c) and dates back to Hazan and Kale (2014). Specifically, the analysis can be directly generalized to randomized algorithm as discussed in Footnote 3 of Wan et al. (2024c).

under this new environment is then given by $x'_t(0) = \mathcal{A}_t(X')$, where $X' \in \{0,1\}^i$ is drawn from $(\mathcal{B}^n_{p'})^i$. Following the argument of Hazan et al. (2007), later re-used in Wan et al. (2024c), we will show that for appropriately chosen values of p and p', the expected regret incurred by agent 0 under at least one of the two distributions must be large.

Lemma D.11 (Lemma 7 in Wan et al. (2024c).)**.** *Fix a block i and let $\epsilon \leq \frac{1}{32\sqrt{i+1}}$ be a parameter. Assume that $p, p' \in \left[\frac{1}{4}, \frac{3}{4}\right]$ such that $2\epsilon \leq |p - p'| \leq 4\epsilon$. Following the above notations, for any $t \in [c_i + 1, c_{i+1}]$, we have:*

$$\mathbb{E}_X\left[\|\mathcal{A}_t(X) - \xi\mathbf{p}\|_2^2\right] + \mathbb{E}_{X'}\left[\left\|\mathcal{A}_t(X') - \xi\mathbf{p}'\right\|_2^2\right] \geq \frac{n(\xi\epsilon)^2}{4},$$

where $\xi = (M+1)D/(N\sqrt{n})$ and $\mathbf{p}' = p'\mathbf{1}$.

Let $K = \lfloor \log_{16}(15Z + 16) - 1 \rfloor$. Assuming that T is sufficiently large such that $16(N+d)+1 \leq T$, we know that $K \geq 1$. To leverage the lemma above, we partition the first $Z' = \frac{1}{15}(16^{K+1} - 16)$ blocks into K epochs, where the k-th epoch spans $r_k = 16^k$ blocks for $k = 1, 2, \ldots, K$. Specifically, epoch k corresponds to the block indices:

$$E_k = \left\{\frac{1}{15}(16^k - 16), \ldots, \frac{1}{15}(16^{k+1} - 16) - 1\right\}.$$

This means that epoch k covers the time steps between $c_{\frac{1}{15}(16^k-16)} + 1$ and $c_{\frac{1}{15}(16^{k+1}-16)}$.

The following lemma proven in Wan et al. (2024c) shows that without the knowledge of the true environment, any algorithm will suffer from a non-trivial gap to the optimal solution $\xi\mathbf{p}$.

Lemma D.12 (Lemma 8 in Wan et al. (2024c))**.** *Following the notation used in Lemma D.11, there exists a collection of nested intervals $\left[\frac{1}{4}, \frac{3}{4}\right] \supseteq I_1 \supseteq I_2 \supseteq \cdots \supseteq I_K$ such that the length of the k-th interval equals to $|I_k| = 4^{-(k+3)}$ and for every $p \in I_k$,*

$$\mathbb{E}_X\left[\|\mathcal{A}_t(X) - \xi\mathbf{p}\|_2^2\right] \geq \frac{16^{-(k+3)}n\xi^2}{8}$$

holds for at least half the rounds t in epoch k.

From Lemma D.12, there exists $p \in \cap_{k=1}^K I_k$ such that:

$$\begin{aligned}
\mathbb{E}_{\mathbf{w}_0,\ldots,\mathbf{w}_Z}[\mathrm{Reg}_T(0)] &\geq \sum_{i=0}^{Z}\sum_{t=c_i+1}^{c_{i+1}} \frac{\alpha N}{2}\left\|x_t(0) - \frac{(M+1)D\mathbf{p}}{N\sqrt{n}}\right\|_2^2 \\
&\geq \sum_{k=1}^{K}\sum_{i\in E_k}\sum_{t=c_i+1}^{c_{i+1}} \mathbb{E}_X\left[\frac{\alpha N}{2}\left\|\mathcal{A}_t(X) - \frac{(M+1)D\mathbf{p}}{N\sqrt{n}}\right\|_2^2\right] \\
&\geq \sum_{k=1}^{K} \frac{\left(c_{\frac{1}{15}(16^{k+1}-16)} - c_{\frac{1}{15}(16^k-16)}\right)16^{-(k+3)}\alpha(M+1)^2D^2}{64N} \\
&= \frac{16^{-4}\alpha K\tau(M+1)^2D^2}{4N} \\
&\geq \frac{16^{-4}\alpha K\tau M^2D^2}{4N}.
\end{aligned} \tag{D.84}$$

From the definitions of K, Z, τ, we get:

$$\begin{aligned}\frac{K\tau M^2}{4N} &\geq \frac{(\log_{16}(15(T-1)/(M/2+d))-2)(M/2+d)M^2}{4N} \\ &\geq \frac{(\log_{16}(15(T-1)/(M/2+d))-2)(M/2+d)M}{16} \\ &\geq \frac{(\log_{16}(15(T-1)/(N+d))-2)(N+d)N}{16^2}. \end{aligned} \tag{D.85}$$

