Group Distributionally Robust Optimization with Flexible Sample Queries

Haomin Bai[1,2] **Dingzhi Yu**[1,2] **Shuai Li**[3,4] **Haipeng Luo**[5] **Lijun Zhang**[1,2*]
[1] State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing, China
[2] School of Artificial Intelligence, Nanjing University, Nanjing, China
[3] The University of Hong Kong, Hong Kong, China
[4] Shanghai Innovation Institute, Shanghai, China
[5] Thomas Lord Department of Computer Science, University of Southern California, USA
{baihm,yudz,zhanglj}@lamda.nju.edu.cn
shuaili@hku.hk haipengl@usc.edu

Abstract

Group distributionally robust optimization (GDRO) aims to learn models that perform well across m distributions simultaneously. Existing methods either (i) query m samples per round, which imposes strict implementation requirements, or (ii) query only a single sample per round, which requires many rounds to converge and therefore leads to long runtime. To avoid these limitations, we introduce a novel setting called *flexible sample queries* (FSQ), which allows the number of samples that can be queried to vary across rounds. Under the FSQ setting, we cast GDRO as a two-player game: one performs non-oblivious online convex optimization, while the other tackles a non-oblivious prediction with limited advice (PLA) problem. In this game, the first player updates its decision using follow-the-regularized-leader with stochastic gradients. For the second player, we develop a new PLA method equipped with an adaptive exploration strategy that can leverage an arbitrary per-round sample size. We then establish the *first* high-probability regret bound for non-oblivious PLA. By integrating these components, we develop an anytime GDRO algorithm that supports FSQ, and derive a high-probability optimization error bound of $O\left(\frac{1}{t}\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}\log m}\right)$, where r_j is the sample size at round j. This result demonstrates that the optimization error decreases as the per-round sample sizes increase, and implies the same near-optimal sample complexity of $O(m\log(m)/\epsilon^2)$ for any fixed sample size $r \in [m]$.

1 Introduction

Traditional machine learning methods typically train models by minimizing the empirical risk over random samples drawn from an unknown distribution [Vapnik, 2000]. However, when the test data distribution deviates from the training data distribution, models often suffer significant performance degradation [Sugiyama et al., 2007, Gama et al., 2014, Kuang et al., 2018]. Distributionally robust optimization (DRO) is a promising approach to mitigating the effect of distribution shifts by minimizing the worst-case risk over an uncertainty set [Delage and Ye, 2010, Duchi and Namkoong, 2021, Rahimian and Mehrotra, 2022].

In this paper, we focus on a special case of DRO, known as group DRO (GDRO) [Oren et al., 2019, Sagawa et al., 2020], where the uncertainty set is defined as a finite collection of distributions.

*Lijun Zhang is the corresponding author.

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

GDRO can be mathematically formulated as a minimax stochastic optimization problem

$$\min_{\mathbf{w}\in\mathcal{W}} \max_{i\in[m]} \left\{R_i(\mathbf{w}) = \mathbb{E}_{\mathbf{z}\sim\mathcal{P}_i}\left[\ell(\mathbf{w};\mathbf{z})\right]\right\}, \tag{1}$$

where $\{\mathcal{P}_i\}_{i\in[m]}$ denotes a set of m distributions, $\mathbf{z}$ is a random sample, $\mathbf{w}$ corresponds to the model, $\mathcal{W}$ represents the hypothesis class, and $\ell(\cdot;\cdot)$ is the loss function. Assuming that each risk function $R_i(\mathbf{w})$ is convex for all $i \in [m]$, the optimization problem in (1) can be cast as a stochastic convex-concave saddle-point problem [Nemirovski et al., 2009]

$$\min_{\mathbf{w}\in\mathcal{W}} \max_{\mathbf{q}\in\Delta_m} \left\{\phi(\mathbf{w},\mathbf{q}) = \sum_{i=1}^{m} q_i R_i(\mathbf{w})\right\}, \tag{2}$$

where $\Delta_m = \{\mathbf{q} \in \mathbb{R}^m | \mathbf{q} \geq \mathbf{0}_m, \sum_{i=1}^m q_i = 1\}$ denotes the $(m-1)$-dimensional simplex.

Recent work has proposed various algorithms for solving GDRO in (1)/(2) and established a nearly optimal sample complexity of $O(m\log(m)/\epsilon^2)$, where ϵ denotes the error tolerance for (1). These methods assume sampling access to any of the m distributions, but restrict the per-round sample size to either m [Nemirovski et al., 2009] or 1 [Haghtalab et al., 2022, Soma et al., 2022, Carmon and Hausler, 2022, Zhang et al., 2023]. In many applications, however, the per-round sampling capacity is constrained by time-varying external resources, such as communication bandwidth in mobile-edge federated learning [Nishio and Yonetani, 2019], human review capacity in content moderation [Allouah et al., 2023], and expert annotation capacity in medical data collection [Wang et al., 2024]. In these settings, the learner can still decide which groups to query under the available sampling capacity. We provide a detailed discussion of these examples in Appendix A.

To capture such scenarios, we introduce a new setting termed *flexible sample queries* (FSQ), where at the beginning of each round t, the environment reveals a potentially time-varying sample size $r_t \in [m]$. After observing r_t, the learner decides which groups to query. Existing GDRO algorithms are not satisfactory in the FSQ setting for the following reasons:

1. Methods whose updates require m samples cannot proceed when the available sample size is less than m. They must wait until enough samples are collected to form an update, which introduces latency.
2. Methods that process one sample per update need to handle the r_t samples sequentially, which substantially increases the number of iteration rounds and leads to a long runtime.

Overall, existing methods are tailored to extreme cases and cannot adapt to varying sample sizes.

In this paper, we study GDRO under the FSQ setting and, following prior work, formulate it as a two-player game [Freund and Schapire, 1997], where the minimizer is the w-player and the maximizer is the q-player. For the w-player, we model its task as "non-oblivious online convex optimization (OCO) with stochastic gradients" [Zhang et al., 2023], and update its decisions using follow-the-regularized-leader (FTRL) [Orabona, 2019]. By constructing an unbiased stochastic gradient from a single queried sample, FTRL yields guarantees independent of the per-round sample size. Our main technical contribution is to design a q-player algorithm that can operate with a variable number of samples per round. To this end, we model the q-player's task as "non-oblivious prediction with limited advice (PLA)". However, existing PLA methods are not applicable to our setting, as they either do not handle non-oblivious losses [Seldin et al., 2014] or cannot accommodate varying sample sizes [Yun et al., 2018]. We therefore propose NPLA-EXP3, an Exp3-style algorithm for non-oblivious PLA, which employs an adaptive exploration strategy based on the sample size r_t. When $r_t \geq 2$, NPLA-EXP3 performs *active* exploration by allocating 1 sample to exploitation and the remaining $r_t - 1$ to exploration, yielding an unbiased loss estimator as in Seldin et al. [2014]. When $r_t = 1$, NPLA-EXP3 resorts to *implicit* exploration by adding an offset to the selection probability, resulting in the biased Implicit-eXploration (IX) estimator [Kocák et al., 2014]. The loss estimators are accumulated over rounds to compute $\mathbf{q}_t$ via FTRL with negative entropy. Based on this design, we establish the *first* high-probability regret bound $O\left(\sqrt{\sum_{j=1}^{t} \frac{m}{r_j}\log m}\right)$ for non-oblivious PLA.

Leveraging the methods for the two players, we develop FLEX-GDRO and establish a high-probability optimization error bound that depends on the per-round sample size r_t. In addition, FLEX-GDRO runs in an *anytime* manner, meaning that it does not require prior knowledge of the number of iterations and can return a solution at any round [Cutkosky, 2019, Zhang et al., 2026]. Our contributions are summarized as follows:

1. We introduce the FSQ setting and develop FLEX-GDRO, the first GDRO algorithm that can process an arbitrary number of samples per round. Then, we establish a high-probability optimization error bound of $O\left(\frac{1}{t}\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}\log m}\right)$.
2. When the per-round sample size is fixed, i.e., $r_t \equiv r$, FLEX-GDRO improves upon existing methods in three respects: (i) it possesses the anytime ability while matching the near-optimal sample complexity $O(m\log(m)/\epsilon^2)$ achieved by non-anytime algorithms [Nemirovski et al., 2009, Zhang et al., 2023]; (ii) it improves the anytime optimization error bound from $O(\log(t)\sqrt{m\log(m)/(rt)})$ [Zhang et al., 2026] to $O(\sqrt{m\log(m)/(rt)})$; and (iii) it supports any $r \in [m]$, rather than $r \in \{1, m\}$.
3. As a by-product, we propose a novel non-oblivious PLA algorithm, NPLA-EXP3, and establish the first high-probability regret bound $O\left(\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}\log m}\right)$, which is of the same order as the expected regret bound against an oblivious adversary [Seldin et al., 2014].

Related Work. We briefly review related work on GDRO and PLA, deferring a more detailed discussion to Appendix B. GDRO is motivated by applications affected by distribution shift [Mohri et al., 2019, Oren et al., 2019, Sagawa et al., 2020]. Existing methods fall into two fixed sampling regimes: querying one sample from each of the m distributions per round [Nemirovski et al., 2009], or querying one sample from a single distribution per round [Carmon and Hausler, 2022, Haghtalab et al., 2022, Soma et al., 2022, Zhang et al., 2023]. Existing anytime variants are also restricted to the two extremes $r \in \{1, m\}$ [Zhang et al., 2026]. The learning problem faced by the q-player is related to PLA [Seldin et al., 2014], which lies between multi-armed bandits (MAB) [Robbins, 1952, Auer et al., 2002] and prediction with expert advice [Freund and Schapire, 1997]. In PLA, the player queries the advice of $r_t \in [m]$ experts in round t, and the loss is evaluated on a subset of the selected experts [Seldin et al., 2014, Saad and Blanchard, 2023]. Existing PLA algorithms are inapplicable to our setting, which requires minimizing regret against a non-oblivious adversary with varying sample sizes, because prior work assumes stochastic adversaries [Saad and Blanchard, 2021], oblivious adversaries [Seldin et al., 2014, Saad and Blanchard, 2023], fixed sample sizes [Thune and Seldin, 2018, Yun et al., 2018] or focuses on pseudo-regret objectives [Kale, 2014, Amin et al., 2015].

2 Preliminaries and Two-Player Game Formulation

We begin by equipping the domain $\mathcal{W}$ with a function $\nu_w(\cdot)$ that is 1-strongly convex with respect to the norm $\|\cdot\|_w$. Similarly, the simplex Δ_m is equipped with the negative entropy function $\nu_q(\mathbf{q}) = \sum_{i=1}^{m} q_i \ln q_i$, which is 1-strongly convex with respect to the ℓ_1-norm $\|\cdot\|_1$. Next, we make the following standard assumptions [Nemirovski et al., 2009, Haghtalab et al., 2022].

Assumption 2.1. $\mathcal{W}$ is convex and its diameter measured by a function $\nu_w(\cdot)$ is bounded by D, i.e.,

$$\max_{\mathbf{w}\in\mathcal{W}} \nu_w(\mathbf{w}) - \min_{\mathbf{w}\in\mathcal{W}} \nu_w(\mathbf{w}) \le D^2. \tag{3}$$

It is immediate that Δ_m is convex, and its diameter measured by $\nu_q(\cdot)$ is bounded by $\sqrt{\ln m}$.

Assumption 2.2. For all $i \in [m]$, we assume

$$\|\nabla \ell(\mathbf{w}; \mathbf{z})\|_{w,*} \le G \quad \forall \mathbf{w} \in \mathcal{W},\ \mathbf{z} \sim \mathcal{P}_i, \tag{4}$$

where $\|\cdot\|_{w,*}$ denotes the dual norm of $\|\cdot\|_w$.

Assumption 2.3. For all $i \in [m]$, we assume

$$0 \le \ell(\mathbf{w}; \mathbf{z}) \le 1 \quad \forall \mathbf{w} \in \mathcal{W},\ \mathbf{z} \sim \mathcal{P}_i. \tag{5}$$

Assumption 2.4. For all $i \in [m]$, the risk function $R_i(\mathbf{w}) = \mathbb{E}_{\mathbf{z}\sim\mathcal{P}_i}[\ell(\mathbf{w}; \mathbf{z})]$ is convex.

For any approximate solution $(\bar{\mathbf{w}}, \bar{\mathbf{q}})$ of problem (2), its optimization error is defined as

$$\epsilon_\phi(\bar{\mathbf{w}}, \bar{\mathbf{q}}) = \max_{\mathbf{q}\in\Delta_m} \phi(\bar{\mathbf{w}}, \mathbf{q}) - \min_{\mathbf{w}\in\mathcal{W}} \phi(\mathbf{w}, \bar{\mathbf{q}}),$$

which controls the optimality of $\bar{\mathbf{w}}$ for the original GDRO problem (1) [Zhang et al., 2023].

Error Decomposition via a Two-Player Game. We cast the problem (2) as a two-player game [Freund and Schapire, 1997]: in each round t, the w-player selects $\mathbf{w}_t \in \mathcal{W}$ to minimize the convex function $\phi(\mathbf{w}, \mathbf{q}_t) = \sum_{i=1}^m q_{t,i} R_i(\mathbf{w})$, and the q-player selects $\mathbf{q}_t \in \Delta_m$ to maximize the linear function $\phi(\mathbf{w}_t, \mathbf{q}) = \sum_{i=1}^m q_i R_i(\mathbf{w}_t)$. Based on iterative decisions, we output the averaged solutions $(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t)$ at each iteration t, defined as

$$\bar{\mathbf{w}}_t = \frac{1}{t}\sum_{j=1}^t \mathbf{w}_j, \text{ and } \bar{\mathbf{q}}_t = \frac{1}{t}\sum_{j=1}^t \mathbf{q}_j. \tag{6}$$

Under Assumption 2.4 and by Jensen's inequality, the optimization error can be decomposed into the sum of two average regrets

$$\begin{aligned}
&\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) \\
\leq &\frac{1}{t}\underbrace{\left(\sum_{j=1}^t \phi(\mathbf{w}_j, \mathbf{q}_j) - \min_{\mathbf{w}\in\mathcal{W}} \sum_{j=1}^t \phi(\mathbf{w}, \mathbf{q}_j)\right)}_{\mathrm{Regret}_{w,t}} + \frac{1}{t}\underbrace{\left(\max_{\mathbf{q}\in\Delta_m} \sum_{j=1}^t \phi(\mathbf{w}_j, \mathbf{q}) - \sum_{j=1}^t \phi(\mathbf{w}_j, \mathbf{q}_j)\right)}_{\mathrm{Regret}_{q,t}},
\end{aligned} \tag{7}$$

where $\mathrm{Regret}_{w,t}$ and $\mathrm{Regret}_{q,t}$ are the regrets of the w-player and the q-player, respectively.

Learning Problems for the Two Players. In the game formulation, both players face *non-oblivious* online learning problems, since the objective functions encountered by each player depend on their past decisions. Moreover, the players cannot observe the exact values of risks $R_i(\mathbf{w})$ or gradients $\nabla R_i(\mathbf{w})$. Instead, they only have access to stochastic estimates using randomly queried samples. Building on prior analyses [Zhang et al., 2023], we observe that for the w-player, 1 sample drawn according to $\mathbf{q}_t$ suffices to form an unbiased stochastic gradient of $\phi(\mathbf{w}, \mathbf{q}_t)$ with low norm under Assumption 2.2. Therefore, we model the problem faced by the w-player as "non-oblivious OCO with stochastic gradients", whose performance does not depend on the per-round sample size r_t. In contrast, for the q-player, r_t has a direct impact on $\mathrm{Regret}_{q,t}$. To make this dependence explicit, we decouple the stochastic nature (arising from stochastic observation) from the adversarial nature (non-obliviousness in the game) as follows, in a manner similar to Haghtalab et al. [2022].

We first transform the maximization problem over $\{\phi(\mathbf{w}_j, \mathbf{q}) = \sum_{i=1}^m q_i R_i(\mathbf{w}_j)\}_{j\in[t]}$ into an equivalent minimization problem over $\{\sum_{i=1}^m q_i(1 - R_i(\mathbf{w}_j))\}_{j\in[t]}$ under Assumption 2.3. Next, we define the surrogate loss

$$\hat{s}_{t,i} = 1 - \ell(\mathbf{w}_t; \mathbf{z}_t^{(i)}) \overset{(5)}{\in} [0, 1], \quad \forall i \in [m], \tag{8}$$

where each $\mathbf{z}_t^{(i)} \sim \mathcal{P}_i$ is a random sample drawn at round t. We note that $\hat{\mathbf{s}}_t$ is introduced for analysis, and only the queried coordinates $\hat{s}_{t,i}$ are observed. For any fixed $t \in \mathbb{N}_+$ and $\delta \in (0, 1)$, concentration inequalities [Cesa-Bianchi and Lugosi, 2006] imply that, with probability at least $1-\delta$,

$$\mathrm{Regret}_{q,t} = \sum_{j=1}^t \langle \mathbf{q}_j, \mathbb{E}[\hat{\mathbf{s}}_j]\rangle - \min_{i\in[m]} \sum_{j=1}^t \mathbb{E}[\hat{s}_{j,i}] \leq \underbrace{\sum_{j=1}^t \langle \mathbf{q}_j, \hat{\mathbf{s}}_j\rangle - \min_{i\in[m]} \sum_{j=1}^t \hat{s}_{j,i}}_{\mathrm{Regret}'_{q,t}} + O\left(\sqrt{t \log \frac{m}{\delta}}\right),$$

where the last term captures the stochastic nature. We then assign the q-player to minimize $\mathrm{Regret}'_{q,t}$ with r_t observations per round, which yields a "non-oblivious PLA" problem.