Recall that for the N-cycle graph, we have $\frac{1}{1-\sigma_2(W)} \leq \frac{N^2}{2}$ as established in Equation (D.61). Moreover, the second-smallest eigenvalue of the Laplacian satisfies $\sigma_{N-1}(\mathrm{Lap}(\mathcal{G})) = 2 - 2\cos\left(\frac{2\pi}{N}\right) \leq \frac{4\pi^2}{N^2} \leq \frac{40}{N^2}$, and the largest eigenvalue is $\sigma_1(\mathrm{Lap}(\mathcal{G})) = 4$. Since $W = I - \frac{1}{4}\mathrm{Lap}(\mathcal{G})$, it follows that $\sigma_2(W) = 1 - \frac{1}{4}\sigma_{N-1}(\mathrm{Lap}(\mathcal{G}))$, so $\frac{1}{1-\sigma_2(W)} \geq \frac{N^2}{10}$. Combining this estimate with Equation (D.84) and Equation (D.85), we conclude that for some realization of $\mathbf{w}_0, \ldots, \mathbf{w}_Z$,

$$\mathrm{Reg}_T(0) \geq 16^{-6}\alpha N D^2 \left(\frac{\sqrt{2}}{\sqrt{1-\sigma_2(W)}} + d\right) \log_{16}\left(\frac{15(T-1)}{\frac{\sqrt{10}}{\sqrt{1-\sigma_2(W)}} + d} - 2\right). \tag{D.86}$$

□

Appendix E

Proof Details for Chapter 6

E.1 Auxiliary results

In this section, we show several auxiliary lemmas that will be helpful.

The following lemma is the concentration bound for the estimation of the observed expected rewards,

Lemma E.1. *Let $\widehat{\mu}_{i,k}(t)$ be the observed empirical average of the expected reward up to the end of round $t-1$. Then,*

$$\Pr\left[|\widehat{\mu}_{i,k}(t) - \mu_{i,k}| > \sqrt{\frac{2\log T}{T_{i,k}(t)}}\right] \leq \frac{2}{T^2}$$

for $i \in [N],\ k \in [K]$ and $t \in [T]$ holds.

Proof of Lemma E.1. This lemma follows immediately from Hoeffding's inequality and a union bound. □

The following lemma is the concentration bound for the estimation of the observed expected rewards.

Lemma E.2. *Let $\widehat{\mu}_{i,k}(t)$ be the observed empirical average of the expected reward for agent i pulling arm k. Then,*

$$\mathbb{P}\left[\left|\sum_{i\in[N]} \widehat{\mu}_{i,k} - \sum_{i\in[N]} \mu_{i,k}\right| > \sqrt{\frac{4N\log T}{\min_{i\in[N]}\{T_{i,k}(t)\}}}\right] \leq \frac{2}{T^2}$$

holds for any $i \in [N], k \in [K]$ and $t \in [T]$.

Proof of Lemma E.2. We sample $T_{i,k}(t)$ number of random variables iid from $\mathbb{P}_{i,k}$ for each $i \in [N]$, each taking values in $[0,1]$, and hence $X_{i,k}(\tau) - \mu_{i,k}$ is 1-subgaussian for each τ.

According to the property of subgaussian variables, we can obtain that

$$\sum_{i\in[N]} (\widehat{\mu}_{i,k}(t) - \mu_{i,k}) \text{ is } \left(\sum_{i\in[N]} \frac{1}{T_{i,k}(t)}\right)^{1/2} \text{- subgaussian.}$$

Moreover, note that $\sum_{i\in[N]} \frac{1}{T_{i,k}(t)} \leq \frac{N}{\min_{i\in[N]} T_{i,k}(t)}$, then by applying Chernoff's bound, we have

$$\mathbb{P}\left(\left|\sum_{i\in[N]} \widehat{\mu}_{i,k}(t) - \mu_{i,k}\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2 \min_{i\in[N]}\{T_{i,k}(t)\}}{2N}\right).$$

Taking $\epsilon = \sqrt{\frac{4N\log T}{\min_{i\in[N]}\{T_{i,k}(t)\}}}$, we obtain that

$$\mathbb{P}\left(\left|\sum_{i\in[N]} \widehat{\mu}_{i,k}(t) - \mu_{i,k}\right| \geq \sqrt{\frac{4N\log T}{\min_{i\in[N]}\{T_{i,k}(t)\}}}\right) \leq \frac{2}{T^2}$$

which ends the proof of Lemma E.1 □

We provide lemmas for the convergence bound of randomised gossip algorithms. Similar proof could be found Lei et al. (2020), Achddou et al. (2024).

Lemma E.3 (Random Graph)**.** *Let the communication protocol based on random graphs defined in Assumption 6.1 hold. For $t = 1 \dots T$, W_t is doubly stochastic matrix and symmetric and i.i.d. $\forall v \in V$, $\forall s, t \in [T]$ such that $t > s$,*

$$\Pr\left(\left\|W_t \cdots W_{s+1} e_v - \frac{1}{N}\mathbf{1}\right\|_2 \geq \delta\right) \leq \frac{\lambda_2(\mathbb{E}[W^2])^{t-s}}{\delta^2}.$$

When $t - s \geq \left\lceil \frac{3\log(T)}{\log \lambda_2(\mathbb{E}[W^2])^{-1}} \right\rceil = \tau'$, we have

$$\Pr\left(\left\|W_t \cdots W_{s+1} e_v - \frac{1}{N}\mathbf{1}\right\|_2 \geq \delta\right) \leq \delta. \tag{E.1}$$

Furthermore, when $t - s \geq \tau^* = \left\lceil \frac{3N\log(T)}{p\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))} \right\rceil \geq \tau'$, Equation (E.1) *still holds.*

Proof. Using Markov's inequality we have

$$\Pr\left(\left\|W_t \cdots W_{s+1} e_v - \frac{1}{N}\mathbf{1}\right\|_2 \geq \delta\right) \leq \frac{\mathbb{E}\left(\left\|W_t \cdots W_{s+1} e_v - \frac{1}{N}\mathbf{1}\right\|_2^2\right)}{\delta^2}.$$