Remark: One may attempt to cast the q-player's problem as a combinatorial semi-bandit [Audibert et al., 2014, Neu and Bartók, 2016]. However, methods for combinatorial semi-bandits are not applicable because they compete with the best subset of r experts in terms of average loss, which does not control $\mathrm{Regret}_{q,t}$ defined against the best *single* expert.

Combining the above results, the following bound holds with probability at least $1 - \delta$ (see Lemma G.1 for a complete statement),

$$\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) \leq \frac{1}{t}\mathrm{Regret}_{w,t} + \frac{1}{t}\mathrm{Regret}'_{q,t} + O\left(\sqrt{\frac{1}{t}\log\frac{m}{\delta}}\right). \tag{9}$$

Therefore, it suffices to design no-regret online learning algorithms for each player, and combining their regret bounds yields the overall optimization error. In particular, the resulting q-player algorithm is established for non-oblivious PLA and may be of independent interest beyond GDRO.

3 FLEX-GDRO for GDRO with FSQ

We introduce FLEX-GDRO through three components: (i) the q-player strategy $\mathcal{A}_q$, (ii) the w-player strategy $\mathcal{A}_w$, and (iii) a meta-scheduler $\mathcal{A}_s$ that interacts with the environment and coordinates the two players. Since the w-player can reuse the samples queried by the q-player (see Section 3.2), we first describe $\mathcal{A}_q$, followed by $\mathcal{A}_w$, and finally $\mathcal{A}_s$. The sample-size protocol is defined as follows.

Definition 3.1 (Sample-size protocol)**.** At the beginning of each round t, the environment reveals a sample size $r_t \in [m]$ before the learner's round-t randomization and before fresh samples are drawn. The sample size r_t may depend on all history before round t and on exogenous randomness.

3.1 NPLA-EXP3 for the q-player

Recall that the q-player needs to solve a non-oblivious PLA problem, to which existing methods are not applicable (see Appendix B for a detailed discussion). A key challenge in achieving low regret is to appropriately balance exploitation and exploration, as in MAB [Bubeck and Cesa-Bianchi, 2012]. To address this challenge, we propose an adaptive exploration strategy based on the per-round sample size r_t:

1. When $r_t \geq 2$, we use 1 sample for exploitation and the remaining $r_t - 1$ samples for exploration via uniform selection over the other $m - 1$ groups. This active exploration strategy, first applied in oblivious PLA [Seldin et al., 2014], yields an unbiased estimator with low variance.
2. When $r_t = 1$, active exploration is infeasible because we only observe the loss of the exploited option. In this case, we adopt *implicit* exploration via the IX estimator [Kocák et al., 2014], which injects a small bias into the loss estimator.

The two loss estimators induced by these regimes are then combined to update $\mathbf{q}_t$. We detail the procedure below and provide theoretical guarantees via a novel analysis.

At the start of each round t, the q-player receives r_t from the meta-scheduler $\mathcal{A}_s$ and then updates $\mathbf{q}_t$ via FTRL as

$$\mathbf{q}_t = \operatorname*{argmin}_{\mathbf{q} \in \Delta_m} \left\{ \langle \eta_{q,t} \mathbf{L}_{t-1}, \mathbf{q} \rangle + \nu_q(\mathbf{q}) \right\}, \tag{10}$$

where $\mathbf{L}_{t-1} \in \mathbb{R}^m$ is the cumulative loss vector initialized as $\mathbf{L}_0 = \mathbf{0}_m$, and the step size is set as

$$\eta_{q,t} = \sqrt{\frac{\ln m}{m \sum_{j=1}^{t} \frac{1}{r_j}}}. \tag{11}$$

With the regularizer $\nu_q(\mathbf{q}) = \sum_{i=1}^{m} q_i \ln q_i$, the FTRL update in (10) reduces to the exponential-weights update rule

$$q_{t,i} = \frac{\exp(-\eta_{q,t} L_{t-1,i})}{\sum_{k=1}^{m} \exp(-\eta_{q,t} L_{t-1,k})}, \quad \forall i \in [m]. \tag{12}$$

We proceed by querying new samples through the following process. The initial step is to select a distribution index $c_t \in [m]$ according to $\mathbf{q}_t$. For rounds where $r_t \geq 2$, we additionally select $r_t - 1$ distribution indices uniformly at random and without replacement from the remaining $m - 1$ elements. These additional indices form the set $\mathcal{I}_t$ and the complete set of selected indices is denoted by $C_t = \{c_t\} \cup \mathcal{I}_t$. The sampling of $\mathcal{I}_t$ can be implemented using the DepRound algorithm [Gandhi et al., 2006] with $O(m)$ time and space complexity, as detailed in Appendix F. Other sampling procedures that yield the required subset size and marginal inclusion probabilities can also be used.

Next, the set C_t together with the exploited index c_t are sent to the scheduler $\mathcal{A}_s$, which returns the realized losses $\hat{s}_{t,i}$ for all $i \in C_t$. We then construct an adaptive loss estimator $\tilde{\mathbf{s}}_t$ based on the per-round sample size r_t for each $i \in [m]$ as

$$\tilde{s}_{t,i} = \begin{cases} \dfrac{\hat{s}_{t,i}}{q_{t,i} + \gamma_t} \mathbb{I}\left[i \in C_t\right], & r_t = 1 \\ \dfrac{\hat{s}_{t,i}}{q_{t,i} + (1 - q_{t,i}) \frac{r_t - 1}{m - 1}} \mathbb{I}\left[i \in C_t\right], & r_t \geq 2 \end{cases}, \tag{13}$$

Algorithm 1 NPLA-EXP3 for non-oblivious PLA

Input: a scheduler $\mathcal{A}_s$

Initialize $\mathbf{L}_0 = \mathbf{0}_m$
for $t = 1, 2, \ldots$ **do**
  Receive r_t from $\mathcal{A}_s$ and compute $\mathbf{q}_t$ according to (12)
  if $r_t = 1$ **then**
    Select $c_t \in [m]$ according to $\mathbf{q}_t$ and set $\mathcal{I}_t = \emptyset$
  else
    Select $c_t \in [m]$ according to $\mathbf{q}_t$ and generate $\mathcal{I}_t = \text{DepRound}(\frac{r_t-1}{m-1}\mathbf{1}_{m-1})$ from $[m]\backslash\{c_t\}$
  end if
  Send $C_t = \{c_t\} \cup \mathcal{I}_t$, c_t and $\mathbf{q}_t$ to $\mathcal{A}_s$
  Receive losses $\hat{s}_{t,i}$ for all $i \in C_t$ from $\mathcal{A}_s$
  Construct $\tilde{\mathbf{s}}_t$ in (13) and update $\mathbf{L}_t$ according to (15)
end for

where $\hat{s}_{t,i}$ is defined in (8), and the IX coefficient γ_t is set to

$$\gamma_t = \frac{1}{2}\eta_{q,t}. \tag{14}$$

As can be seen from (13), when $r_t = 1$, we adopt the (biased) IX estimator [Kocák et al., 2014], which replaces the true observation probability $q_{t,i}$ with the IX-adjusted term $q_{t,i}+\gamma_t$. The parameter γ_t implements *implicit* exploration and prevents overly aggressive importance weighting when $q_{t,i}$ is small. When $r_t \geq 2$, we instead use the true observation probability induced by the sampling rule, which yields an unbiased estimator [Seldin et al., 2014].

At the end of each round t, the cumulative loss is updated as

$$\mathbf{L}_t = \mathbf{L}_{t-1} + \tilde{\mathbf{s}}_t. \tag{15}$$

Since our algorithm is designed for non-oblivious PLA and adopts an EXP3-style update rule [Auer et al., 2002] with adaptive exploration, we term it NPLA-EXP3 and present it in Algorithm 1.

We now state the theoretical guarantee for Algorithm 1. Although in rounds with $r_t \geq 2$ we employ the same loss estimator as Seldin et al. [2014], their analysis does not apply to our setting since it establishes only an expected regret guarantee against an *oblivious* adversary, whereas our goal is to derive a high-probability bound in the more challenging *non-oblivious* setting. Moreover, NPLA-EXP3 alternates between two loss estimators: on the one hand, we need to control the bias induced by implicit exploration; on the other hand, it is also necessary to exploit the low variance offered by the multi-sample loss estimator. We decompose the error according to whether $r_t = 1$ or $r_t \geq 2$, control the two contributions via different analytical tools, and then carefully combine them to obtain an r_t-dependent regret bound. The result is summarized in Lemma 3.2, which provides, to the best of our knowledge, the *first* high-probability regret bound for non-oblivious PLA.

Lemma 3.2. *Let $\eta_{q,t}$ and γ_t be defined in* (11) *and* (14) *for* NPLA-EXP3. *Under Assumption 2.3, for any fixed $t \in \mathbb{N}^+$ and any sample-size process $\{r_j\}_{j\in[t]}$ generated according to Definition 3.1, with probability at least $1-\delta$,*

$$\text{Regret}'_{q,t} = O\left(\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}}\left(\sqrt{\log m} + \frac{\log\frac{1}{\delta}}{\sqrt{\log m}}\right)\right).$$

Remark: The above high-probability regret bound for non-oblivious PLA matches the order of the expected regret bound in the oblivious setting [Seldin et al., 2014]. For fixed $r_t \equiv r < m$, it achieves the optimal polynomial dependence on t and m/r up to logarithmic factors, since the lower bound for oblivious PLA [Seldin et al., 2014, Theorem 2] also applies to our setting. At the full-information endpoint $r = m$, the bound reduces to $O(\sqrt{t\log m})$, which matches the minimax optimal dependence for prediction with expert advice [Audibert and Bubeck, 2010, Table 1].

3.2 FTRL for the w-player

According to the decomposition in (7), we need to control the standard regret of the w-player. To obtain an anytime guarantee, one may follow prior work and use SMD with time-varying step sizes.

Algorithm 2 FTRL for non-oblivious OCO

Input: a scheduler $\mathcal{A}_s$

Initialize $\mathbf{F}_0 = \mathbf{0}_d$
for $t = 1, 2, \ldots$ **do**
  Update $\mathbf{w}_t$ according to (16) and send $\mathbf{w}_t$ to $\mathcal{A}_s$
  Receive $\mathbf{g}_w(\mathbf{w}_t, \mathbf{q}_t)$ in (17) from $\mathcal{A}_s$ and update $\mathbf{F}_t$ according to (19)
end for

However, existing analyses only control a weighted average regret [Zhang et al., 2026], while the standard regret may become linear under Assumption 2.1 [Orabona and Pál, 2018]. We therefore adopt FTRL with time-varying step sizes, which yields the desired standard regret guarantee.

To implement FTRL, we maintain the cumulative gradient vector $\mathbf{F}_t \in \mathbb{R}^d$, where d is the dimension of $\mathcal{W}$, and initialize $\mathbf{F}_0 = \mathbf{0}_d$. In each round t, we update $\mathbf{w}_t$ as

$$\mathbf{w}_t = \operatorname*{argmin}_{\mathbf{w} \in \mathcal{W}} \left\{ \left\langle \eta_{w,t} \mathbf{F}_{t-1}, \mathbf{w} \right\rangle + \nu_w(\mathbf{w}) \right\}, \tag{16}$$

where $\eta_{w,t} > 0$ is the step size. If we set $\nu_w(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|_2^2$, then (16) has a closed-form solution $\mathbf{w}_t = \Pi_{\mathcal{W}}\left[-\eta_{w,t}\mathbf{F}_{t-1}\right]$, where $\Pi_{\mathcal{W}}[\cdot]$ denotes the Euclidean projection onto $\mathcal{W}$. To construct an unbiased stochastic gradient for the w-player, we can reuse the sample $\mathbf{z}_t^{(c_t)} \sim \mathcal{P}_{c_t}$ queried by the q-player. Specifically, the w-player sends[2] $\mathbf{w}_t$ to $\mathcal{A}_s$, which then returns a stochastic gradient

$$\mathbf{g}_w(\mathbf{w}_t, \mathbf{q}_t) = \nabla \ell\left(\mathbf{w}_t; \mathbf{z}_t^{(c_t)}\right). \tag{17}$$

Under Assumption 2.2, $\mathbf{g}_w(\mathbf{w}_t, \mathbf{q}_t)$ is bounded, and since $c_t \sim \mathbf{q}_t$, it is an unbiased estimator of $\nabla_{\mathbf{w}} \phi(\mathbf{w}_t, \mathbf{q}_t)$:

$$\|\mathbf{g}_w(\mathbf{w}_t, \mathbf{q}_t)\|_{w,*} \overset{(4)}{\leq} G, \ \mathbb{E}_{t-1}\left[\mathbf{g}_w(\mathbf{w}_t, \mathbf{q}_t)\right] = \nabla_{\mathbf{w}} \phi(\mathbf{w}_t, \mathbf{q}_t), \tag{18}$$

where $\mathbb{E}_{t-1}[\cdot]$ denotes the expectation conditioned on the randomness up to round $t-1$ and the value of r_t in round t. Then, we update $\mathbf{F}_t$ as

$$\mathbf{F}_t = \mathbf{F}_{t-1} + \mathbf{g}_w(\mathbf{w}_t, \mathbf{q}_t). \tag{19}$$

Algorithm 2 summarizes FTRL for non-oblivious OCO.

We then establish the regret bound for Algorithm 2 in Lemma 3.3. To address the challenge of non-obliviousness, we incorporate the "ghost iterate" technique introduced by Nemirovski et al. [2009] into the FTRL analysis.

Lemma 3.3. *Under Assumptions 2.1, 2.2 and 2.4, and setting $\eta_{w,t} = \frac{\sqrt{2}D}{\sqrt{5}G}\frac{1}{\sqrt{t}}$ for Algorithm 2, we have for any fixed $t \in \mathbb{N}^+$, with probability at least $1-\delta$,*

$$\mathrm{Regret}_{w,t} \leq DG\sqrt{t}\left(2\sqrt{10} + 8\sqrt{\ln\frac{1}{\delta}}\right). \tag{20}$$

Remark: The high-probability regret bound (20) is of the same order as the non-anytime bound established by Zhang et al. [2023, Theorem 5]. Due to different definitions of regret, our result is not comparable to the anytime weighted average regret bound of Zhang et al. [2026].

3.3 The meta-scheduler and FLEX-GDRO

We present the meta-scheduler $\mathcal{A}_s$ in Algorithm 3, which obtains r_t from the environment and draws samples for the updates of both players. By instantiating $\mathcal{A}_q$ and $\mathcal{A}_w$ in $\mathcal{A}_s$ as NPLA-EXP3 (Algorithm 1) and FTRL (Algorithm 2), respectively, we obtain FLEX-GDRO, an anytime algorithm that can process an arbitrary number of samples per round. We discuss its computational cost in Appendix E. Combining Lemmas 3.2 and 3.3 with the decomposition in (9), we establish the following optimization error bound.

[2]The "send" operation is notational and incurs no communication cost.

Algorithm 3 Meta-scheduler for GDRO with FSQ

Instantiate the OCO subroutine $\mathcal{A}_w$ and the PLA subroutine $\mathcal{A}_q$, and run them in parallel.

for $t = 1, 2, \dots$ **do**
  Receive r_t from the environment and send r_t to $\mathcal{A}_q$
  Receive C_t, c_t, and $\mathbf{q}_t$ from $\mathcal{A}_q$
  Receive $\mathbf{w}_t$ from $\mathcal{A}_w$
  Draw samples $\{\mathbf{z}_t^{(i)}\}_{i \in C_t}$
  Construct and send $\hat{s}_{t,i}$ for all $i \in C_t$ to $\mathcal{A}_q$
  Construct and send $\nabla \ell(\mathbf{w}_t; \mathbf{z}_t^{(c_t)})$ to $\mathcal{A}_w$
  Compute the averages $\bar{\mathbf{w}}_t$ and $\bar{\mathbf{q}}_t$ in (6)
end for

Theorem 3.4. *For* FLEX-GDRO *with the step sizes in Lemmas 3.2 and 3.3, under Assumptions 2.1–2.4, for any fixed $t \in \mathbb{N}^+$ and any sample-size process $\{r_j\}_{j \in [t]}$ generated according to Definition 3.1, with probability at least $1 - \delta$,*

$$\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) = O\left(\frac{1}{t}\sqrt{\sum_{j=1}^{t} \frac{m}{r_j}}\left(\sqrt{\log m} + \frac{\log \frac{1}{\delta}}{\sqrt{\log m}}\right)\right). \tag{21}$$

Remark: The high-probability bound in (21) does not require knowing the total number of rounds in advance. In practice, r_t may be revealed online according to available resources. Our result accommodates this setting and shows that larger sample sizes lead to smaller optimization error.

Remark: When the sample size is fixed, namely $r_j \equiv r$ for all $j \in [t]$, the upper bound (21) simplifies to $O(\sqrt{m \log(m)/(rt)})$, implying a sample complexity of $O(m \log(m)/\epsilon^2)$ for all $r \in [m]$. This matches the sample complexity of non-anytime GDRO algorithms in the extreme cases $r = 1$ [Zhang et al., 2023] and $r = m$ [Nemirovski et al., 2009], and nearly attains the lower bound $\Omega(m/\epsilon^2)$ [Soma et al., 2022].[3] Moreover, our rate improves upon the SMD-based anytime guarantees [Zhang et al., 2026] by a factor of $O(\log t)$, which stems from using FTRL with the averaged outputs (6), thus avoiding the accumulation of $\{1/j\}_{j=1}^t$. Finally, the sample size is extended from $r \in \{1, m\}$ to all $r \in [m]$.

We further strengthen Theorem 3.4 by deriving a time-uniform bound that holds for all $t \in \mathbb{N}^+$.