Let $\widetilde{W}_k = W_k - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$ and assume

$$\mathbb{E}\left[\left\|W_{k-1} \cdots W_{s+1} e_v - \frac{\mathbf{1}}{N}\right\|_2^2\right] \leq e_v^T e_v \left\|\mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right\|_{\mathrm{op}}^{k-s-1} \qquad \text{for some } k-1 > s.$$

Let $\mathcal{F}_{k-1}$ be the σ-algebra generated by all random events up to time $k-1$. We have that

$$\begin{aligned}
\mathbb{E}\left[\left\|W_k^\top \cdots W_{s+1}^\top e_v - \frac{\mathbf{1}}{N}\right\|_2^2\right] &= \mathbb{E}\left[e_v^T \widetilde{W}_{s+1}^\top \cdots \widetilde{W}_{k-1}^\top \widetilde{W}_k^\top \widetilde{W}_k \widetilde{W}_{k-1} \cdots \widetilde{W}_{s+1} e_v\right] \\
&= \mathbb{E}\left[e_v^T \widetilde{W}_{s+1}^\top \cdots \widetilde{W}_{k-1}^\top \mathbb{E}[\widetilde{W}_k^\top \widetilde{W}_k \mid \mathcal{F}_{k-1}] \widetilde{W}_{k-1} \cdots \widetilde{W}_{s+1} e_v\right] \\
&= \mathbb{E}\left[e_v^T \widetilde{W}_{s+1}^\top \cdots \widetilde{W}_{k-1}^\top \mathbb{E}[\widetilde{W}_1^\top \widetilde{W}_1] \widetilde{W}_{k-1} \cdots \widetilde{W}_{s+1} e_v\right] \\
&\qquad \text{(by independence of } W_k\text{)}
\end{aligned}$$

$$\leq \left\|\mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right\|_{\text{op}} e_v^T e_v \left\|\mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right\|_{\text{op}}^{k-s-1}$$
$$\leq \lambda_2(\mathbb{E}[W^2])^{t-s} e_v^T e_v \left\|\mathbb{E}[W_1 W_1^\top] - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right\|_{\text{op}}^{k-s-1}$$

which by induction, suffices to prove the lemma. Moreover, the proof of $\tau^* \geq \tau'$ follows from the fact that

$$\frac{1}{log\lambda_2^{-1}} \leq \frac{1}{1-\lambda_2}, \lambda_2(\mathbb{E}[W^2]) \leq 1 - \frac{p\lambda_{N-1}(\text{Lap}(\mathcal{G}))}{N} \text{ and } \frac{1}{\log(1-p)^{-1}} \leq \frac{1}{p},$$

where the second inequality is taken from Theorem 6.1 in Achddou et al. (2024). □

The following lemmas guarantee local consistency between agents.

Lemma E.4. *Let assumption 6.1 hold. With probability $1 - N^2T\delta$, Algorithm 6.1 guarantees that for fixed arm $k \in [K]$, and for every $i, j \in [N]$ and for every $t \in [T]$,*

$$|T_{i,k}(t) - T_{j,k}(t)| \leq KNL_p(\delta)$$

where $L_p(\delta) = \left\lceil \frac{\log(\delta)}{\log(1-p)} \right\rceil$ denotes the maximum number of rounds each edge within base graph $\mathcal{G}$ is connected in the communication graph $\mathcal{G}_t$ with probability $1 - \delta$.

Proof. Fix an agent $i \in [N]$ and a time step $t \in [T]$. According to assumption 6.1 and Algorithm 6.1, p be the probability that agent i communicates with a fixed neighbour $j \in \mathcal{N}_i(t)$ in any given step, independently of the past. For a non-negative integer L, we have

$$\Pr(i \text{ does not contact j during the next } L+1 \text{ steps}) = (1-p)^{L+1}.$$

The communication gap length at time t is the number of steps starting from time t until agent i next successfully communicates with agent j. Choose a confidence parameter $\delta \in (0, 1)$, we have

$$\Pr(\text{time until first contact between agents } i \text{ and } j \text{ exceeds } L_p(\delta)) \leq \delta.$$

Applying a union bound to gives

$$\Pr(\exists\, i \in [N], j \in \mathcal{N}_i, t \in [T] : \text{time until first contact between } i \text{ and } j \text{ after time } t \text{ exceeds } L_p(\delta)) \leq N^2T\delta.$$

Note that we have $\mathcal{N}_i(t) \subseteq \mathcal{N}_i$, where $\mathcal{N}_i$ is a fixed superset of possible neighbors. Hence with probability at least $1 - N^2T\delta$, for every agent $i \in [N]$, every time step $t \in [T]$, and every neighbor $j \in \mathcal{N}_i$, the time until the next communication between i and j is at most $L_p(\delta)$ time steps.

For any two agents i and j, let $d(i, j) \leq N$ denote the shortest path length between i and j in the base graph $\mathcal{G}$. Because with high probability, information can gossip across each edge within at most $L_p(\delta)$ steps, it follows that information originating at i at time step t reaches j at most by time step

$$t + L_p(\delta) \cdot d(i, j) \leq t + L_p(\delta) \cdot N. \tag{E.2}$$

In order to be cautious for notations, we use $\mathcal{S}_i(t)$ to denote the active set in round t for Algorithm 6.1. According to Algorithm 6.1, during every communication, each agent i replaces its active set by the intersection with its neighbors:

$$\mathcal{S}_i(t+1) = \bigcap_{j \in \mathcal{N}_i(t) \cup \{i\}} \mathcal{S}_j(t).$$

because for all $\mathcal{S}_i(0) = [K]$, at most K distinct arms can ever be removed. Each arm can start a new wave of disagreement. Each wave of disagreement at most last $L_p(\delta) \cdot N$. In the worst case, the waves do not overlap. Consenquently the longest possible sequence of rounds in which $\mathcal{S}_i(t) \neq \mathcal{S}_j(t)$ is $K \cdot N \cdot L_p(\delta)$. During the disagreement period, for a fixed arm k we could upper bound of maximum pull of k is $KNL_p(\delta)$ and lower bound of maximum pull of k is 0. Hence we have

$$|T_{i,k}(t) - T_{j,k}(t)| \leq KNL_p(\delta).$$

□

E.2 Omitted details in Section 6.4

In this section, we show the omitted details in Section 6.4.

Proof of Lemma 6.1. When $T_{i,k}(t) \leq KL^*$, the bound trivially holds.

Now we consider the case when $T_{i,k}(t) > KL^*$, We first define the following event:

$$E \coloneqq \left\{ \bigcap_{\substack{k \in [K] \\ t \in [T]}} \left\{ \frac{\left| \sum_{j \in [N]} \left(\widehat{\mu}_{j,k}(t) - \mu_{j,k} \right) \right|}{N} \leq \sqrt{\frac{4 \log(T)}{N \min_{j \in [N]} \{T_{j,k}(t)\}}} \right\} \right\}$$

$$\cap \left\{ \bigcap_{\substack{j \in [N] \\ \tau^* \leq t-s \leq T}} \left\| W_t \cdots W_{s+1} e_j - \frac{1}{N} \mathbf{1} \right\|_2 \leq \frac{1}{T^2} \right\} \cap \left\{ \bigcap_{\substack{i,j \in [N] \\ k \in [K] \\ t \in [T]}} |T_{i,k}(t) - T_{j,k}(t)| \leq KL^* \right\}. \quad \text{(E.3)}$$

In Equation (E.3), the first event bounds the global estimation error for each arm, the second ensures near-uniform information mixing across agents after sufficient communication rounds, and the third guarantees that the number of pulls for any arm remains approximately balanced across agents at each time step. By applying Lemma E.2, E.3, E.4 and union bound, we obtain that $\mathbb{P}[E^c] \leq \frac{K}{T} + \frac{N}{T} + N^2 T\delta$ holds for Erdős–Rényi random model defined in Equation (6.1).

We define $\widetilde{\mu}_k(t) = \frac{1}{N} \sum_{j \in [N]} z_{j,k}(t)$ to be an intermediate variable that has access to each agent's average mean on arm k at time t.

For any agent i, we have

$$\begin{aligned} |z_{i,k}(t) - \mu_k| &= |z_{i,k}(t) - \widetilde{\mu}_k(t) + \widetilde{\mu}_k(t) - \mu_k| \\ &\leq |\widetilde{\mu}_k(t) - z_{i,k}(t)| + |\widetilde{\mu}_k(t) - \mu_k| && \text{(Triangle inequality)} \end{aligned}$$

$$= \underbrace{|\widetilde{\mu}_k(t) - z_{i,k}(t)|}_{\text{Consensus Error}} + \underbrace{\frac{\left|\sum_{j\in[N]} (\widehat{\mu}_{j,k}(t) - \mu_{j,k})\right|}{N}}_{\text{Estimation Error}}, \tag{E.4}$$

The last equality is due to the definition of the global mean reward and $\widetilde{\mu}_k(t)$. Let us first focus on Consensus Error. For any $i \in [N]$, According to the update in Equation (6.3) we have

$$\begin{aligned}
z_{i,k}(t) &= \sum_{j\in[N]} [W_{t-1}]_{i,j} z_{i,k}(t-1) + \widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2) \\
&= \sum_{j\in[N]} [W_{t-1}\cdots W_{t-s}]_{i,j} z_{i,k}(t-s) + \sum_{\tau=t-s}^{t-2} \sum_{j\in[N]} [W_{t-2}\cdots W_{\tau+1}]_{i,j} \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)\right) \\
&\quad + \widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2) \qquad \text{(E.5)} \\
&= \sum_{j\in[N]} [W_{t-1}\cdots W_1]_{i,j} z_{i,k}(1) + \sum_{\tau=1}^{t-2} \sum_{j\in[N]} [W_{t-2}\cdots W_{\tau+1}]_{i,j} \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)\right) \\
&\quad + \widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2). \qquad \text{(setting } s = t-1\text{)}
\end{aligned}$$