Theorem 3.5. *Let* FLEX-GDRO *be run with the step sizes specified in Lemmas 3.2 and 3.3. Let $\{r_t\}_{t \geq 1}$ be any sample-size process generated according to Definition 3.1. Under Assumptions 2.1–2.4, with probability at least $1 - \delta$, the following bound holds simultaneously for all $t \in \mathbb{N}^+$:*

$$\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) = O\left(\frac{1}{t}\sqrt{\sum_{j=1}^{t} \frac{m}{r_j}}\left(\sqrt{\log m} + \frac{\log \frac{\log t}{\delta}}{\sqrt{\log m}}\right)\right).$$

Remark: The time-uniform guarantee incurs a minor additional cost compared to the fixed-time bound in (21), increasing the factor from $O(\sqrt{\log m})$ to $O(\max\{\sqrt{\log m}, \log\log(t)/\sqrt{\log m}\})$. Compared with taking a union bound over all t with varying confidence levels, which incurs an $O(\sqrt{\log t})$ cost, our method leverages time-uniform concentration inequalities [Lee et al., 2025, Lee and Oh, 2025] to reduce the cost to $O(\sqrt{\log\log t})$.

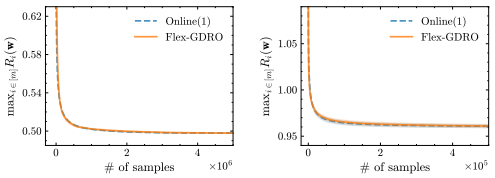
4 Experiments

In this section, we evaluate the sample and time efficiency of FLEX-GDRO in dynamic-resource and fixed-resource settings and examine the variance of its loss estimator.

4.1 Experimental Setup

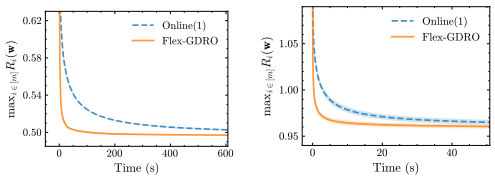
We consider four datasets: a synthetic binary dataset ($m = 20$), a multi-class diabetes dataset ($m = 12$) [Strack et al., 2014], CivilComments-WILDS ($m = 16$) [Koh et al., 2021], and FEMNIST

[3] This lower bound applies here because it is stated in terms of the total number of stochastic-oracle calls.



(a) The synthetic dataset (b) The diabetes dataset

Figure 1: Dynamic-resource setting: the maximum risk versus the number of samples.



(a) The synthetic dataset (b) The diabetes dataset

Figure 2: Dynamic-resource setting: the maximum risk versus the running time.

Table 1: Comparison of ONLINE(1) and FLEX-GDRO on the synthetic and diabetes datasets. The experiments use five independent runs. *Time* is measured in seconds. *Speedup* is the runtime ratio of ONLINE(1) to FLEX-GDRO for achieving the same risk. Gray shading highlights FLEX-GDRO.

Risk (Synthetic)	Time (ONLINE(1))	Time (FLEX-GDRO)	Speedup	Risk (Diabetes)	Time (ONLINE(1))	Time (FLEX-GDRO)	Speedup
0.525	98.192	9.107	10.78	0.99	4.842	0.569	8.50
0.52	129.708	12.129	10.69	0.985	6.479	0.876	7.39
0.515	175.201	15.152	11.56	0.98	9.406	1.236	7.61
0.51	256.080	21.191	12.08	0.975	13.495	1.902	7.09
0.505	432.405	33.274	13.00	0.97	22.839	3.488	6.55
0.5	1123.537	108.804	10.33	0.965	51.470	8.710	5.91

($m = 100$) [Cohen et al., 2017, Caldas et al., 2018]. We evaluate FLEX-GDRO on all four datasets under the dynamic-resource setting with $r_t \in [m-1]$, and additionally consider the fixed-resource setting with $r \in [m]$ on the synthetic and diabetes datasets. We also conduct synthetic experiments with larger numbers of groups to examine the variance of the loss estimator of FLEX-GDRO.

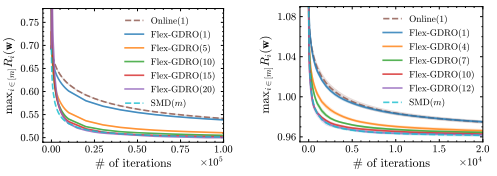
All tasks train a linear classifier with cross-entropy loss. We compare FLEX-GDRO (Algorithm 3 instantiated with Algorithms 1 and 2) with SMD(m) and ONLINE(1), which are anytime variants of Algorithms 1 and 2 in Zhang et al. [2026] and use m samples and 1 sample per round, respectively. For the synthetic and diabetes datasets, we report mean risk curves, with standard deviations shown as shaded regions. Further experimental details are provided in Appendix C.

4.2 Dynamic-resource Scenario

We compare FLEX-GDRO with ONLINE(1) on all four datasets under varying per-round sample budgets, where SMD(m) is not applicable without a buffering scheme.

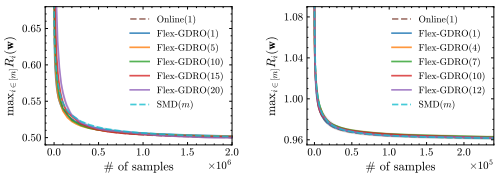
For the synthetic and diabetes datasets, we uniformly sample $r_t \in [1, m-1]$ in each round. Figs. 1 and 2 present the maximum risk versus the number of samples and the running time, respectively. As shown in Fig. 1, the risk curves of FLEX-GDRO and ONLINE(1) are closely aligned, suggesting similar sample complexities. Fig. 2 demonstrates that the risk of FLEX-GDRO decreases more rapidly than that of ONLINE(1). This is because FLEX-GDRO processes multiple samples per update, whereas ONLINE(1) is restricted to 1 sample. As a result, given a comparable sample complexity, FLEX-GDRO reaches the target risk in fewer updates and therefore in less wall-clock time. Table 1 further reports the runtime required to reach the same risk, showing that ONLINE(1) requires approximately 5 to 13 times longer than FLEX-GDRO. These results highlight the advantage of processing multiple samples simultaneously rather than handling them one by one.

On CivilComments-WILDS, we freeze a DistilBERT [Sanh et al., 2019] encoder, and train only a linear classification head. We consider $m = 16$ groups and uniformly sample $r_t \in [1, m-1]$ in each round. Table 2 reports results at five target risks. Across these thresholds, ONLINE(1) takes 8.40 to 11.20 times longer than FLEX-GDRO. Moreover, FLEX-GDRO uses fewer samples at risks 0.65 and 0.64. Implementation details are provided in Appendix C.2.



(a) The synthetic dataset (b) The diabetes dataset

Figure 3: Fixed-resource setting: the maximum risk versus the number of iterations.



(a) The synthetic dataset (b) The diabetes dataset

Figure 4: Fixed-resource setting: the maximum risk versus the number of samples.

Table 2: Comparison of ONLINE(1) and FLEX-GDRO on CivilComments-WILDS. *Time* is measured in seconds. *Speedup* is the runtime ratio of ONLINE(1) to FLEX-GDRO for achieving the same risk. Gray shading highlights FLEX-GDRO.

Risk	Samples (ONLINE(1))	Samples (FLEX-GDRO)	Time (ONLINE(1))	Time (FLEX-GDRO)	Speedup
0.68	61.35k	69.33k	40.26	4.79	8.40
0.67	67.55k	75.34k	44.34	5.20	8.52
0.66	77.75k	81.36k	51.03	5.62	9.08
0.65	97.81k	87.39k	64.19	6.03	10.64
0.64	120.00k	101.84k	78.80	7.03	11.20

We further evaluate FLEX-GDRO on FEMNIST with 100 writer groups under dynamic resources. At a maximum group risk of 0.94, it uses nearly the same number of samples as ONLINE(1) while requiring less optimization time. Detailed setups and results are provided in Appendix C.3.

4.3 Fixed-resource Scenario

We consider a scenario with a fixed number of samples $r \in \{1, 5, 10, 15, 20\}$ for the synthetic dataset and $r \in \{1, 4, 7, 10, 12\}$ for the diabetes dataset. For brevity, we denote FLEX-GDRO with r samples as FLEX-GDRO(r). Figs. 3 and 4 report the maximum risk versus the number of iterations and the number of queried samples, respectively. As shown in Fig. 3, larger r leads to faster risk reduction, which is consistent with the optimization error bound of $O(\sqrt{m\log(m)/(rt)})$ for any fixed $r \in [m]$. Fig. 4 shows that the risk curves of FLEX-GDRO(r) nearly overlap across different values of r, in line with the theoretical sample complexity of $O(m\log(m)/\epsilon^2)$. Moreover, in Fig. 4, the risk curves of SMD(m) and ONLINE(1) closely match those of FLEX-GDRO(r) across different values of r, indicating similar sample complexities.

4.4 Estimator Variance

For synthetic experiments with $m \in \{50, 100, 200\}$ and fixed ratios $m/r \in \{5, 10, 25\}$, the $\mathbf{q}_t$-weighted variance of the loss estimator remains similar as m increases at a fixed ratio, and increases with m/r. This supports the conclusion that the regret-relevant variance is governed primarily by m/r_t. Appendix C.4 reports the estimation protocol and the results.

5 Conclusion and future work

This paper studies GDRO in a novel FSQ setting, where the per-round sample size may vary over time. We formulate the problem as a two-player game, with one player solving non-oblivious OCO via FTRL and the other solving non-oblivious PLA via our proposed NPLA-EXP3. By integrating the two components, we develop FLEX-GDRO, an anytime GDRO algorithm that supports FSQ, and derive optimization error bounds that depend on the per-round sample sizes. Future work could extend FSQ to other optimization formulations, such as MRO [Agarwal and Zhang, 2022, Zhang et al., 2024a], and explore its applications to federated learning [Mohri et al., 2019, Deng et al., 2020, Bai et al., 2026b], recommendation systems [Wang et al., 2025], and large language models [Xie et al., 2023, Qiu et al., 2024, Ramesh et al., 2024, Panaganti et al., 2026].

Acknowledgments and Disclosure of Funding

This work was partially supported by National Science and Technology Major Project (2022ZD0114801), NSFC (62361146852), the "111 Center" (No. B26023), and the Collaborative Innovation Center of Novel Software Technology and Industrialization.

The authors declare no competing interests.

References

Alekh Agarwal and Tong Zhang. Minimax regret optimization for robust machine learning under distribution shift. In *Proceedings of the 35th Conference on Learning Theory*, pages 2704–2729, 2022.

Amine Allouah, Christian Kroer, Xuan Zhang, Vashist Avadhanula, Nona Bohanon, Anil Dania, Caner Gocmen, Sergey Pupyrev, Parikshit Shah, Nicolas Stier-Moses, and Ken Rodríguez Taarup. Fair allocation over time, with applications to content moderation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 25–35. Association for Computing Machinery, 2023.

Noga Alon, Nicolò Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Proceedings of the 28th Conference on Learning Theory*, volume 40, pages 23–35, 2015.

Kareem Amin, Satyen Kale, Gerald Tesauro, and Deepak Turaga. Budgeted prediction with expert advice. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1):2490–2496, 2015.

Jean-Yves Audibert and Sébastien Bubeck. Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research*, 11(94):2785–2836, 2010.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2014.

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, 2002.

Haomin Bai, Junyan Sun, Quan Cheng, Yuanyu Wan, and Lijun Zhang. Group distributionally robust optimization over decentralized networks. *ArXiv e-prints*, arXiv:2610.29261, 2026a.

Haomin Bai, Junyan Sun, Sifan Yang, Bo Xue, and Lijun Zhang. Communication-efficient agnostic federated learning via faster convergence and compression. *ArXiv e-prints*, arXiv:2609.36610, 2026b.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Sebastian Caldas, Sai Meher Karthik Duddu, Peter Wu, Tian Li, Jakub Konečný, H. Brendan McMahan, Virginia Smith, and Ameet Talwalkar. LEAF: A benchmark for federated settings. *ArXiv e-prints*, arXiv:1812.01097, 2018.

Yair Carmon and Danielle Hausler. Distributionally robust optimization via ball oracle acceleration. In *Advances in Neural Information Processing Systems 35*, pages 35866–35879, 2022.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Gregory Cohen, Saeed Afshar, Jonathan Tapson, and André van Schaik. EMNIST: Extending MNIST to handwritten letters. In *2017 International Joint Conference on Neural Networks*, pages 2921–2926, 2017.

Ashok Cutkosky. Anytime online-to-batch, optimism and acceleration. In *Proceedings of the 36th International Conference on Machine Learning*, pages 1446–1454, 2019.

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. Distributionally robust federated averaging. In *Advances in Neural Information Processing Systems 33*, pages 15111–15122, 2020.

John C. Duchi and Hongseok Namkoong. Learning models with uniform performance via distributionally robust optimization. *The Annals of Statistics*, 49(3):1378–1406, 2021.

David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3(1):100–118, 1975.

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Joao Gama, Indre Zliobaite, Albert Bifet, Mykol Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4):1–37, 2014.

Rajiv Gandhi, Samir Khuller, Srinivasan Parthasarathy, and Aravind Srinivasan. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM*, 53(3):324–360, 2006.

A. Gouissem, Z. Chkirbene, and R. Hamila. A comprehensive survey on client selections in federated learning. In *Innovation and Technological Advances for Sustainability*, pages 417–428. CRC Press, 2024.

Nika Haghtalab, Michael I. Jordan, and Eric Zhao. On-demand sampling: Learning optimally from multiple distributions. In *Advances in Neural Information Processing Systems 35*, pages 406–419, 2022.

Satyen Kale. Multiarmed bandits with limited expert advice. In *Proceedings of the 27th Conference on Learning Theory*, pages 107–122, 2014.

Tomáš Kocák, Gergely Neu, Michal Valko, and Rémi Munos. Efficient learning by implicit exploration in bandit problems with side observations. In *Advances in Neural Information Processing Systems 27*, pages 613–621, 2014.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, Sara M Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. WILDS: A benchmark of in-the-wild distribution shifts. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 5637–5664, 2021.

Fang Kong, Yichi Zhou, and Shuai Li. Simultaneously learning stochastic and adversarial bandits with general graph feedback. In *Proceedings of the 39th International Conference on Machine Learning*, pages 11473–11482, 2022.

Kun Kuang, Peng Cui, Susan Athey, Ruoxuan Xiong, and Bo Li. Stable prediction across unknown environments. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1617–1626, 2018.

Harin Lee and Min-hwan Oh. Minimax optimal reinforcement learning with quasi-optimism. In *International Conference on Learning Representations*, pages 13184–13231, 2025.

Harin Lee, Taehyun Hwang, and Min-hwan Oh. Lasso bandit with compatibility condition on optimal arm. In *International Conference on Learning Representations*, pages 9907–9964, 2025.

Alexander Libin, Jonah Treitler, Tadas Vasaitis, and Yijun Shao. Evaluating and reducing subgroup disparity in ai models predicting pediatric covid-19 test outcomes. In *2024 IEEE International Conference on Big Data*, pages 5033–5042, 2024.

Shiyin Lu, Yao Hu, and Lijun Zhang. Stochastic bandits with graph feedback in non-stationary environments. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10):8758–8766, 2021a.

Shiyin Lu, Guanghui Wang, and Lijun Zhang. Stochastic graphical bandits with adversarial corruptions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10):8749–8757, 2021b.

Haipeng Luo and Robert Schapire. Towards minimax online learning with unknown time horizon. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 226–234, 2014.

Haipeng Luo, Hanghang Tong, Mengxiao Zhang, and Yuheng Zhang. Improved high-probability regret for adversarial bandits with time-varying feedback graphs. In *Proceedings of the 34th International Conference on Algorithmic Learning Theory*, volume 201, pages 1074–1100, 2023.

Shie Mannor and Ohad Shamir. From bandits to experts: On the value of side-observations. In *Advances in Neural Information Processing Systems 24*, pages 684–692, 2011.

Mehryar Mohri, Gary Sivek, and Ananda Theertha Suresh. Agnostic federated learning. In *Proceedings of the 36th International Conference on Machine Learning*, pages 4615–4625, 2019.

Hongseok Namkoong and John C Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. In *Advances in Neural Information Processing Systems 29*, pages 2208–2216, 2016.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

Gergely Neu. Explore no more: Improved high-probability regret bounds for non-stochastic bandits. In *Advances in Neural Information Processing Systems 28*, pages 3168–3176, 2015.

Gergely Neu and Gábor Bartók. Importance weighting without importance weights: An efficient algorithm for combinatorial semi-bandits. *Journal of Machine Learning Research*, 17(154):1–21, 2016.

Quan M. Nguyen, Nishant A Mehta, and Cristóbal A Guzmán. Beyond minimax rates in group distributionally robust optimization via a novel notion of sparsity. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 46073–46115, 2025.

Takayuki Nishio and Ryo Yonetani. Client selection for federated learning with heterogeneous resources in mobile edge. In *ICC 2019 - 2019 IEEE International Conference on Communications*, pages 1–7, 2019.

Francesco Orabona. A modern introduction to online learning. *ArXiv e-prints*, arXiv:1912.13213v6, 2019.

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

Yonatan Oren, Shiori Sagawa, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust language modeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 4227–4237, 2019.

Kishan Panaganti, Zhenwen Liang, Wenhao Yu, Haitao Mi, and Dong Yu. Group distributionally robust optimization-driven RL for LLM reasoning. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

Stephen R. Pfohl, Haoran Zhang, Yizhe Xu, Agata Foryciarz, Marzyeh Ghassemi, and Nigam H. Shah. A comparison of approaches to improve worst-case predictive model performance over patient subpopulations. *Scientific Reports*, 12(1):3254, 2022.