We also have

$$\begin{aligned}
\widetilde{\mu}_k(t) &= \frac{1}{N} \sum_{j\in[N]} z_{j,k}(t) \\
&= \widetilde{\mu}_k(t-s) + \frac{1}{N} \sum_{\tau=t-s}^{t-1} \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)\right) \\
&= \widetilde{\mu}_k(1) + \frac{1}{N} \sum_{\tau=1}^{t-1} \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)\right) \qquad \text{(setting } s = t-1\text{)} \\
&= \frac{1}{N} \sum_{j\in[N]} z_{j,k}(1) + \frac{1}{N} \sum_{\tau=1}^{t-1} \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)\right) \\
&= \frac{1}{N} \sum_{j\in[N]} z_{j,k}(1) + \frac{1}{N} \sum_{\tau=1}^{t-2} \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)\right) \\
&\quad + \frac{1}{N} \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(t-1) - \widehat{\mu}_{j,k}(t-2)\right) \qquad \text{(E.6)}
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\widetilde{\mu}_k(t) - z_{i,k}(t) &= \sum_{\tau=1}^{t-2} \left(\sum_{j\in[N]} \left(\frac{1}{N} - [W_{t-2}\cdots W_{\tau+1}]_{i,j} \right) \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)\right) \right) \\
&\quad + \frac{1}{N} \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(t-1) - \widehat{\mu}_{j,k}(t-2)\right) - \left(\widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2)\right) \\
&\quad + \frac{1}{N} \sum_{j\in[N]} z_{j,k}(1) - \sum_{j\in[N]} [W_{t-1}\cdots W_1]_{i,j} z_{j,k}(1).
\end{aligned}$$

$$= \sum_{\tau=1}^{t-2} \left(\sum_{j\in[N]} \left(\frac{1}{N} - [W_{t-2}\cdots W_{\tau+1}]_{i,j} \right) (\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)) \right) + \frac{1}{N} \sum_{j\in[N]} (\widehat{\mu}_{j,k}(t-1) - \widehat{\mu}_{j,k}(t-2)) - (\widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2)),$$

.Taking absolute values on both sides, we have

$$\begin{aligned}
|\widetilde{\mu}_k(t) - z_{i,k}(t)| &\le \left| \sum_{\tau=1}^{t-2} \left(\sum_{j\in[N]} \left(\frac{1}{N}[W_{t-2}\cdots W_{\tau+1}]_{i,j} \right) (\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)) \right) \right| \\
&\quad + \left| \frac{1}{N} \sum_{j\in[N]} (\widehat{\mu}_{j,k}(t-1) - \widehat{\mu}_{j,k}(t-2)) - (\widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2)) \right| \\
&\le \underbrace{\left| \sum_{\tau=1}^{t-\tau^*-2} \left(\sum_{j\in[N]} \left(\frac{1}{N} - [W_{t-2}\cdots W_{\tau+1}]_{i,j} \right) (\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)) \right) \right|}_{\heartsuit} \\
&\quad + \underbrace{\left| \sum_{\tau=t-\tau^*-1}^{t-2} \left(\sum_{j\in[N]} \left(\frac{1}{N} - [W_{t-2}\cdots W_{\tau+1}]_{i,j} \right) (\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)) \right) \right|}_{\spadesuit} \\
&\quad + \underbrace{\left| \frac{1}{N} \sum_{j\in[N]} (\widehat{\mu}_{j,k}(t-1) - \widehat{\mu}_{j,k}(t-2)) - (\widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2)) \right|}_{\clubsuit}. \qquad \text{(E.7)}
\end{aligned}$$

Now we analyze three terms on the right-hand side of Equation (E.7).

Bounding term $\heartsuit$. Conditioning on event E, we obtain

$$\begin{aligned}
\heartsuit &= \left| \sum_{\tau=1}^{t-\tau^*-2} \left(\sum_{j\in[N]} \left(\frac{1}{N} - [W_{t-2}\cdots W_{\tau+1}]_{i,j} \right) (\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1)) \right) \right| \\
&\le \sum_{\tau=1}^{t-\tau^*-2} \sum_{j\in[N]} \left| \frac{1}{N} - [W_{t-2}\cdots W_{\tau+1}]_{i,j} \right| \qquad \text{(rewards are bounded in the interval } [0,1]) \\
&= \sum_{\tau=1}^{t-\tau^*-2} \left\| W_{t-2}\cdots W_{\tau+1} e_i - \frac{\mathbf{1}}{N} \right\|_1 \\
&\le \sqrt{N} \cdot \sum_{\tau=1}^{t-\tau^*-2} \left\| W_{t-2}\cdots W_{\tau+1} e_i - \frac{\mathbf{1}}{N} \right\|_2 \\
&\le \frac{\sqrt{N}(t-\tau^*)}{T^2} \\
&\le \frac{\sqrt{N}}{T} \\
&\le \frac{\sqrt{N}}{T_{i,k}(t)}, \qquad \text{(E.8)}
\end{aligned}$$

where the third inequality comes from the condition that $t-\tau-1 \in [\tau^*, t-3]$ and the event E.