Zi-Hao Qiu, Siqi Guo, Mao Xu, Tuo Zhao, Lijun Zhang, and Tianbao Yang. To cool or not to cool? temperature network meets large foundation models via DRO. In *Proceedings of the 41st International Conference on Machine Learning*, pages 41604–41643, 2024.

Hamed Rahimian and Sanjay Mehrotra. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3:1–85, 2022.

Shyam Sundhar Ramesh, Yifan Hu, Iason Chaimalas, Viraj Mehta, Pier Giuseppe Sessa, Haitham Bou Ammar, and Ilija Bogunovic. Group robust preference optimization in reward-free RLHF. In *Advances in Neural Information Processing Systems 37*, pages 37100–37137, 2024.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

Christophe Roux, Elias Wirth, Sebastian Pokutta, and Thomas Kerdreux. Efficient online-bandit strategies for minimax learning problems. *ArXiv e-prints*, arXiv:2105.13939, 2021.

El Mehdi Saad and Gilles Blanchard. Fast rates for prediction with limited expert advice. In *Advances in Neural Information Processing Systems 34*, pages 23582–23591, 2021.

El Mehdi Saad and Gilles Blanchard. Constant regret for sequence prediction with limited advice. In *Proceedings of the 34th International Conference on Algorithmic Learning Theory*, pages 1343–1386, 2023.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *International Conference on Learning Representations*, 2020.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv e-prints*, arXiv:1910.01108, 2019.

Yevgeny Seldin, Peter Bartlett, Koby Crammer, and Yasin Abbasi-Yadkori. Prediction with limited advice and multiarmed bandits with paid observations. In *Proceedings of the 31st International Conference on Machine Learning*, pages 280–287, 2014.

Agnieszka Słowik and Léon Bottou. On distributionally robust optimization and data rebalancing. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, pages 1283–1297, 2022.

Tasuku Soma, Khashayar Gatmiry, and Stefanie Jegelka. Optimal algorithms for group distributionally robust optimization and beyond. *ArXiv e-prints*, arXiv:2212.13669, 2022.

Beata Strack, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore. Impact of hba1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed Research International*, 2014(1):781670, 2014.

Masashi Sugiyama, Matthias Krauledat, and Klaus-Robert Müller. Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8(35):985–1005, 2007.

Tobias Sommer Thune and Yevgeny Seldin. Adaptation to easy data in prediction with limited advice. In *Advances in Neural Information Processing Systems 31*, pages 2909–2918, 2018.

Taishi Uchiya, Atsuyoshi Nakamura, and Mineichi Kudo. Algorithms for adversarial bandit problems with multiple plays. In *Proceedings of the 21st International Conference on Algorithmic Learning Theory*, pages 375–389, 2010.

Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, second edition, 2000.

Nuri Mert Vural, Hakan Gokcesu, Kaan Gokcesu, and Suleyman S. Kozat Kozat. Minimax optimal algorithms for adversarial bandit problem with multiple plays. *IEEE Transactions on Signal Processing*, 67(16):4383–4398, 2019.

Haoran Wang, Qiuye Jin, Shiman Li, Siyu Liu, Manning Wang, and Zhijian Song. A comprehensive survey on deep active learning in medical image analysis. *Medical Image Analysis*, 95:103201, 2024.

Yibo Wang, Yingchun Jian, Wenhao Yang, Shiyin Lu, Lei Shen, Bing Wang, Xiaoyi Zeng, and Lijun Zhang. Towards unbiased information extraction and adaptation in cross-domain recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(12):12757–12765, 2025.

Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy S Liang, Quoc V Le, Tengyu Ma, and Adams Wei Yu. Doremi: Optimizing data mixtures speeds up language model pretraining. In *Advances in Neural Information Processing Systems 36*, pages 69798–69818, 2023.

Meng Ye, Karan Sikka, Katherine Atwell, Sabit Hassan, Ajay Divakaran, and Malihe Alikhani. Multilingual content moderation: A case study on Reddit. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3828–3844. Association for Computational Linguistics, 2023.

Dingzhi Yu, Yunuo Cai, Wei Jiang, and Lijun Zhang. Efficient algorithms for empirical group distributionally robust optimization and beyond. In *Proceedings of the 41st International Conference on Machine Learning*, pages 57384–57414, 2024.

Donggyu Yun, Alexandre Proutiere, Sumyeong Ahn, Jinwoo Shin, and Yung Yi. Multi-armed bandit with additional observations. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2(1):1–22, 2018.

Lijun Zhang, Peng Zhao, Zhenhua Zhuang, Tianbao Yang, and Zhi-Hua Zhou. Stochastic approximation approaches to group distributionally robust optimization. In *Advances in Neural Information Processing Systems 36*, pages 52490–52522, 2023.

Lijun Zhang, Haomin Bai, Wei-Wei Tu, Ping Yang, and Yao Hu. Efficient stochastic approximation of minimax excess risk optimization. In *Proceedings of the 41st International Conference on Machine Learning*, pages 58599–58630, 2024a.

Lijun Zhang, Haomin Bai, Peng Zhao, and Zhi-Hua Zhou. Stochastic approximation approaches to group distributionally robust optimization and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(10):12880–12891, 2026.

Mengxiao Zhang, Yuheng Zhang, Haipeng Luo, and Paul Mineiro. Efficient contextual bandits with uninformed feedback graphs. In *Proceedings of the 41st International Conference on Machine Learning*, pages 60336–60358, 2024b.

Xiongtao Zhang, Xiaomin Zhu, Ji Wang, Hui Yan, Huangke Chen, and Weidong Bao. Federated learning with adaptive communication compression under dynamic bandwidth and unreliable networks. *Information Sciences*, 540:242–262, 2020.

A Representative Applications of FSQ

In this section, we present three representative applications of the FSQ setting. In all three applications, the per-round sample size r_t is externally determined by potentially time-varying resource constraints, while the learner can still choose which groups to query.

Federated learning in mobile edge computing (MEC). In MEC-based federated learning, client participation is coordinated by the server through centralized resource allocation [Nishio and Yonetani, 2019, Gouissem et al., 2024]. To improve robustness and fairness across clients, one may adopt agnostic federated learning [Mohri et al., 2019], whose objective coincides with GDRO when each client is treated as a group. Since wireless resources are limited and communication conditions may fluctuate over time, for example due to bandwidth variations [Zhang et al., 2020], only a finite and potentially time-varying number of clients can participate in each round. In this case, we view the number of clients that the server can query in round t as r_t.

Multilingual content moderation. In multilingual content moderation, the massive volume of online content makes an efficient moderation system essential [Ye et al., 2023, Allouah et al., 2023]. A practical goal is to train a model that predicts whether content complies with platform policies. In this setting, human judgment is indispensable for providing high-quality labels for model improvement. Since data availability differs substantially across languages, especially between English and other languages, it is necessary to balance overall accuracy and fairness [Ye et al., 2023]. Under this objective, a GDRO formulation over language groups is well motivated. Meanwhile, human

review capacity is limited and time-varying due to changing workloads, reviewer/partner capacities, and staffing requirements [Allouah et al., 2023]. We therefore model the number of labels that can be queried in round t as r_t, and let the platform allocate these label queries across language groups. Given the large scale of the data, it is reasonable to assume that a sample is available from each group in every round.

Medical image analysis. In medical image analysis, the goal is to train a deep model from a pool of unlabeled medical images. When the model is trained using empirical risk minimization, performance may vary substantially across patient subgroups, raising concerns about subgroup disparity and fairness in downstream clinical use [Pfohl et al., 2022, Libin et al., 2024]. It is therefore necessary to account for both overall predictive performance and subgroup robustness. Under this objective, applying GDRO over patient subgroups, such as those defined by race, sex, and age, is a natural choice [Pfohl et al., 2022]. Since expert annotation is expensive and time-consuming [Wang et al., 2024], only a limited number of annotations can be obtained in each round. We therefore model the annotation capacity in round t as r_t, and allow it to vary over time to reflect changing operating conditions. Given that the unlabeled pool is typically much larger than the available expert capacity, it is reasonable to assume that each group has an available instance in every round.

B Related Work

In this section, we briefly review GDRO and PLA.

B.1 Group Distributionally Robust Optimization

GDRO is motivated by applications affected by distribution shift, including federated learning [Mohri et al., 2019], language modeling [Oren et al., 2019], and neural network training [Sagawa et al., 2020]. A key algorithmic objective is to minimize the *sample complexity*, i.e., the number of samples required to attain an ϵ-optimal solution. Soma et al. [2022] establish a lower bound of $\Omega(m/\epsilon^2)$, and state-of-the-art methods achieve a near-optimal bound of $O(m\log(m)/\epsilon^2)$. These methods fall into two sampling paradigms: (i) m-sample methods that query one sample from each of the m distributions per round, e.g., stochastic mirror descent (SMD) [Nemirovski et al., 2009]; (ii) 1-sample methods that draw a single sample per round, including gradient-clipped SMD [Carmon and Hausler, 2022] and online-learning-based two-player game formulations [Haghtalab et al., 2022, Soma et al., 2022, Zhang et al., 2023]. Beyond sample efficiency, another practical consideration is whether an algorithm is *anytime* [Cutkosky, 2019, Luo and Schapire, 2014]. In contrast to the above non-anytime methods, which require the number of iterations to be specified in advance, Zhang et al. [2026] propose anytime GDRO algorithms with optimization error of $O(\log(t)\sqrt{m\log(m)/(rt)})$ for $r \in \{1, m\}$. Overall, existing GDRO methods are restricted to the two extremes of processing either 1 or m samples per round, leaving the intermediate and variable sample sizes under-explored.

Several works also explore GDRO under additional assumptions or different formulations. Nguyen et al. [2025] exploit sparsity to reduce the dependence on m in the sample complexity. Agarwal and Zhang [2022] introduce minimax regret optimization (MRO) to prevent the worst-case objective from being dominated by a single distribution, which can be viewed as a special case of calibrated DRO [Słowik and Bottou, 2022]. Subsequent studies develop efficient algorithms for MRO in finite-distribution settings [Zhang et al., 2024a, Yu et al., 2024] and algorithms for GDRO in decentralized settings [Bai et al., 2026a].

B.2 Prediction with Limited Advice

PLA [Seldin et al., 2014] can be regarded as an intermediate case between multi-armed bandit [Robbins, 1952, Auer et al., 2002], where the player selects 1 expert per round and observes its advice, and prediction with expert advice [Freund and Schapire, 1997], where the player observes the advice of all m experts. In PLA, the player can query the advice of $r_t \in [m]$ experts in round t, and the loss is evaluated on a subset of the selected experts [Seldin et al., 2014, Saad and Blanchard, 2023]. Seldin et al. [2014] study oblivious PLA with varying sample sizes and establish an expected regret bound of $O\left(\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}\log m}\right)$. Subsequent work investigates PLA across multiple settings, including access to distributions over experts for selection [Kale, 2014], non-uniform

expert costs [Amin et al., 2015], fixed per-round sample sizes [Yun et al., 2018], stochastic loss sequences, or adversarial ones with a small loss range [Thune and Seldin, 2018], excess generalization error in stochastic settings [Saad and Blanchard, 2021], and convex-combination predictions over experts [Saad and Blanchard, 2023].

Despite extensive studies, existing PLA algorithms are inapplicable to our setting, which requires minimizing regret against a non-oblivious adversary with varying sample sizes, because prior work assumes stochastic adversaries [Saad and Blanchard, 2021], oblivious adversaries [Seldin et al., 2014, Saad and Blanchard, 2023], fixed sample sizes [Thune and Seldin, 2018, Yun et al., 2018] or focuses on pseudo-regret objectives [Kale, 2014, Amin et al., 2015].

Beyond PLA, another line of work allowing multiple observations per round studies graph-structured partial feedback, where selecting one action may reveal the losses of neighboring actions through a feedback graph [Mannor and Shamir, 2011, Alon et al., 2015, Lu et al., 2021a,b, Kong et al., 2022, Luo et al., 2023, Zhang et al., 2024b].

C Experimental Details

All the experiments are implemented in Python and conducted on a server running Ubuntu 20.04, equipped with dual Intel(R) Xeon(R) Platinum 8358P CPUs (2.60 GHz).

During training, synthetic samples are generated on the fly, while samples from diabetes, CivilComments-WILDS, and FEMNIST are drawn uniformly with replacement within each queried group. During evaluation, we estimate each group's expected loss by an empirical average and report the maximum over groups. Since NPLA-EXP3 requires bounded loss feedback, we clip and normalize the cross-entropy loss used in the q-player update in all experiments. Let $\ell_{\mathrm{ce}}(w; z)$ denote the cross-entropy loss. We define

$$\ell_C(w; z) = \frac{\min\{\ell_{\mathrm{ce}}(w; z), C\}}{C},$$

where the clipping threshold is set to $C = 5$. The q-player then forms the surrogate loss as $\hat{s}_{t,i} = 1 - \ell_C(w_t; z_t^{(i)})$, so its feedback lies in $[0, 1]$. For the w-player update and evaluation, we use the original, unclipped cross-entropy loss.

Data and model licenses. Diabetes 130-US Hospitals [Strack et al., 2014][4], CivilComments-WILDS [Koh et al., 2021][5], and the `distilbert-base-uncased` checkpoint [Sanh et al., 2019][6] are distributed under CC BY 4.0, CC0 1.0, and Apache 2.0, respectively. Federated EMNIST [Cohen et al., 2017, Caldas et al., 2018][7] is provided through TensorFlow Federated and is derived from the LEAF preprocessing of EMNIST. The TensorFlow Federated and LEAF codebases are released under Apache 2.0 and BSD 2-Clause, respectively.

C.1 Synthetic and Diabetes Datasets

We construct the synthetic dataset following prior work [Namkoong and Duchi, 2016, Soma et al., 2022, Zhang et al., 2023, 2024a, 2026]. Specifically, we consider $m = 20$ distributions, each associated with a ground-truth classifier $\mathbf{w}_i^* \in \mathbb{R}^{500}$. The set $\{\mathbf{w}_i^*\}_{i\in[m]}$ is constructed following the approach outlined for the synthetic dataset by Zhang et al. [2026]. For each distribution, we randomly generate samples $(\mathbf{x}, y)$, where $\mathbf{x}$ is drawn from the standard normal distribution $\mathcal{N}(0, I)$ and y is generated as $y = \mathrm{sign}(\mathbf{x}^\top \mathbf{w}_i^*)$ with probability 0.9, or $y = -\mathrm{sign}(\mathbf{x}^\top \mathbf{w}_i^*)$ with probability 0.1.

For diabetes, we employ the dataset of Strack et al. [2014] to perform a multi-class classification task. The goal is to predict a patient's hospital readmission status, categorized as readmitted within 30 days, readmitted after 30 days, or not readmitted, based on treatment-related features. The original dataset contains 101,766 instances and 47 attributes. After preprocessing by removing three

[4] `https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008`

[5] `https://wilds.stanford.edu/datasets/#civilcomments`

[6] `https://huggingface.co/distilbert/distilbert-base-uncased`

[7] `https://www.tensorflow.org/federated/api_docs/python/tff/simulation/datasets/emnist`

Table 3: Comparison of ONLINE(1) and FLEX-GDRO on FEMNIST with 100 writer groups under dynamic resources. Entries are mean ± sample standard deviation over five seeds at the first saved checkpoint reaching each threshold. *Time* is measured in seconds for head optimization. *Speedup* is the ratio of mean optimization times of ONLINE(1) to FLEX-GDRO for achieving the same risk. Gray shading highlights FLEX-GDRO.

Risk	Samples ($\times 10^3$) (ONLINE(1))	Samples ($\times 10^3$) (FLEX-GDRO)	Time (ONLINE(1))	Time (FLEX-GDRO)	Speedup
0.98	52.06 ± 0.65	63.43 ± 3.33	78.45 ± 0.90	23.61 ± 0.94	3.32
0.95	71.22 ± 1.19	73.78 ± 3.81	106.50 ± 1.69	27.46 ± 1.18	3.88
0.94	79.53 ± 0.30	79.51 ± 5.01	118.76 ± 0.54	29.48 ± 1.65	4.03
0.93	89.73 ± 1.68	83.68 ± 6.89	133.66 ± 2.70	31.00 ± 2.37	4.31

features with predominantly missing values, three highly sparse features unsuitable for one-hot encoding, and two irrelevant features, we obtain a refined dataset of 99,492 instances with no missing values and a 102-dimensional expanded feature representation. We partition the dataset into $m = 12$ groups based on the combination of three sensitive attributes: race {Caucasian, African American, Other}, age $\{\leq 60, > 60\}$, and gender {Male, Female}. Each distribution $\mathcal{P}_i$ corresponds to the empirical distribution of the samples in the i-th group.

C.2 CivilComments-WILDS

We use the official training, validation, and test splits of CivilComments-WILDS. The task is binary toxicity classification, with groups formed by crossing eight identity attributes with the binary toxicity label. We use the frozen `distilbert-base-uncased` encoder and precompute its representations once. The reported wall-clock time measures only the subsequent optimization of the linear classification head. The reported run uses $30{,}000$ rounds. Both methods use the same initialization and r_t sequence, with ONLINE(1) processing samples sequentially. Hyperparameters for each method are selected from the same fixed candidate set using validation performance.

C.3 FEMNIST with 100 Writer Groups

For the FEMNIST experiments, we use the TensorFlow Federated (TFF) partition of EMNIST for 62-class handwritten-character classification [Cohen et al., 2017, Caldas et al., 2018]. The `train` and `test` splits contain the same $3{,}400$ writers, with $671{,}585$ training samples and $77{,}483$ test samples. We rank writers by their number of training samples and select the top 100 as the target groups. The non-target pool consists of all $3{,}300$ remaining writers, which we split into $2{,}970$ writers for CNN training and 330 writers for CNN validation, corresponding to a $90\%/10\%$ split at the writer level. We pretrain a two-layer CNN using ERM on the $2{,}970$ writers. After removing the final classification layer, we freeze the encoder and train only a linear softmax head on the target groups with cross-entropy loss, retaining a convex optimization objective for the head.

We run FLEX-GDRO for $10{,}000$ rounds, with $r_t = \lceil 100\alpha_t \rceil$ and α_t drawn uniformly from $\{0.02, 0.05, 0.10, 0.20\}$. We report maximum test risk over the target writers. Table 3 gives the mean and sample standard deviation. At risk 0.94, the methods require nearly the same number of samples, while FLEX-GDRO takes less optimization time. At risk 0.93, it uses fewer samples as well. The optimization-time ratios range from 3.32 to 4.31 across the reported thresholds. These results show that FLEX-GDRO maintains competitive sample efficiency and a clear optimization-time advantage with 100 groups.

C.4 Empirical Estimator Variance

The adaptive loss estimator $\tilde{\mathbf{s}}_t$ serves as the stochastic gradient of the q-player. Its entropy-FTRL regret is controlled by the $\mathbf{q}_t$-weighted second moment $M_{q,t} := \sum_{i=1}^m q_{t,i} \mathbb{E}_{t-1}[\tilde{s}_{t,i}^2]$. For $r_t \geq 2$, the estimator analysis in (13) gives $M_{q,t} \leq \frac{m}{r_t}$. Hence, the corresponding $\mathbf{q}_t$-weighted variance $V_{q,t} := \sum_{i=1}^m q_{t,i} \operatorname{Var}_{t-1}(\tilde{s}_{t,i})$ satisfies $V_{q,t} \leq M_{q,t} \leq m/r_t$. Thus, the variance quantity relevant to the regret analysis depends on the observation ratio m/r_t, rather than on m alone.

Table 4: Empirical $\mathbf{q}_t$-weighted variance $\widehat{V}_{q,t}$ of the loss estimator on synthetic data. Entries are mean $\pm$ standard deviation over five seeds.

m/r	m	r	$T = 1{,}000$	$T = 5{,}000$	$T = 10{,}000$	$T = 20{,}000$
5	50	10	2.476 ± 0.133	2.586 ± 0.081	2.680 ± 0.092	2.688 ± 0.130
5	100	20	2.502 ± 0.009	2.669 ± 0.038	2.694 ± 0.048	2.703 ± 0.073
5	200	40	2.525 ± 0.081	2.646 ± 0.038	2.691 ± 0.082	2.730 ± 0.107
10	50	5	5.464 ± 0.111	5.687 ± 0.160	5.932 ± 0.194	6.060 ± 0.225
10	100	10	5.538 ± 0.159	5.927 ± 0.115	5.925 ± 0.064	6.066 ± 0.090
10	200	20	5.751 ± 0.053	5.862 ± 0.147	6.043 ± 0.138	6.044 ± 0.214
25	50	2	14.242 ± 0.462	14.736 ± 0.414	15.469 ± 0.230	15.586 ± 0.812
25	100	4	14.603 ± 0.133	15.463 ± 0.354	15.794 ± 0.245	15.739 ± 0.477
25	200	8	15.038 ± 0.315	15.430 ± 0.299	15.730 ± 0.389	16.145 ± 0.490

Table 5: Unweighted mean per-coordinate variance $\widehat{V}_t$ of the loss estimator on synthetic data. Entries are mean $\pm$ standard deviation over five seeds.

m/r	m	r	$T = 1{,}000$	$T = 5{,}000$	$T = 10{,}000$	$T = 20{,}000$
5	50	10	2.537 ± 0.127	2.635 ± 0.068	2.725 ± 0.069	2.739 ± 0.136
5	100	20	2.549 ± 0.010	2.682 ± 0.054	2.719 ± 0.049	2.747 ± 0.076
5	200	40	2.534 ± 0.049	2.667 ± 0.041	2.693 ± 0.057	2.753 ± 0.109
10	50	5	5.597 ± 0.092	5.961 ± 0.239	6.146 ± 0.174	6.228 ± 0.250
10	100	10	5.690 ± 0.205	6.099 ± 0.140	6.088 ± 0.084	6.181 ± 0.153
10	200	20	5.801 ± 0.054	5.966 ± 0.091	6.134 ± 0.141	6.217 ± 0.232
25	50	2	15.895 ± 0.531	16.461 ± 0.626	17.147 ± 0.340	17.189 ± 0.888
25	100	4	15.296 ± 0.342	16.411 ± 0.267	16.531 ± 0.290	16.836 ± 0.514
25	200	8	15.560 ± 0.283	15.985 ± 0.259	16.306 ± 0.297	16.605 ± 0.567

To examine this dependence empirically, we conduct fixed-resource experiments on the synthetic dataset with $m \in \{50, 100, 200\}$ and $m/r \in \{5, 10, 25\}$. At checkpoints $t \in \{1{,}000, 5{,}000, 10{,}000, 20{,}000\}$, we fix $(\mathbf{w}_t, \mathbf{q}_t)$ and generate 500 independent realizations of $\tilde{\mathbf{s}}_t$. We estimate the weighted variance by $\widehat{V}_{q,t} := \sum_{i=1}^{m} q_{t,i} \widehat{\mathrm{Var}}(\tilde{s}_{t,i})$, where $\widehat{\mathrm{Var}}(\tilde{s}_{t,i})$ denotes the empirical variance. Table 4 reports the mean and standard deviation over five independent seeds.

The results are consistent with the $O(m/r)$ dependence in analysis. For example, with $r = 20$, increasing m from 100 to 200 doubles m/r from 5 to 10 and increases the weighted variance at $t = 20{,}000$ from 2.703 to 6.044. By contrast, when r grows proportionally with m so that $m/r = 5$, the values are 2.688, 2.703, and 2.730 for $m = 50$, 100, and 200, respectively. These results indicate that the regret-relevant estimator variance is primarily determined by m/r.

We also consider the unweighted average per-coordinate variance $V_t := \frac{1}{m} \sum_{i=1}^{m} \mathrm{Var}(\tilde{s}_{t,i})$, with empirical estimate $\widehat{V}_t := \frac{1}{m} \sum_{i=1}^{m} \widehat{\mathrm{Var}}(\tilde{s}_{t,i})$. This quantity directly measures the average coordinate-wise fluctuation as m increases. Table 5 shows the same qualitative dependence on m/r.

Table 6 further reports the mean per-coordinate variance of the w-player stochastic gradient. Across all considered values of m and r, the means at $t = 20{,}000$ range from 0.172 to 0.183. This is consistent with our design, since the w-player uses a single gradient sampled according to $\mathbf{q}_t$, whose norm bound and regret guarantee are independent of the per-round sample size.

D Limitations

Our FSQ framework focuses on settings where the available sample size may vary due to external resource constraints, while the learner is still allowed to decide which groups to query. This modeling captures applications such as centrally coordinated client selection, human review allocation, and expert annotation scheduling, as discussed in Appendix A. However, it does not cover settings in which workers or clients autonomously decide whether to join the learning process. Extending FSQ to such self-selected participation settings is an interesting direction for future work.

Table 6: Mean per-coordinate variance of the w-player stochastic gradient on synthetic data. Entries are mean $\pm$ standard deviation over five seeds.

m/r	m	r	$T = 1{,}000$	$T = 5{,}000$	$T = 10{,}000$	$T = 20{,}000$
5	50	10	0.220 ± 0.012	0.189 ± 0.009	0.180 ± 0.013	0.183 ± 0.011
5	100	20	0.212 ± 0.003	0.194 ± 0.008	0.179 ± 0.006	0.177 ± 0.012
5	200	40	0.218 ± 0.010	0.192 ± 0.007	0.190 ± 0.019	0.183 ± 0.007
10	50	5	0.213 ± 0.004	0.192 ± 0.012	0.180 ± 0.010	0.183 ± 0.015
10	100	10	0.218 ± 0.011	0.191 ± 0.012	0.180 ± 0.008	0.172 ± 0.012
10	200	20	0.213 ± 0.005	0.189 ± 0.005	0.180 ± 0.011	0.182 ± 0.009
25	50	2	0.218 ± 0.015	0.192 ± 0.018	0.184 ± 0.015	0.182 ± 0.009
25	100	4	0.218 ± 0.002	0.190 ± 0.012	0.179 ± 0.005	0.173 ± 0.014
25	200	8	0.220 ± 0.009	0.191 ± 0.008	0.182 ± 0.015	0.180 ± 0.010

E Computational cost of FLEX-GDRO

The per-round computational cost of FLEX-GDRO consists of three components: the q-player update, the w-player update, and group selection with sample acquisition. The first two components have the same order of complexity as existing GDRO methods, while the additional group-selection step requires $O(m)$ time per round. We detail these operations below.

1. **q-player update.** The exponential-weights update (12) requires $O(m)$ elementary operations. Accumulating the observed losses requires another $O(r_t)$ operations, so the overall cost is $O(m)$ since $r_t \leq m$. Entropy-based updates in existing GDRO methods have the same order of computational cost [Zhang et al., 2026].
2. **w-player update.** With Euclidean regularizer, the FTRL update (16) requires $O(d)$ arithmetic operations and one projection onto $\mathcal{W}$, in addition to computing one stochastic gradient, where d is the model dimension. This is comparable to a Euclidean OMD update, which also requires $O(d)$ arithmetic operations and one projection. The stochastic gradient reuses the exploitation sample and thus requires no additional sample query.
3. **Group selection and sample acquisition.** Drawing c_t from $\mathbf{q}_t$ and selecting the exploration set $\mathcal{I}_t$ can be implemented in $O(m)$ time, e.g., DepRound in Appendix F. Each round then acquires r_t samples and evaluates their losses, while computing only one stochastic gradient for the w-player.

F Supporting Algorithm: DepRound

The DepRound algorithm has been extensively employed in combinatorial semi-bandit algorithms [Uchiya et al., 2010, Vural et al., 2019, Roux et al., 2021]. DepRound takes as input a vector $\mathbf{p} \in \mathbb{R}^m$ satisfying $0 \leq p_i \leq 1$ and $\|\mathbf{p}\|_1 = k \in \mathbb{Z}^+$, and outputs a set $\mathcal{I}$ such that $\Pr[i \in \mathcal{I}] = p_i$ and $|\mathcal{I}| = k$, with a computational complexity $O(m)$ for both time and space. The DepRound procedure is outlined in Algorithm 4.

Algorithm 4 DepRound

Input: vector $\mathbf{p} \in \mathbb{R}^m$ that satisfies $\mathbf{0} \leq \mathbf{p} \leq \mathbf{1}$ and $\|\mathbf{p}\|_1 = k \in \mathbb{Z}^+$

while $\exists i \in [m]$ s.t. $p_i \in (0,1)$ **do**
  Choose any $i, j \in [m]$ with $i \neq j$ and $p_i \in (0,1), p_j \in (0,1)$
  Set $\alpha = \min\{1 - p_i, p_j\}$ and $\beta = \min\{p_i, 1 - p_j\}$
  Update

$$(p_i, p_j) = \begin{cases} (p_i + \alpha, p_j - \alpha) \text{ with probability } \frac{\beta}{\alpha+\beta} \\ (p_i - \beta, p_j + \beta) \text{ with probability } \frac{\alpha}{\alpha+\beta} \end{cases}$$

end while
return $\mathcal{I} = \{i \in [m] \mid p_i = 1\}$

G Analysis

In analysis, $\mathcal{F}_{t-1}$ contains the history through round $t-1$ and the currently revealed sample size r_t, before the learner's round-t randomization and fresh sampling. In particular, r_t, $\mathbf{w}_t$, $\mathbf{q}_t$, $\eta_{q,t}$, $\eta_{w,t}$, and γ_t are $\mathcal{F}_{t-1}$-measurable. We write $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_{t-1}]$.

G.1 Supporting Lemmas

Based on the discussion in Section 2, we first state an error-decomposition lemma.

Lemma G.1. *Under Assumptions 2.1-2.4, suppose that*

1. *The w-player is equipped with a non-oblivious OCO algorithm $\mathcal{A}_w$, and for any fixed $t \in \mathbb{N}^+$, with probability at least $1-\delta$, the regret $\mathrm{Regret}_{w,t}$ is upper bounded by $U_w(t,\delta)$.*
2. *The q-player is equipped with a non-oblivious PLA algorithm $\mathcal{A}_q$, and for any fixed $t \in \mathbb{N}^+$, with probability at least $1-\delta$, the regret $\mathrm{Regret}'_{q,t}$ is upper bounded by $U_q(t,\delta)$.*

Then, for any fixed $t \in \mathbb{N}^+$, with probability at least $1-\delta$, we have

$$\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) \le \frac{1}{t} U_w\left(t, \frac{\delta}{4}\right) + \frac{1}{t} U_q\left(t, \frac{\delta}{4}\right) + 2\sqrt{\frac{2}{t} \ln \frac{4m}{\delta}}. \tag{22}$$

Then, we give some supporting lemmas as follows.

Lemma G.2. *[Seldin et al., 2014, Lemma 7] For any probability distribution $\mathbf{q} \in \Delta_m$ and any $r \in [m]$, we have*

$$\sum_{i=1}^{m} \frac{q_i(m-1)}{q_i(m-r)+r-1} \le \frac{m}{r}. \tag{23}$$

Lemma G.3. *[Seldin et al., 2014, Lemma 8] For any sample-size process of non-negative numbers $a_1, a_2, \ldots, a_t$, such that $a_1 > 0$, and any power $\gamma \in (0,1)$ we have*

$$\sum_{j=1}^{t} \frac{a_j}{\left(\sum_{s=1}^{j} a_s\right)^\gamma} \le \frac{1}{1-\gamma} \left(\sum_{j=1}^{t} a_j\right)^{1-\gamma}. \tag{24}$$

We present Azuma's inequality and an adaptive-budget Bernstein inequality as follows.

Lemma G.4. *[Cesa-Bianchi and Lugosi, 2006, Lemma A.7] Let $V_1, V_2, \ldots$ be a martingale difference sequence with respect to some sequence $X_1, X_2, \ldots$ such that $V_i \in [A_i, A_i + c_i]$ for some random variable A_i, measurable with respect to $X_1, \ldots, X_{i-1}$ and a positive constant c_i. If $S_n = \sum_{i=1}^{n} V_i$, then for any $t > 0$,*

$$\Pr[S_n > t] \le \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right).$$

Lemma G.5. *Fix $t \in \mathbb{N}^+$, an integer $m \ge 2$, and $\delta \in (0,1)$. Let $\{X_j\}_{j=1}^t$ be a martingale difference sequence with $|X_j| \le b$ for $b > 0$. Let $\{r_j\}_{j=1}^t$ follow Definition 3.1. Suppose a fixed nonnegative and nondecreasing function $\Psi : (0,\infty) \to [0,\infty)$ satisfies, for every $B > 0$, $\sum_{j=1}^{t} \mathbb{E}_{j-1}\left[(X_j \mathbb{I}\left[\sum_{s=1}^{j} \frac{m}{r_s} \le B\right])^2\right] \le \Psi(B)$. Then, with probability at least $1-\delta$,*

$$\sum_{j=1}^{t} X_j \le \sqrt{2\Psi\left(2\sum_{j=1}^{t} \frac{m}{r_j}\right) \ln \frac{m}{\delta}} + \frac{2b}{3} \ln \frac{m}{\delta}. \tag{25}$$

The proof is given in Appendix G.7.

Note that both Lemmas G.4 and G.5 hold only for a fixed time horizon t. In the following, we present two time-uniform versions to establish a time-uniform guarantee.

Lemma G.6. *[Lee et al., 2025, Lemma 27] Let $\{X_t\}_{t=1}^{\infty}$ be a real-valued martingale difference sequence adapted to a filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. Assume that $\{X_t\}_{t=1}^{\infty}$ is conditionally σ-sub-Gaussian, i.e., $\mathbb{E}[e^{sX_t} \mid \mathcal{F}_{t-1}] \leq e^{\frac{s^2\sigma^2}{2}}$ for all $s \in \mathbb{R}$. Then, the following inequality holds with probability at least $1-\delta$, for all $n \in \mathbb{N}^+$:*

$$\sum_{t=1}^{n} X_t \leq 2^{\frac{3}{4}}\sigma\sqrt{n \ln \frac{7(\ln 2n)^2}{2\delta}}.$$

Although a time-uniform version of Bernstein's inequality is presented by Lee and Oh [2025, Lemma 37], we adopt the Theorem 3 from the author's blog [8] as it provides a simpler bound. For completeness, we provide its proof in Appendix G.8.

Lemma G.7. *Let $\{X_n\}_{n=1}^{\infty}$ be a martingale difference sequence with respect to a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$. Suppose $X_n \leq 1$ holds almost surely for all n. Let $T_n = \sum_{t=1}^{n} \mathbb{E}[X_t^2|\mathcal{F}_{t-1}]$. Then, the following inequality holds with probability at least $1-\delta$, for all $n \in \mathbb{N}^+$:*

$$\sum_{i=1}^{n} X_i \leq 2\ln\frac{2(1+\ln^+ T_n)^2}{\delta} + 2\sqrt{T_n \ln\frac{2(1+\ln^+ T_n)^2}{\delta}}$$

where $\ln^+ x = \max\{\ln x, 0\}$ for $x > 0$, with the convention $\ln^+ 0 = 0$.

G.2 Proof of Lemma 3.2

We first present the complete form of Lemma 3.2.