Bounding term ♠. We have

$$
\begin{aligned}
\spadesuit &= \left| \sum_{\tau=t-\tau^*-1}^{t-2} \left(\sum_{j\in[N]} \left(\frac{1}{N} - [W_{t-1}\cdots W_{\tau+1}]_{i,j} \right) \left(\widehat{\mu}_{j,k}(\tau) - \widehat{\mu}_{j,k}(\tau-1) \right) \right) \right| \\
&\leq \sum_{\tau=t-\tau^*-1}^{t-2} \left(\sum_{j\in[N]} \left| \left(\frac{1}{N} - [W_{t-1}\cdots W_{\tau+1}]_{i,j} \right) \right| \left| \left(\frac{\sum_{s=1}^{\tau} \mathbb{I}\{A_j(s)=k\} X_{j,k}(s)}{T_{j,k}(\tau)} - \frac{\sum_{s=1}^{\tau-1} \mathbb{I}\{A_j(s)=k\} X_{j,k}(s)}{T_{j,k}(\tau-1)} \right) \right| \right) \\
&\qquad\qquad (\text{ definition of } \widehat{\mu}_{j,k}(t)) \\
&\leq \sum_{\tau=t-\tau^*-1}^{t-2} \left(\sum_{j\in[N]} \left| \left(\frac{1}{N} - [W_{t-1}\cdots W_{\tau+1}]_{i,j} \right) \right| \cdot \right. \\
&\left. \left| \left(\frac{\sum_{s=1}^{\tau-1} \mathbb{I}\{A_j(s)=k\} X_{j,k}(s) + X_{j,k}(\tau)}{T_{j,k}(\tau)} - \frac{\sum_{s=1}^{\tau-1} \mathbb{I}\{A_j(s)=k\} X_{j,k}(s)}{T_{j,k}(\tau)-1} \right) \right| \right) \\
&\leq \sum_{\tau=t-\tau^*-1}^{t-2} \left(\sum_{j\in[N]} \left| \left(\frac{1}{N} - [W_{t-1}\cdots W_{\tau+1}]_{i,j} \right) \right| \left| \left(\frac{-\sum_{s=1}^{\tau-1} \mathbb{I}\{A_j(s)=k\} X_{j,k}(s) + (T_{j,k}(\tau)-1) X_{j,k}(\tau)}{T_{j,k}(\tau)\left(T_{j,k}(\tau)-1\right)} \right) \right| \right) \\
&\leq \sum_{\tau=t-\tau^*-1}^{t-2} \sum_{j\in[N]} \left| \left(\frac{1}{N} - [W_{t-1}\cdots W_{\tau+1}]_{i,j} \right) \right| \frac{1}{T_{j,k}(\tau)} \quad (\text{rewards are bounded in the interval } [0,1]) \\
&\leq \sum_{\tau=t-\tau^*-1}^{t-2} \sum_{j\in[N]} \left| \left(\frac{1}{N} - [W_{t-1}\cdots W_{\tau+1}]_{i,j} \right) \right| \frac{1}{\max\left\{T_{i,k}(\tau) - KL^*, 1\right\}} \qquad (\text{event } E) \\
&\leq \frac{4\tau^*}{\max\left\{T_{i,k}(t) - KL^*, 1\right\}},
\end{aligned}
$$

where the second inequality follows from the fact that there are only two possible cases for $T_{j,k}(\tau)$ and $T_{j,k}(\tau-1)$. When $T_{j,k}(\tau) = T_{j,k}(\tau-1)$ the absolute difference of means is trivially 0, so the only non-trivial case to consider is when $T_{j,k}(\tau) = T_{j,k}(\tau-1)+1$. The last inequality is due to that the upper bound of the total-variation distance from any distribution to the uniform distribution is 1.

Bounding term ♣. Conditioning on E, we obtain

$$
\begin{aligned}
\clubsuit &= \left| \frac{1}{N} \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(t-1) - \widehat{\mu}_{j,k}(t-2) \right) - \left(\widehat{\mu}_{i,k}(t-1) - \widehat{\mu}_{i,k}(t-2) \right) \right| \\
&\leq \sum_{j\in[N]} \frac{2}{N T_{j,k}(t-1)} + \frac{2}{T_{i,k}(t-1)} \\
&\leq \frac{4}{\max\left\{T_{i,k}(t-1) - KL^*, 1\right\}},
\end{aligned}
$$

Next, let us analyse Estimation Error. Conditioned on E, for all $k \in [K]$ and $t \in [T]$ we obtain

$$
\begin{aligned}
\frac{\left| \sum_{j\in[N]} \left(\widehat{\mu}_{j,k}(t) - \mu_{j,k} \right) \right|}{N} &\leq \sqrt{\frac{4\log(T)}{N \min_{j\in[N]}\{T_{j,k}(t)\}}} \\
&\leq \sqrt{\frac{4\log(T)}{N \max\{T_{i,k}(t) - KL^*, 1\}}},
\end{aligned} \tag{E.9}
$$

where the last inequality is due to the event E.

Combining all the results collected so far, we can finally derive the concentration bound conditioned on the event E. For any $i \in [N]$ and $k \in [K]$, we obtain

$$|z_{i,k}(t) - \mu_k| \leq \sqrt{\frac{4\log(T)}{N\max\{T_{i,k}(t) - KL^*, 1\}}} + \frac{4\left(\sqrt{N} + \tau^*\right)}{\max\{T_{i,k}(t) - KL^*, 1\}}. \tag{E.10}$$

□

Proof of Lemma 6.2. First, for all agents we consider the cases under the event E. According to Algorithm 6.1, if arm k is eliminated, there are only two possible cases: 1) there exists some k' such that $z_{i,k}(t) + c_{i,k}(t) \leq z_{i,k'}(t) - c_{i,k'}(t)$; 2) When Algorithm 6.1 updates the active set $\mathcal{S}_i(t+1) \leftarrow \bigcap_{j \in \mathcal{N}_i(t) \cup \{i\}} \mathcal{S}_j(t)$, $k \notin \mathcal{S}_j$ for any $j \in \mathcal{N}_i(t)$.