Lemma G.8. *Let $\eta_{q,t}$ and γ_t be defined in* (11) *and* (14) *for Algorithm 1. Under Assumption 2.3, we have for any fixed $t \in \mathbb{N}^+$ and any sample-size process $\{r_j\}_{j\in[t]}$ generated according to Definition 3.1, with probability at least $1-\delta$,*

$$\begin{aligned}\mathrm{Regret}'_{q,t} \leq& \sqrt{\sum_{j=1}^{t}\frac{m}{r_j}}\left(9\sqrt{2\ln m} + 4\sqrt{2\ln\frac{5}{\delta}}\right) + 2^{1/4}m\sqrt{\ln m\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}\ln\frac{5m}{\delta}}}\\ &+\left(m+1+\frac{m^2\sqrt{\ln m}}{3}+\sqrt{\frac{2\sum_{j=1}^{t}\frac{m}{r_j}}{\ln m}}\right)\ln\frac{5}{\delta}+\left(\frac{m^2\sqrt{\ln m}}{3}+\frac{5m}{3}\right)\ln m.\end{aligned}$$

By defining $\nu_{q,t}(\mathbf{q}) = \frac{1}{\eta_{q,t}}\left(\nu_q(\mathbf{q}) - \min_{\mathbf{q}\in\Delta_m}\nu_q(\mathbf{q})\right)$ with the negative-entropy regularizer $\nu_q(\mathbf{q}) = \sum_{i=1}^{m} q_i \log q_i$, the update rule in (12) can be reformulated as the standard FTRL update rule

$$\mathbf{q}_t = \operatorname*{argmin}_{\mathbf{q}\in\Delta_m}\left\{\langle \mathbf{L}_{t-1}, \mathbf{q}\rangle + \nu_{q,t}(\mathbf{q})\right\}.$$

Since $\eta_{q,t}$ is a non-increasing sequence, by the results from Orabona [2019, Remark 7.4., Lemma 7.14 and §7.5], we can derive for each $i \in [m]$,

$$\sum_{j=1}^{t}\left(\langle \mathbf{q}_j, \tilde{\mathbf{s}}_j\rangle - \tilde{s}_{j,i}\right) \leq \frac{\ln m}{\eta_{q,t}} + \sum_{j=1}^{t}\frac{\eta_{q,j}}{2}\sum_{i=1}^{m} q_{j,i}(\tilde{s}_{j,i})^2. \tag{26}$$

Using the notation $k^* \in \operatorname{argmin}_{i\in[m]}\sum_{j=1}^{t}\hat{s}_{j,i}$, $S_t = \{j | r_j = 1, j \in [t]\}$ and $M_t = \{j | r_j \geq 2, j \in [t]\}$, we have

$$\begin{aligned}\mathrm{Regret}'_{q,t} =& \sum_{j=1}^{t}\langle \mathbf{q}_j, \hat{\mathbf{s}}_j\rangle - \min_{i\in[m]}\sum_{j=1}^{t}\hat{s}_{j,i} = \sum_{j=1}^{t}\langle \mathbf{q}_j, \hat{\mathbf{s}}_j\rangle - \sum_{j=1}^{t}\hat{s}_{j,k^*}\\ =& \sum_{j=1}^{t}\left(\langle \mathbf{q}_j, \tilde{\mathbf{s}}_j\rangle - \tilde{s}_{j,k^*}\right) + \sum_{j=1}^{t}\left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j\rangle - \langle \mathbf{q}_j, \tilde{\mathbf{s}}_j\rangle\right) + \sum_{j=1}^{t}\left(\tilde{s}_{j,k^*} - \hat{s}_{j,k^*}\right)\\ \overset{(26)}{\leq}& \frac{\ln m}{\eta_{q,t}} + \sum_{j=1}^{t}\frac{\eta_{q,j}}{2}\sum_{i=1}^{m} q_{j,i}(\tilde{s}_{j,i})^2 + \sum_{j=1}^{t}\left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j\rangle - \langle \mathbf{q}_j, \tilde{\mathbf{s}}_j\rangle\right) + \sum_{j=1}^{t}\left(\tilde{s}_{j,k^*} - \hat{s}_{j,k^*}\right).\end{aligned} \tag{27}$$

[8] https://harinboy.github.io/posts/FreedmansInequality/

Then, we analyze the bias introduced by $\tilde{s}_{j,i}$ when $j \in S_t$. For $r_j = 1$, the selected distribution set $C_j = \{c_j\}$. Using Neu [2015, (5)], for $j \in S_t$, we have

$$\langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle = \sum_{i=1}^{m} \mathbb{I}\left[i = c_j\right] \frac{\hat{s}_{j,i}\left(q_{j,i} + \gamma_j\right)}{q_{j,i} + \gamma_j} - \gamma_j \sum_{i=1}^{m} \mathbb{I}\left[i = c_j\right] \frac{\hat{s}_{j,i}}{q_{j,i} + \gamma_j} = \hat{s}_{j,c_j} - \gamma_j \sum_{i=1}^{m} \tilde{s}_{j,i}. \quad (28)$$

By combining (28) with (27), and noting that $S_t \cup M_t = [t]$ and $q_{j,i}\tilde{s}_{j,i} \le 1$ for all $j \in S_t, i \in [m]$, we obtain

$$\begin{aligned}
\mathrm{Regret}'_{q,t} \le & \frac{\ln m}{\eta_{q,t}} + \sum_{j \in S_t} \frac{\eta_{q,j}}{2} \sum_{i=1}^{m} \tilde{s}_{j,i} + \sum_{j \in M_t} \frac{\eta_{q,j}}{2} \sum_{i=1}^{m} q_{j,i} (\tilde{s}_{j,i})^2 \\
& + \sum_{j \in S_t} \left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle\right) + \sum_{j \in M_t} \left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle\right) + \sum_{j=1}^{t} \left(\tilde{s}_{j,k^*} - \hat{s}_{j,k^*}\right) \\
\overset{(28)}{\le} & \frac{\ln m}{\eta_{q,t}} + \underbrace{\sum_{j \in S_t} \left(\frac{\eta_{q,j}}{2} + \gamma_j\right) \sum_{i=1}^{m} \tilde{s}_{j,i}}_{\texttt{term(d)}} + \underbrace{\sum_{j \in M_t} \frac{\eta_{q,j}}{2} \sum_{i=1}^{m} q_{j,i} (\tilde{s}_{j,i})^2}_{\texttt{term(e)}} \\
& + \underbrace{\sum_{j \in S_t} \left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \hat{s}_{j,c_j}\right) + \sum_{j \in M_t} \left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle\right)}_{\texttt{term(f)}} \\
& + \underbrace{\sum_{j \in S_t} \left(\tilde{s}_{j,k^*} - \hat{s}_{j,k^*}\right)}_{\texttt{term(g)}} + \underbrace{\sum_{j \in M_t} \left(\tilde{s}_{j,k^*} - \hat{s}_{j,k^*}\right)}_{\texttt{term(h)}}.
\end{aligned} \quad (29)$$

For $\texttt{term(d)}$, we invoke Lemma 2 in Zhang et al. [2023] and extend it to any subset of $[t]$ as follows.

Lemma G.9. *Let $\mathcal{F}_{j-1}$ denote the analysis history before round-j sampling, including all past loss vectors and the currently revealed sample size. Let $\mathbf{p}_j \in \Delta_m$ be $\mathcal{F}_{j-1}$-measurable, and let $\hat{\xi}_{j,i} \in [0,1]$ with $\Pr(i_j = i \mid \mathcal{F}_{j-1}, \hat{\xi}_{j,1}, \ldots, \hat{\xi}_{j,m}) = p_{j,i}$. Define the IX estimator $\tilde{\xi}_{j,i} = \frac{\hat{\xi}_{j,i}}{p_{j,i}+\gamma_j}\mathbb{I}[i_j = i]$. Let $\{\gamma_j\}_{j=1}^{t}$ be positive and non-increasing, with γ_j being $\mathcal{F}_{j-1}$-measurable for every $j \in [t]$, and let $\alpha_{j,i}$ be non-negative $\mathcal{F}_{j-1}$-measurable random variables satisfying $\alpha_{j,i} \le 2\gamma_j$. Let $S \subseteq [t]$ have $\mathcal{F}_{j-1}$-measurable indicators $\mathbb{I}[j \in S]$. Then, with probability at least $1-\delta$,*

$$\sum_{j \in S} \sum_{i=1}^{m} \alpha_{j,i} (\tilde{\xi}_{j,i} - \hat{\xi}_{j,i}) \le \ln \frac{1}{\delta}. \quad (30)$$

For our application, the full vector of fresh losses can be generated independently of the current index draw, conditional on the past and r_j, with only the queried coordinates revealed. Moreover, r_j is $\mathcal{F}_{j-1}$-measurable by Definition 3.1, so $\mathbb{I}[r_j = 1]$ is $\mathcal{F}_{j-1}$-measurable; γ_j and $\eta_{q,j}$ are also $\mathcal{F}_{j-1}$-measurable by (11) and (14). Thus, for $j \in S_t$, our construction satisfies the requirements of Lemma G.9. Setting $S = S_t$ and $\alpha_{j,i} = \eta_{q,j}$ for all $j \in [t]$ and $i \in [m]$, we have, with probability at least $1-\delta$,

$$\begin{aligned}
\sum_{j \in S_t} \left(\frac{\eta_{q,j}}{2} + \gamma_j\right) \sum_{i=1}^{m} \tilde{s}_{j,i} \overset{(14)}{=} \sum_{j \in S_t} \eta_{q,j} \sum_{i=1}^{m} \tilde{s}_{j,i} \overset{(30)}{\le} \sum_{j \in S_t} \eta_{q,j} \sum_{i=1}^{m} \hat{s}_{j,i} + \ln \frac{1}{\delta} \overset{(8)}{\le} \sum_{j \in S_t} \eta_{q,j} m + \ln \frac{1}{\delta} \\
\overset{(11)}{=} \sum_{j \in S_t} \sqrt{\frac{m \ln m}{\sum_{s=1}^{j} \frac{1}{r_s}}} + \ln \frac{1}{\delta} \le \sum_{j \in S_t} \sqrt{\frac{m \ln m}{|[j] \cap S_t|}} + \ln \frac{1}{\delta} \le 2\sqrt{|S_t| m \ln m} + \ln \frac{1}{\delta},
\end{aligned} \quad (31)$$

where the last inequality follows from the fact $\sum_{j=1}^{t} \frac{1}{\sqrt{j}} \le \int_0^t \frac{1}{\sqrt{x}} \mathrm{d}x = 2\sqrt{x}|_0^t = 2\sqrt{t}$.

For $\texttt{term(e)}$, we denote $Q_j = \frac{\eta_{q,j}}{2} \sum_{i=1}^{m} q_{j,i} (\tilde{s}_{j,i})^2$ for $j \in [t]$. For clarity and ease of reference, we present the following lemma.

Lemma G.10. *Under Assumption 2.3 and the sampling rule of Algorithm 1 with the sample-size protocol in Definition 3.1, let $\eta_{q,j}$ be defined in* (11)*, $Q_j = \frac{\eta_{q,j}}{2}\sum_{i=1}^{m} q_{j,i}\tilde{s}_{j,i}^2$, and $M_t = \{j \in [t] : r_j \geq 2\}$. Then, for any fixed $t \in \mathbb{N}_+$ and $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\sum_{j\in M_t} Q_j \leq 2^{1/4} m \sqrt{\ln m \sqrt{\sum_{j=1}^{t} \frac{m}{r_j} \ln \frac{m}{\delta}} + \frac{m^2\sqrt{\ln m}}{3} \ln \frac{m}{\delta}} + \frac{m}{2} \sum_{j\in M_t} \frac{\eta_{q,j}}{r_j}. \tag{32}$$

By setting $\gamma = \frac{1}{2}$ in Lemma G.3 and applying it to the subsequence indexed by M_t, we have

$$\begin{aligned} \frac{m}{2} \sum_{j\in M_t} \frac{\eta_{q,j}}{r_j} &\overset{(11)}{=} \frac{\sqrt{m\ln m}}{2} \sum_{j\in M_t} \frac{\frac{1}{r_j}}{\sqrt{\sum_{s=1}^{j} \frac{1}{r_s}}} \\ &\leq \frac{\sqrt{m\ln m}}{2} \sum_{j\in M_t} \frac{\frac{1}{r_j}}{\sqrt{\sum_{s\in[j]\cap M_t} \frac{1}{r_s}}} \overset{(24)}{\leq} \sqrt{\sum_{j\in M_t} \frac{m}{r_j} \ln m}. \end{aligned} \tag{33}$$

Combining Lemma G.10 with (33), we obtain, with probability at least $1-\delta$,

$$\sum_{j\in M_t} Q_j \leq 2^{1/4} m \sqrt{\ln m \sqrt{\sum_{j=1}^{t} \frac{m}{r_j} \ln \frac{m}{\delta}} + \frac{m^2\sqrt{\ln m}}{3} \ln \frac{m}{\delta}} + \sqrt{\sum_{j\in M_t} \frac{m}{r_j} \ln m}. \tag{34}$$

For $\texttt{term(f)}$, we define $V_j = \langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \hat{s}_{j,c_j}$ when $j \in S_t$ and $V_j = \langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle$ when $j \in M_t$. Then we know $\mathbb{E}_{j-1}[V_j] = 0$. By Lemma G.2, (50), (8), $m \geq 2$ and $r_j \geq 2$ when $j \in M_t$, we know $|V_j| \leq \frac{m}{2}$ for all $j \in [t]$. Moreover, for $j \in S_t$, we have

$$\mathbb{E}_{j-1}\left[V_j^2\right] = \mathbb{E}_{j-1}\left[\left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \hat{s}_{j,c_j}\right)^2\right] \leq 1,$$

and for $j \in M_t$, we have

$$\begin{aligned} \mathbb{E}_{j-1}\left[V_j^2\right] &= \mathbb{E}_{j-1}\left[\left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle - \langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle\right)^2\right] = \mathbb{E}_{j-1}\left[\left(\langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle\right)^2\right] - \mathbb{E}_{j-1}\left[\left(\langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle\right)^2\right] \\ &\leq \mathbb{E}_{j-1}\left[\left(\langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle\right)^2\right] \overset{(23)}{\leq} \frac{m}{r_j} \mathbb{E}_{j-1}\left[\langle \mathbf{q}_j, \tilde{\mathbf{s}}_j \rangle\right] \leq \frac{m}{r_j}. \end{aligned}$$

In both regimes, $\mathbb{E}_{j-1}[V_j^2] \leq \frac{m}{r_j}$. To apply Lemma G.5, we instantiate its martingale difference sequence as $X_j = V_j$. Let $A_j = \sum_{s=1}^{j} \frac{m}{r_s}$. For any $B > 0$, since $\mathbb{I}[A_j \leq B]$ is $\mathcal{F}_{j-1}$-measurable, we have

$$\sum_{j=1}^{t} \mathbb{E}_{j-1}\left[\left(V_j \mathbb{I}[A_j \leq B]\right)^2\right] = \sum_{j=1}^{t} \mathbb{I}[A_j \leq B] \mathbb{E}_{j-1}[V_j^2] \leq \sum_{j=1}^{t} \frac{m}{r_j} \mathbb{I}[A_j \leq B] \leq B.$$

Therefore, the conditions of (25) hold with $\Psi(B) = B$ and $b = m/2$. It follows that, with probability at least $1-\delta$,

$$\sum_{j=1}^{t} V_j \leq 2\sqrt{\sum_{j=1}^{t} \frac{m}{r_j} \ln \frac{m}{\delta}} + \frac{m}{3} \ln \frac{m}{\delta}. \tag{35}$$

For $\texttt{term(g)}$, we need to control the unweighted sum $\sum_{j\in S_t}(\tilde{s}_{j,i} - \hat{s}_{j,i})$ for each $i \in [m]$. Lemma G.9, however, controls weighted sums with weights bounded by $2\gamma_j$. To obtain the desired unweighted bound, we apply this lemma with a family of scale-dependent common weights. Let $A_j = \sum_{s=1}^{j} \frac{m}{r_s}$. Fix t, let $K = 1 + \lfloor \log_2 m \rfloor \leq m$, and set $B_k = 2^{k+1}t$ and $\lambda_k = \sqrt{\ln(m)/B_k}$ for $k = 0, \ldots, K-1$. For each $i \in [m]$, define $\alpha_{j,\ell}^{(k,i)} = \lambda_k \mathbb{I}[A_j \leq B_k]\mathbb{I}[\ell = i]$. Since A_j is $\mathcal{F}_{j-1}$-measurable and $\lambda_k \leq \sqrt{\ln(m)/A_j} = 2\gamma_j$ whenever $A_j \leq B_k$, these weights satisfy the conditions

of Lemma G.9. Applying the lemma with $S = S_t$ and failure probability $\delta/(mK)$, followed by a union bound over i and k, gives, with probability at least $1-\delta$,

$$\lambda_k \sum_{j \in S_t} \mathbb{I}[A_j \le B_k](\tilde{s}_{j,i} - \hat{s}_{j,i}) \le \ln \frac{mK}{\delta} \le \ln \frac{m^2}{\delta},$$

simultaneously for all $i \in [m]$ and $k = 0, \dots, K-1$. Since $t \le A_t \le mt$, there exists $k \in \{0, \dots, K-1\}$ such that $2^k t \le A_t < B_k \le 2A_t$. For this k, $A_j \le A_t < B_k$ for every $j \le t$, so all indicators above equal one. Moreover, $\lambda_k^{-1} = \sqrt{B_k / \ln m} \le \sqrt{2A_t / \ln m}$. Therefore, simultaneously for all $i \in [m]$,

$$\sum_{j \in S_t} (\tilde{s}_{j,i} - \hat{s}_{j,i}) \le \sqrt{\frac{2A_t}{\ln m}} \ln \frac{m^2}{\delta} = 2\sqrt{2 \sum_{j=1}^{t} \frac{m}{r_j} \ln m} + \sqrt{\frac{2 \sum_{j=1}^{t} \frac{m}{r_j}}{\ln m}} \ln \frac{1}{\delta}. \tag{36}$$