For Case 1, Since we have $z_{i,k}(t) + c_{i,k}(t) \leq \mu_k + 2c_{i,k}(t)$ as well as $z_{i,k'}(t) - c_{i,k'}(t) \geq \mu_{k'} - 2c_{i,k'}(t)$, then when

$$2(c_{i,k}(t) + c_{i,k'}(t)) \leq \mu_{k'} - \mu_k \leq \Delta_k$$

arm k will be essentially eliminated. Due to the pulling rule of Algorithm 6.1, we have $|T_{i,k}(t) - T_{i,k'}(t)| \leq 1$. Thus when

$$\Delta_k \geq 2\left(\sqrt{\frac{4\log(T)}{N\max\{T_{i,k}(t) - KL^*, 1\}}} + \frac{4\left(\sqrt{N} + \tau^*\right)}{\max\{T_{i,k}(t) - KL^*, 1\}}\right).$$

Hence, here we know that when

$$T_{i,k}(t) \geq \frac{64\log(T)}{N\Delta_k^2} + \frac{16\left(\sqrt{N} + \tau^*\right)}{\Delta_k} + KL^*,$$

arm k will be essentially eliminated.

For Case 2,the optimal arm k^* it cannot be eliminated, because for agents i we consider the cases under the event E.

For all agents cases under event E^c, we have

$$T \cdot \mathbb{P}(E^c)\Delta_{\max} \leq 3KN\Delta_{\max}.$$

Therefore, combining all inequalities above, we have

$$\mathrm{Reg}_{i,T}(\pi) \leq \sum_{k:\Delta_k > 0}\left(\frac{64\log(T)}{N\Delta_k} + 16\left(\sqrt{N} + \tau^*\right)\right) + KL^* + 3KN\Delta_{\max}$$

which ends the proof. □

Proof of Theorem 6.1. By adding up Lemma 6.2 for all agents $i \in \mathcal{N}$, Theorem 6.1 can be proved

through the facts that

$$\frac{1}{log\lambda_2^{-1}} \leq \frac{1}{1-\lambda_2}, \lambda_2(\mathbb{E}[W^2]) \leq 1 - \frac{p\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))}{N} \text{ and } \frac{1}{\log(1-p)^{-1}} \leq \frac{1}{p},$$

where the second inequality is taken from Theorem 6.1 in Achddou et al. (2024). □

Proof of Corollary 6.1. If $\mathcal{G}$ is complete, then $d(i,j) = 1$ for all $i, j \in \mathcal{N}$, and the gossip time simplifies to $t + L$. Recall Equation (E.2), with the same steps we can prove that

$$|T_{i,k}(t) - T_{j.k}(t)| \leq KL_p(\delta) = K\left\lceil \frac{\log(\delta)}{\log(1-p)} \right\rceil$$

for any arm $k \in [K]$ and agents $i, j \in [N]$ for any $t \in [T]$.

Hence by the fact that $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G})) = N$ and let L^* as defined in Corollary 6.1 we can prove the case for complete graph.

For the rest two cases, we can just prove which by substituting the exact values of $\lambda_{N-1}(\mathrm{Lap}(\mathcal{G}))$ in Theorem 6.1. □

Proof of Theorem 6.2. First, note that

$$\mathrm{Reg}_T^{\nu}(\pi) = \sum_{i\in[N]} \mathrm{Reg}_{i,T}^{\nu}(\pi) = \sum_{i\in[N]} \sum_{k\in[K]} \mathbb{E}[T_{i,k}(T)]\Delta_k = \sum_{k\in[K]} \Delta_k \sum_{i\in[N]} \mathbb{E}[T_{i,k}(T)].$$

In order to show Theorem 6.2, we only need to prove that

$$\lim_{T\to\infty} \frac{\sum_{i\in[N]} \mathbb{E}[T_{i,k}(T)]}{\log T} \geq \frac{2(1-s)}{(1+\varepsilon)^2\Delta_k^2} \tag{E.11}$$

for p and ε and some sub-optimal arm $k \neq k^*$.

Suppose the distribution over instance ν is given by $\mathcal{P} = (\mathbb{P}_{i,a})_{i\in[N],a\in[K]}$. Consider another instance ν' with $\mathcal{P}' = \left(\mathbb{P}'_{i,a}\right)_{i\in[N],a\in[K]}$ such that

$$\mathbb{P}'_{i,a} = \begin{cases} \mathcal{N}(\mu_{i,a},\, 1),\ a \neq k \\ \mathcal{N}(\mu_{i,a} + (1+\varepsilon)\Delta_a,\, 1),\ a = k \end{cases}$$

for $i \in [N]$ where $\mathbb{P}_{i,a} = \mathcal{N}(\mu_{i,a}, 1)$ and $\varepsilon \in (0, 1]$.

According to the consistency of policy π, it holds that

$$\mathrm{Reg}_T^{\nu}(\pi) + \mathrm{Reg}_T^{\nu'}(\pi) \leq 2CT^s$$

for some constant C and $p \in (0, 1)$.

Simultaneously, let event $A_i = \left\{T_{i,k}(T) \geq \frac{T}{2}\right\}$, then

$$\begin{aligned} \mathrm{Reg}_{i,T}^{\nu}(\pi) + \mathrm{Reg}_{i,T}^{\nu'}(\pi) &\geq \frac{T}{2} \cdot \Delta_k \cdot \mathbb{P}_{\nu,\pi}[A_i] + \frac{T}{2} \cdot \varepsilon \cdot \Delta_k \cdot \mathbb{P}_{\nu',\pi}[A_i^c] \\ &\geq \frac{T}{2}\varepsilon\Delta_k \left(\mathbb{P}_{\nu,\pi}[A_i] + \mathbb{P}_{\nu',\pi}[A_i^c]\right) \end{aligned}$$

$$
\begin{aligned}
&\geq \frac{T}{4}\varepsilon\Delta_k \exp\left(-\mathrm{KL}(\nu,\nu')\right) \\
&= \frac{T}{4}\varepsilon\Delta_k \exp\left(\sum_{j\in[N]} \mathbb{E}[T_{j,k}(T)] \cdot \mathrm{KL}(\mathbb{P}_{j,a}, \mathbb{P}'_{j,a})\right) \\
&= \frac{T}{4}\varepsilon\Delta_k \exp\left(-\sum_{j\in[N]} \mathbb{E}[T_{j,k}(T)] \cdot \frac{(1+\varepsilon)^2\Delta_k^2}{2}\right).
\end{aligned}
$$

Hence, summing the above inequality for all agents $i \in [N]$, we obtain

$$
\mathrm{Reg}_T^{\nu}(\pi) + \mathrm{Reg}_T^{\nu'}(\pi) \geq \frac{NT}{4}\varepsilon\Delta_k \exp\left(-\sum_{j\in[N]} \mathbb{E}[T_{j,k}(T)] \cdot \frac{(1+\varepsilon)^2\Delta_k^2}{2}\right). \tag{E.12}
$$

Rearranging the terms in Equation (E.12), it holds that

$$
\begin{aligned}
\sum_{j\in[N]} \mathbb{E}[T_{j,k}(T)] \cdot \frac{(1+\varepsilon)^2\Delta_k^2}{2} &\geq \log\left(\frac{NT\varepsilon\Delta_k/4}{\mathrm{Reg}_T^{\nu}(\pi) + \mathrm{Reg}_T^{\nu'}(\pi)}\right) \\
&\geq \log\left(\frac{NT\varepsilon\Delta_k}{8CT^s}\right) = (1-s)\log(T) + \log\left(\frac{N\varepsilon\Delta_k}{8C}\right).
\end{aligned}
$$

Therefore, we can show

$$
\lim_{T\to\infty} \frac{\sum_{i\in[N]} \mathbb{E}[T_{j,k}(T)]}{\log T} \geq \frac{2(1-s)}{(1+\varepsilon)^2\Delta_k^2}
$$

which is the goal in Equation (E.11). □

E.3 Estimation of unknown link probability

Since $\mathcal{G}$ is connected, each agent has at least one neighbor. This allows every agent to estimate the edge activation probability p by observing its connectivity status over multiple rounds. We design the following procedure for each agent $i \in [N]$ to compute an estimate $\widehat{p}$ of p.

Algorithm E.1: Burn-in Phase for Estimating p by Agent $i \in [N]$

Input: Confidence level $\delta \in (0,1)$, any fixed neighbor $n_i \in [N]_i$ of agent i in the base graph $\mathcal{G}$
Initialization: Set $\widetilde{\tau}_i \leftarrow 0$, $t \leftarrow 0$, $\widehat{p}_i(0) \leftarrow 0$ and $\mathrm{CI}_i(0) \leftarrow \infty$
while $\widehat{p}_i(t) - 3\mathrm{CI}_i(t) \leq 0$ **do**
    Increment time $t \leftarrow t+1$ and observe $\mathcal{N}_i(t)$
    if $n_i \in [N]_i(t)$ **then**
        $\widetilde{\tau}_i \leftarrow \widetilde{\tau}_i + 1$
    Select an arm uniformly at random
    Update $\widehat{p}_i(t) = \frac{\widetilde{\tau}_i}{t}$ and $\mathrm{CI}_i(t) = \sqrt{\frac{\log(2/\delta)}{2t}}$
Output: $\widehat{p}_i = \widehat{p}_i(t) - \mathrm{CI}_i(t)$

Lemma E.5. *For the agent i and each t, it holds that*

$$
\mathbb{P}\left[|\widehat{p}_i(t) - p| \leq \mathrm{CI}_i(t)\right] \geq 1 - \delta.
$$

Proof of Lemma E.5. Lemma E.5 comes directly from Hoeffding's inequality. □

Theorem E.1. *Let $\delta = \frac{2}{T^2}$. Then, with probability at least $1 - \frac{2N}{T}$, the estimate $\widehat{p}_i$ returned by Algorithm E.1 satisfies $\widehat{p}_i \in \left(\frac{p}{2}, p\right]$ for every agent $i \in [N]$. Furthermore, the cumulative regret incurred during the burn-in phase across all agents is bounded by $\mathcal{O}\left(\frac{N \log T}{p^2}\right)$.*

Proof of Theorem E.1. Suppose the end round for Algorithm E.1 is t^*. As $\widehat{p}_i = \widehat{p}_i(t) - \mathrm{CI}_i(t) \leq p$ according to Lemma E.5, we only need to show $\widehat{p}_i > \frac{p}{2}$. This can be verified by

$$\widehat{p}_i = \widehat{p}_i(t) - \mathrm{CI}_i(t) > \frac{\widehat{p}_i(t) + \mathrm{CI}_i(t)}{2} \geq \frac{p}{2}.$$

By combining each $t \in [T]$ and $i \in [N]$ and union bound, we can obtain the first part of Theorem E.1. Moreover, for $t^* > \frac{16 \log(T)}{p^2}$, we have

$$\widehat{p}_i(t) - 3\mathrm{CI}_i(t) \geq p - 4\mathrm{CI}_i(t) = p - 4\sqrt{\frac{\log T}{t^*}} > 0,$$

hence the stopping condition in Algorithm E.1 holds. By the fact that $\Delta_{\max} \leq 1$ for all arms, we can obtain the second part of Theorem E.1 through adding the regret for all agents $i \in [N]$. □