For $\texttt{term(h)}$, define $X_{j,i} = (\tilde{s}_{j,i} - \hat{s}_{j,i})\mathbb{I}[r_j \ge 2]$ for all $j \in [t]$ and $i \in [m]$. Then $\mathbb{E}_{j-1}[X_{j,i}] = 0$ and $|X_{j,i}| \le m$. For $j \in M_t$,

$$\mathbb{E}_{j-1}[X_{j,i}^2] \le \mathbb{E}_{j-1}[\tilde{s}_{j,i}^2] \le \frac{1}{q_{j,i} + (1 - q_{j,i})\frac{r_j - 1}{m-1}} \le \frac{m}{r_j - 1} \le \frac{2m}{r_j}.$$

Hence, for any $B > 0$,

$$\sum_{j=1}^{t} \mathbb{E}_{j-1}\left[(X_{j,i}\mathbb{I}[A_j \le B])^2\right] \le 2 \sum_{j=1}^{t} \frac{m}{r_j} \mathbb{I}[A_j \le B] \le 2B.$$

Applying (25) with $\Psi(B) = 2B$, $b = m$, and failure probability δ/m, and then taking a union bound over $i \in [m]$, we obtain, with probability at least $1-\delta$, simultaneously for all $i \in [m]$,

$$\sum_{j \in M_t} (\tilde{s}_{j,i} - \hat{s}_{j,i}) \le 2\sqrt{2 \sum_{j=1}^{t} \frac{m}{r_j} \ln \frac{m^2}{\delta}} + \frac{2m}{3} \ln \frac{m^2}{\delta}. \tag{37}$$

Combining (31)–(37) with (29) and allocating failure probability $\delta/5$ to each term, and further using (11), $\ln(5m/\delta) = \ln m + \ln(5/\delta)$, $\ln(5m^2/\delta) = 2\ln m + \ln(5/\delta)$, and $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, we obtain, with probability at least $1-\delta$,

$$\begin{aligned}
\mathrm{Regret}'_{q,t} \le{}& \sqrt{\sum_{j=1}^{t} \frac{m}{r_j}} \left(9\sqrt{2\ln m} + 4\sqrt{2\ln \frac{5}{\delta}}\right) + 2^{1/4} m \sqrt{\ln m} \sqrt{\sum_{j=1}^{t} \frac{m}{r_j} \ln \frac{5m}{\delta}} \\
&+ \left(m + 1 + \frac{m^2\sqrt{\ln m}}{3} + \sqrt{\frac{2\sum_{j=1}^{t} \frac{m}{r_j}}{\ln m}}\right) \ln \frac{5}{\delta} + \left(\frac{m^2\sqrt{\ln m}}{3} + \frac{5m}{3}\right) \ln m.
\end{aligned}$$

Here, we also use $|S_t| m \le \sum_{j=1}^{t} \frac{m}{r_j}$ and

$$\sqrt{\sum_{j \in M_t} \frac{m}{r_j} \ln m} + \sqrt{|S_t| m \ln m} \le \sqrt{2 \sum_{j=1}^{t} \frac{m}{r_j} \ln m}, \tag{38}$$

where the latter follows from $S_t \cup M_t = [t]$ and $r_j = 1$ for $j \in S_t$.

G.3 Proof of Lemma 3.3

By convexity of $\phi(\cdot, q_j)$, we have

$$\begin{aligned}
&\sum_{j=1}^{t} [\phi(\mathbf{w}_j, \mathbf{q}_j) - \phi(\mathbf{w}, \mathbf{q}_j)] \le \sum_{j=1}^{t} \langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{w} \rangle \\
\le{}& \sum_{j=1}^{t} \langle \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{w} \rangle + \sum_{j=1}^{t} \langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{w} \rangle.
\end{aligned} \tag{39}$$

We bound the first term as follows. By defining $\nu_{w,t}(\mathbf{w}) = \frac{1}{\eta_{w,t}}\left(\nu_w(\mathbf{w}) - \min_{\mathbf{w}\in\mathcal{W}} \nu_w(\mathbf{w})\right)$, we rewrite (16) as

$$\mathbf{w}_t = \operatorname*{argmin}_{\mathbf{w}\in\mathcal{W}} \left\{ \left\langle \sum_{i=1}^{t-1} \mathbf{g}_w(\mathbf{w}_i, \mathbf{q}_i), \mathbf{w} \right\rangle + \nu_{w,t}(\mathbf{w}) \right\}.$$

By the standard analysis of FTRL [Orabona, 2019, Corollary 7.9.] and the fact that ν_w is 1-strongly convex with respect to a certain norm $\|\cdot\|_w$, for all $t \in \mathbb{Z}^+$ and $\mathbf{w} \in \mathcal{W}$, we have

$$\begin{aligned}\sum_{j=1}^{t} \langle \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{w}\rangle \leq & \frac{\nu_w(\mathbf{w}) - \min_{\mathbf{w}\in\mathcal{W}} \nu_w(\mathbf{w})}{\eta_{w,t}} + \frac{1}{2}\sum_{j=1}^{t} \eta_{w,j} \|\mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*}^2 \\ \overset{(3),(18)}{\leq} & \frac{D^2}{\eta_{w,t}} + \frac{G^2}{2}\sum_{j=1}^{t} \eta_{w,j}.\end{aligned} \tag{40}$$

Taking the maximum over the two sides in (39) over $\mathbf{w} \in \mathcal{W}$, we obtain

$$\begin{aligned}&\max_{\mathbf{w}\in\mathcal{W}} \left\{ \sum_{j=1}^{t} [\phi(\mathbf{w}_j, \mathbf{q}_j) - \phi(\mathbf{w}, \mathbf{q}_j)] \right\} = \sum_{j=1}^{t} \phi(\mathbf{w}_j, \mathbf{q}_j) - \min_{\mathbf{w}\in\mathcal{W}} \sum_{j=1}^{t} \phi(\mathbf{w}, \mathbf{q}_j) \\ \overset{(40)}{\leq} & \frac{D^2}{\eta_{w,t}} + \frac{G^2}{2}\sum_{j=1}^{t} \eta_{w,j} + \max_{\mathbf{w}\in\mathcal{W}} \left\{ \underbrace{\sum_{j=1}^{t} \langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{w}\rangle}_{=F_t(\mathbf{w})} \right\}.\end{aligned} \tag{41}$$

To bound $\max_{\mathbf{w}\in\mathcal{W}} F_t(\mathbf{w})$ in (41), we cannot directly treat the summands of $F_t(\widetilde{\mathbf{w}})$ as martingale differences, because $\widetilde{\mathbf{w}} = \operatorname{argmax}_{\mathbf{w}\in\mathcal{W}} F_t(\mathbf{w})$ may depend on the gradient noise from rounds $1, \ldots, t$ [Zhang et al., 2023]. To address this challenge, we employ the "ghost iterate" technique [Nemirovski et al., 2009, proof of Lemma 3.1], which decouples the dependency between $\widetilde{\mathbf{w}}$ and $F_t(\mathbf{w})$. It is important to note that the referenced work applies the "ghost iterate" to SMD, whereas our analysis constructs an FTRL-based variant. Specifically, we introduce a virtual sequence of variables that performs FTRL as

$$\mathbf{v}_t = \operatorname*{argmin}_{\mathbf{w}\in\mathcal{W}} \left\{ \left\langle \sum_{j=1}^{t-1} \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w} \right\rangle + \nu_{w,t}(\mathbf{w}) \right\}, \tag{42}$$

where $\mathbf{v}_1 = \mathbf{w}_1$. By repeating the derivation of (40), we obtain for all $\mathbf{w} \in \mathcal{W}$,

$$\begin{aligned}&\sum_{j=1}^{t} \langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{v}_j - \mathbf{w}\rangle \\ \leq & \frac{\nu_w(\mathbf{w}) - \min_{\mathbf{w}\in\mathcal{W}} \nu_w(\mathbf{w})}{\eta_{w,t}} + \frac{1}{2}\sum_{j=1}^{t} \eta_{w,j} \|\nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*}^2 \\ \leq & \frac{D^2}{\eta_{w,t}} + 2G^2 \sum_{j=1}^{t} \eta_{w,j},\end{aligned} \tag{43}$$

where the last inequality is due to (3) and

$$\begin{aligned}&\|\nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*} \leq \|\nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*} + \|\mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*} \\ \leq & \mathbb{E}_{j-1}[\|\mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*}] + \|\mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*} \overset{(18)}{\leq} 2G.\end{aligned} \tag{44}$$

Then, we have

$$\begin{aligned}\max_{\mathbf{w}\in\mathcal{W}} F_t(\mathbf{w}) = \max_{\mathbf{w}\in\mathcal{W}} &\left\{ \sum_{j=1}^{t} \langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{v}_j - \mathbf{w} \rangle \right\} \\ &+ \sum_{j=1}^{t} \langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{v}_j \rangle \\ &\overset{(43)}{\leq} \frac{D^2}{\eta_{w,t}} + 2G^2 \sum_{j=1}^{t} \eta_{w,j} + \sum_{j=1}^{t} V_j,\end{aligned} \tag{45}$$

where $V_j = \langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{v}_j \rangle$. Since $\mathbf{v}_j$ and $\mathbf{w}_j$ are $\mathcal{F}_{j-1}$-measurable by (42) and (16), and $\mathbb{E}_{j-1}[\mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)] = \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j)$ by (18), we have $\mathbb{E}_{j-1}[V_j] = 0$ for every $j \in [t]$. Hence $\{V_j\}_{j=1}^t$ is a martingale difference sequence.

To establish a high-probability bound for any fixed $t \in \mathbb{N}^+$, we first note that

$$\begin{aligned}|V_j| &= |\langle \nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j), \mathbf{w}_j - \mathbf{v}_j \rangle| \\ &\leq \|\nabla_{\mathbf{w}} \phi(\mathbf{w}_j, \mathbf{q}_j) - \mathbf{g}_w(\mathbf{w}_j, \mathbf{q}_j)\|_{w,*} \|\mathbf{w}_j - \mathbf{v}_j\|_w \\ &\overset{(44)}{\leq} 2G \|\mathbf{w}_j - \mathbf{v}_j\|_w \leq 2G \left(\|\mathbf{w}_j - \mathbf{w}_1\|_w + \|\mathbf{v}_j - \mathbf{w}_1\|_w \right) \\ &\leq 2G \left(\sqrt{2B_w(\mathbf{w}_j, \mathbf{w}_1)} + \sqrt{2B_w(\mathbf{v}_j, \mathbf{w}_1)} \right) \leq 4\sqrt{2} DG,\end{aligned} \tag{46}$$

where the last step is by defining

$$B_w(\mathbf{u}, \mathbf{v}) = \nu_w(\mathbf{u}) - \left[\nu_w(\mathbf{v}) + \langle \nabla \nu_w(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle \right], \ \forall \mathbf{u}, \mathbf{v} \in \mathcal{W},$$

and the inequality [Nemirovski et al., 2009, (2.42)]

$$\max_{\mathbf{w}\in\mathcal{W}} B_w(\mathbf{w}, \mathbf{w}_1) \leq \max_{\mathbf{w}\in\mathcal{W}} \nu_w(\mathbf{w}) - \min_{\mathbf{w}\in\mathcal{W}} \nu_w(\mathbf{w}) \overset{(3)}{\leq} D^2.$$

By applying Lemma G.4, we obtain for any fixed $t \in \mathbb{N}^+$, with probability at least $1 - \delta$,

$$\sum_{j=1}^{t} V_j \leq 8DG \sqrt{t \ln \frac{1}{\delta}}. \tag{47}$$

Combining (41), (45) and (47), it follows that, with probability at least $1 - \delta$,

$$\begin{aligned}\sum_{j=1}^{t} \phi(\mathbf{w}_j, \mathbf{q}_j) - \min_{\mathbf{w}\in\mathcal{W}} \sum_{j=1}^{t} \phi(\mathbf{w}, \mathbf{q}_j) &\leq \frac{2D^2}{\eta_{w,t}} + \frac{5G^2}{2} \sum_{j=1}^{t} \eta_{w,j} + 8DG \sqrt{t \ln \frac{1}{\delta}} \\ &\leq DG\sqrt{t} \left(2\sqrt{10} + 8\sqrt{\ln \frac{1}{\delta}} \right),\end{aligned}$$

where in the last inequality, we set $\eta_{w,j} = \frac{\sqrt{2}D}{\sqrt{5}G} \frac{1}{\sqrt{j}}$ and use the fact that

$$\sum_{j=1}^{t} \frac{1}{\sqrt{j}} \leq \int_0^t \frac{1}{\sqrt{x}} \mathrm{d}x = 2\sqrt{x}\Big|_0^t = 2\sqrt{t}.$$

G.4 Proof of Theorem 3.4

The complete form of Theorem 3.4 is given below.

Theorem G.11. *Let* FLEX-GDRO *be Algorithm 3 instantiated with Algorithm 1 as $\mathcal{A}_q$ and Algorithm 2 as $\mathcal{A}_w$, with step sizes specified by Lemmas 3.2 and 3.3. Under Assumptions 2.1–2.4, for*

any fixed $t \in \mathbb{N}^+$ and any sample-size process $\{r_j\}_{j\in[t]}$ generated according to Definition 3.1, we have, with probability at least $1-\delta$,

$$
\begin{aligned}
&\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) \\
\leq &\frac{DG}{\sqrt{t}}\left(2\sqrt{10} + 8\sqrt{\ln\frac{4}{\delta}}\right) + \frac{1}{t}\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}}\left(9\sqrt{2\ln m} + 4\sqrt{2\ln\frac{20}{\delta}}\right) \\
&+ \frac{2^{1/4}m}{t}\sqrt{\ln m\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}}\ln\frac{20m}{\delta}} + \frac{1}{t}\left(m + 1 + \frac{m^2\sqrt{\ln m}}{3} + \sqrt{\frac{2\sum_{j=1}^{t}\frac{m}{r_j}}{\ln m}}\right)\ln\frac{20}{\delta} \\
&+ \frac{1}{t}\left(\frac{m^2\sqrt{\ln m}}{3} + \frac{5m}{3}\right)\ln m + 2\sqrt{\frac{2}{t}\ln\frac{4m}{\delta}}.
\end{aligned}
$$

By substituting the results of Lemmas 3.3 and 3.2 into Lemma G.1, we complete the proof.

G.5 Proof of Theorem 3.5

The full statement of Theorem 3.5 is given below.

Theorem G.12. *Let* FLEX-GDRO *be Algorithm 3 instantiated with Algorithm 1 as $\mathcal{A}_q$ and Algorithm 2 as $\mathcal{A}_w$, with step sizes specified by Lemmas 3.2 and 3.3. Let $\{r_t\}_{t\geq 1}$ be any process generated according to Definition 3.1. Under Assumptions 2.1–2.4, with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,*

$$
\begin{aligned}
\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) \leq &\frac{DG}{\sqrt{t}}\left(2\sqrt{10} + 10\sqrt{\ln\frac{16}{\delta} + 2\ln\ln(2t)}\right) + 4\sqrt{\frac{1}{t}\left(\ln\frac{16m}{\delta} + 2\ln\ln(2t)\right)} \\
&+ \frac{1}{t}\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}}\left(8\sqrt{\ln m} + 2\sqrt{\ln\frac{20C_{2,t}}{\delta}}\right) + \frac{m}{t}\sqrt{2\ln m\sqrt{\sum_{j\in M_t}\frac{m}{r_j}}\ln\left(\frac{20C_{1,t}}{\delta}\right)} \\
&+ \frac{2}{t}\sqrt{2\sum_{j\in M_t}\frac{m}{r_j}\ln\frac{20C_{3,t}}{\delta}} + \frac{m}{t}\ln\frac{20C_{2,t}}{\delta} + \frac{1}{t}\left(1 + \sqrt{\frac{2\sum_{j=1}^{t}\frac{m}{r_j}}{\ln m}}\right)\ln\frac{80(\ln 2t)^2}{\delta} \\
&+ \frac{m^2\sqrt{\ln m}}{t}\ln\left(\frac{20C_{1,t}}{\delta}\right) + \frac{2m\ln m}{t} + \frac{2m}{t}\ln\frac{20C_{3,t}}{\delta},
\end{aligned}
$$

where $C_{1,t} = 2\left(1 + \ln^+\left(\frac{2}{m}\sqrt{\sum_{j\in M_t}\frac{1}{r_j}}\right)\right)^2$, $C_{2,t} = 2\left(1 + \ln^+\left(\frac{4}{m^2}\left(\sum_{j\in M_t}\frac{m}{r_j} + |S_t|\right)\right)\right)^2$, $C_{3,t} = 2\left(1 + \ln^+\left(\sum_{j\in M_t}\frac{2}{mr_j}\right)\right)^2$ with $\ln^+ x = \max\{\ln x, 0\}$ for $x > 0$ and $\ln^+ 0 = 0$, $S_t = \{j \in [t] \mid r_j = 1\}$ and $M_t = \{j \in [t] \mid r_j \geq 2\}$.

The analysis proceeds by decomposing the time-uniform optimization error and then bounding regrets for both the w-player and the q-player.

Lemma G.13. *Under Assumptions 2.1-2.4, suppose that*

1. *The w-player is equipped with a non-oblivious OCO algorithm $\mathcal{A}_w$, and with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$, the regret $\mathrm{Regret}_{w,t}$ is upper bounded by $U_w(t,\delta)$.*
2. *The q-player is equipped with a non-oblivious PLA algorithm $\mathcal{A}_q$, and with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$, the regret $\mathrm{Regret}'_{q,t}$ is upper bounded by $U_q(t,\delta)$.*

Then, with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$
\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) \leq \frac{1}{t}U_w\left(t, \frac{\delta}{4}\right) + \frac{1}{t}U_q\left(t, \frac{\delta}{4}\right) + 4\sqrt{\frac{1}{t}\left(\ln\frac{16m}{\delta} + 2\ln\ln(2t)\right)}.
$$

Lemma G.14. *Let $\eta_{q,t}$ and γ_t be defined in* (11) *and* (14) *for Algorithm 1. Let $\{r_t\}_{t\geq 1}$ be any process generated according to Definition 3.1. Under Assumption 2.3, we obtain with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,*

$$\begin{aligned}\mathrm{Regret}'_{q,t} \leq & \sqrt{\sum_{j=1}^{t} \frac{m}{r_j}} \left(8\sqrt{\ln m} + 2\sqrt{\ln \frac{5C_{2,t}}{\delta}}\right) + m\sqrt{2\ln m \sqrt{\sum_{j\in M_t} \frac{m}{r_j} \ln\left(\frac{5C_{1,t}}{\delta}\right)}} \\ & + 2\sqrt{2\sum_{j\in M_t} \frac{m}{r_j} \ln \frac{5C_{3,t}}{\delta}} + m\ln\frac{5C_{2,t}}{\delta} + \left(1 + \sqrt{\frac{2\sum_{j=1}^{t}\frac{m}{r_j}}{\ln m}}\right) \ln \frac{20(\ln 2t)^2}{\delta} \\ & + m^2\sqrt{\ln m}\ln\left(\frac{5C_{1,t}}{\delta}\right) + 2m\ln m + 2m\ln\frac{5C_{3,t}}{\delta}.\end{aligned}$$

where $C_{1,t} = 2\left(1 + \ln^+\left(\frac{2}{m}\sqrt{\sum_{j\in M_t}\frac{1}{r_j}}\right)\right)^2$, $C_{2,t} = 2\left(1 + \ln^+\left(\frac{4}{m^2}\left(\sum_{j\in M_t}\frac{m}{r_j} + |S_t|\right)\right)\right)^2$, $C_{3,t} = 2\left(1 + \ln^+\left(\sum_{j\in M_t}\frac{2}{mr_j}\right)\right)^2$ *with* $\ln^+ x = \max\{\ln x, 0\}$ *for* $x > 0$ *and* $\ln^+ 0 = 0$, $S_t = \{j \in [t] \mid r_j = 1\}$ *and* $M_t = \{j \in [t] \mid r_j \geq 2\}$.

Lemma G.15. *Under Assumptions 2.1, 2.2 and 2.4, and setting $\eta_{w,t} = \frac{\sqrt{2}D}{\sqrt{5}G}\frac{1}{\sqrt{t}}$ for Algorithm 2, we have with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,*

$$\mathrm{Regret}_{w,t} \leq DG\sqrt{t}\left(2\sqrt{10} + 10\sqrt{\ln\frac{4}{\delta} + 2\ln\ln(2t)}\right).$$

By substituting the results of Lemmas G.14 and G.15 into Lemma G.13, we complete the proof.

G.6 Proof of Lemma G.1

By Jensen's inequality and the outputs $\bar{\mathbf{w}}_t = \frac{1}{t}\sum_{j=1}^{t}\mathbf{w}_j$ and $\bar{\mathbf{q}}_t = \frac{1}{t}\sum_{j=1}^{t}\mathbf{q}_j$, we have

$$\begin{aligned}&\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) = \max_{\mathbf{q}\in\Delta_m}\phi(\bar{\mathbf{w}}_t, \mathbf{q}) - \min_{\mathbf{w}\in\mathcal{W}}\phi(\mathbf{w}, \bar{\mathbf{q}}_t) \\ \leq & \frac{1}{t}\left(\max_{\mathbf{q}\in\Delta_m}\sum_{j=1}^{t}\phi(\mathbf{w}_j, \mathbf{q}) - \min_{\mathbf{w}\in\mathcal{W}}\sum_{j=1}^{t}\phi(\mathbf{w}, \mathbf{q}_j)\right) = \frac{1}{t}\mathrm{Regret}_{q,t} + \frac{1}{t}\mathrm{Regret}_{w,t},\end{aligned} \tag{48}$$

where $\mathrm{Regret}_{q,t}$ and $\mathrm{Regret}_{w,t}$ are defined in (7). For the first term $\mathrm{Regret}_{q,t}$, by the definition of $\phi(\mathbf{w}, \mathbf{q})$ in (2) and the property of linear optimization over the simplex Δ_m, we have

$$\begin{aligned}&\mathrm{Regret}_{q,t} \\ =& \max_{\mathbf{q}\in\Delta_m}\sum_{j=1}^{t}\phi(\mathbf{w}_j, \mathbf{q}) - \sum_{j=1}^{t}\phi(\mathbf{w}_j, \mathbf{q}_j) = \max_{\mathbf{q}\in\Delta_m}\sum_{i=1}^{m} q_i\left(\sum_{j=1}^{t}R_i(\mathbf{w}_j)\right) - \sum_{j=1}^{t}\sum_{i=1}^{m} q_{j,i}R_i(\mathbf{w}_j) \\ =& \sum_{j=1}^{t}R_{k^*}(\mathbf{w}_j) - \sum_{j=1}^{t}\sum_{i=1}^{m} q_{j,i}R_i(\mathbf{w}_j) = \sum_{j=1}^{t}\sum_{i=1}^{m} q_{j,i}s_{j,i} - \sum_{j=1}^{t}s_{j,k^*} \\ =& \sum_{j=1}^{t}\langle\mathbf{q}_j, \mathbf{s}_j\rangle - \sum_{j=1}^{t}\langle\mathbf{q}_j, \hat{\mathbf{s}}_j\rangle + \sum_{j=1}^{t}\langle\mathbf{q}_j, \hat{\mathbf{s}}_j\rangle - \sum_{j=1}^{t}\hat{s}_{j,k^*} + \sum_{j=1}^{t}\hat{s}_{j,k^*} - \sum_{j=1}^{t}s_{j,k^*} \\ \leq & \underbrace{\sum_{j=1}^{t}\langle\mathbf{q}_j, \mathbf{s}_j\rangle - \sum_{j=1}^{t}\langle\mathbf{q}_j, \hat{\mathbf{s}}_j\rangle}_{\mathtt{term(e_1)}} + \underbrace{\sum_{j=1}^{t}\langle\mathbf{q}_j, \hat{\mathbf{s}}_j\rangle - \min_{i\in[m]}\sum_{j=1}^{t}\hat{s}_{j,i}}_{\mathrm{Regret}'_{q,t}} + \underbrace{\sum_{j=1}^{t}\hat{s}_{j,k^*} - \sum_{j=1}^{t}s_{j,k^*}}_{\mathtt{term(e_2)}},\end{aligned} \tag{49}$$

where $k^* = \mathrm{argmax}_{k\in[m]}\sum_{j=1}^{t}R_k(\mathbf{w}_j)$, $\hat{\mathbf{s}}_j \in \mathbb{R}^m$ is defined in (8) and the vector $\mathbf{s}_j \in \mathbb{R}^m$ is defined as

$$s_{j,i} = \mathbb{E}_{j-1}[\hat{s}_{j,i}] = 1 - R_i(\mathbf{w}_j) = 1 - \mathbb{E}_{\mathbf{z}\sim\mathcal{P}_i}[\ell(\mathbf{w}_j; \mathbf{z})] \overset{(5)}{\in} [0, 1],\ \forall i \in [m]. \tag{50}$$

For $\texttt{term}(\texttt{e}_1)$, we utilize the Hoeffding-Azuma inequality [Cesa-Bianchi and Lugosi, 2006] in Lemma G.4 to establish a high-probability bound. Denoting $V_j = \langle \mathbf{q}_j, \mathbf{s}_j \rangle - \langle \mathbf{q}_j, \hat{\mathbf{s}}_j \rangle$, by (50) and (8), we know $\mathbb{E}_{j-1}[V_j] = 0$ and $|V_j| \leq 1$ for all $j \in [t]$. By Lemma G.4, we have for any fixed $t \in \mathbb{N}^+$, with probability at least $1 - \delta$,

$$\texttt{term}(\texttt{e}_1) = \sum_{j=1}^{t} V_j \leq \sqrt{2t \ln \frac{1}{\delta}}. \tag{51}$$

Similarly, for $\texttt{term}(\texttt{e}_2)$, we can verify that for all $i \in [m]$, $V'_{j,i} = \hat{s}_{j,i} - s_{j,i}$ is a martingale difference sequence bounded by 1, i.e., $\mathbb{E}_{j-1}[V'_{j,i}] = 0$ and $|V'_{j,i}| \leq 1$ for all $j \in [t]$. Thus, by Lemma G.4 and taking the union bound over $i \in [m]$, we have for any fixed $t \in \mathbb{N}^+$, with probability at least $1 - \delta$,

$$\texttt{term}(\texttt{e}_2) = \sum_{j=1}^{t} V'_{j,k^*} \leq \sqrt{2t \ln \frac{m}{\delta}}. \tag{52}$$

By substituting (51) and (52) into (49), and subsequently combining the result with (48), we obtain the high-probability bound stated in (22).

G.7 Proof of Lemma G.5

We first recall the classical Bernstein inequality [Cesa-Bianchi and Lugosi, 2006, Lemma A.8]. Let $\{X_j\}_{j \in [t]}$ be a bounded martingale difference sequence satisfying $\mathbb{E}_{j-1}[X_j] = 0$, $|X_j| \leq b$, and $\sum_{j=1}^{t} \mathbb{E}_{j-1}[X_j^2] \leq V$, where b and V are positive values. Then, with probability at least $1 - \delta$,

$$\sum_{j=1}^{t} X_j \leq \sqrt{2V \ln \frac{1}{\delta}} + \frac{2b}{3} \ln \frac{1}{\delta}.$$

We now prove the adaptive-budget version. Fix any $t \in \mathbb{N}_+$ and denote $A_j = \sum_{s=1}^{j} \frac{m}{r_s}$ for $j \in [t]$. In particular, $A_t = \sum_{j=1}^{t} \frac{m}{r_j}$ is the cumulative budget quantity appearing in our variance bound. Since the sample sizes r_j may depend on the history, A_t may also vary with the realized history, so we cannot directly use $\Psi(A_t)$ as the variance parameter in the classical Bernstein inequality.

We handle this issue by dividing the possible range of A_t into multiple scales. For each scale B_k, we stop the martingale when the cumulative budget A_j exceeds B_k. This gives a martingale whose cumulative conditional variance is bounded by $\Psi(B_k)$, so the classical Bernstein inequality can be applied at this scale. We then take a union bound over all scales, so that all these bounds hold simultaneously. Finally, after the realized value of A_t is known, we choose a scale B_k satisfying $A_t < B_k \leq 2A_t$, which gives the desired bound in terms of the realized budget A_t.

Let $L = 1 + \lfloor \log_2 m \rfloor \leq m$ and set $B_k = 2^{k+1}t$ for $k = 0, \ldots, L-1$. For each B_k, define $X_j^{(B_k)} = X_j \mathbb{I}[A_j \leq B_k]$. Since A_j is $\mathcal{F}_{j-1}$-measurable under Definition 3.1, we have $\mathbb{E}_{j-1}[X_j^{(B_k)}] = \mathbb{I}[A_j \leq B_k]\mathbb{E}_{j-1}[X_j] = 0$ and $|X_j^{(B_k)}| \leq b$. Moreover, by the assumption on Ψ,

$$\sum_{j=1}^{t} \mathbb{E}_{j-1}\left[(X_j^{(B_k)})^2\right] = \sum_{j=1}^{t} \mathbb{I}[A_j \leq B_k]\mathbb{E}_{j-1}[X_j^2] \leq \Psi(B_k).$$

If $\Psi(B_k) = 0$, then $X_j^{(B_k)} = 0$ for every $j \in [t]$. Otherwise, applying the classical Bernstein inequality to $\{X_j^{(B_k)}\}_{j \in [t]}$ with failure probability δ/m, we obtain

$$\Pr\left[\sum_{j=1}^{t} X_j^{(B_k)} > \sqrt{2\Psi(B_k) \ln \frac{m}{\delta}} + \frac{2b}{3} \ln \frac{m}{\delta}\right] \leq \frac{\delta}{m}.$$

Taking a union bound over $k = 0, \ldots, L-1$ and using $L \leq m$, with probability at least $1 - \delta$, the following inequality holds simultaneously for all $k = 0, \ldots, L-1$:

$$\sum_{j=1}^{t} X_j^{(B_k)} \leq \sqrt{2\Psi(B_k) \ln \frac{m}{\delta}} + \frac{2b}{3} \ln \frac{m}{\delta}.$$

It remains to select the scale corresponding to the realized value of A_t. Since $t \le A_t \le mt$, there exists $k \in \{0, \dots, L-1\}$ such that $2^k t \le A_t < 2^{k+1} t = B_k \le 2A_t$. For this k, we have $A_j \le A_t < B_k$ for every $j \in [t]$, and hence $X_j^{(B_k)} = X_j$ for every $j \in [t]$. Therefore, on the common high-probability event above,

$$\sum_{j=1}^{t} X_j \le \sqrt{2\Psi(B_k) \ln \frac{m}{\delta}} + \frac{2b}{3} \ln \frac{m}{\delta}.$$

Since Ψ is nondecreasing and $B_k \le 2A_t$, we further have $\Psi(B_k) \le \Psi(2A_t)$. Recalling $A_t = \sum_{j=1}^{t} \frac{m}{r_j}$, we conclude that, with probability at least $1-\delta$,

$$\sum_{j=1}^{t} X_j \le \sqrt{2\Psi\left(2\sum_{j=1}^{t} \frac{m}{r_j}\right) \ln \frac{m}{\delta}} + \frac{2b}{3} \ln \frac{m}{\delta},$$

which proves (25).

G.8 Proof of Lemma G.7

We begin by presenting a concentration inequality [Freedman, 1975, Lee and Oh, 2025]. The proof in this section follows an online blog [9].

Lemma G.16. *[Lee and Oh, 2025, Proposition 4.] Let $\{X_t\}_{t=1}^{\infty}$ be a martingale difference sequence with respect to a filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. Suppose $X_t \le 1$ holds almost surely for all $t \ge 1$. Let $V_t := \mathbb{E}\left[X_t^2 \mid \mathcal{F}_{t-1}\right]$ for $t \ge 1$ and take $\lambda > 0$ and $\delta \in (0,1]$ arbitrarily. Then, the following inequality holds with probability at least $1-\delta$, for all $n \in \mathbb{N}$:*

$$\sum_{t=1}^{n} X_t \le \frac{e^\lambda - 1 - \lambda}{\lambda} \sum_{t=1}^{n} V_t + \frac{1}{\lambda} \log \frac{1}{\delta}. \tag{53}$$

Suppose $0 < \lambda < 3$. Then, by Taylor expansion, it holds that

$$g(\lambda) = \sum_{i=0}^{\infty} \frac{\lambda^i}{(i+2)!} \le \sum_{i=0}^{\infty} \frac{\lambda^i}{2 \cdot 3^i} \le \frac{1}{2\left(1 - \frac{\lambda}{3}\right)}$$

where the first inequality holds by $n! = 2 \cdot 3 \dots n \ge 2 \cdot 3 \dots 3 = 2 \cdot 3^{n-2}$ for $n \ge 2$ and the last inequality holds since $0 < \lambda < 3$. Then, it holds that $\frac{e^\lambda - 1 - \lambda}{\lambda} \le \frac{\lambda}{2\left(1-\frac{\lambda}{3}\right)}$. Plugging in this bound to (53), we obtain that for fixed $0 < \lambda < 3$ and $T_n = \sum_{t=1}^{n} \mathbb{E}[X_t^2 | \mathcal{F}_{t-1}]$, it holds that

$$\mathbb{P}\left(\exists n \in \mathbb{N} : \sum_{t=1}^{n} X_t \ge \frac{\lambda T_n}{2\left(1 - \frac{\lambda}{3}\right)} + \frac{1}{\lambda} \log \frac{1}{\delta}\right) \le \delta \tag{54}$$

Taking the union bound over the following values of $\lambda = \lambda_k$ for $k = 0, 1, \dots$, we have, with probabilities $\delta_k = \frac{\delta}{2(1+k)^2}$,

$$\lambda_k = \frac{3\sqrt{\log \frac{1}{\delta_k}}}{\sqrt{\log \frac{1}{\delta_k}} + 3e^{\frac{k}{2}}}.$$

Plugging in this value to λ, we have

$$\frac{\lambda_k T_n}{2\left(1 - \frac{\lambda_k}{3}\right)} + \frac{1}{\lambda_k} \log \frac{1}{\delta_k} = \frac{T_n \sqrt{\log \frac{1}{\delta_k}}}{2e^{\frac{k}{2}}} + e^{\frac{k}{2}} \sqrt{\log \frac{1}{\delta_k}} + \frac{1}{3} \log \frac{1}{\delta_k}.$$

By inequality (54) and taking the union bound, the probability of $\sum_{t=1}^{n} X_t$ exceeding this value for any k, or equivalently exceeding the minimum of these values over k, is less than δ, i.e.,

$$\mathbb{P}\left(\exists n \in \mathbb{N} : \sum_{i=1}^{n} X_i \ge \min_{k=0,1,\dots} \left\{ \left(\frac{T_n}{2e^{\frac{k}{2}}} + e^{\frac{k}{2}}\right) \sqrt{\log \frac{1}{\delta_k}} + \frac{1}{3} \log \frac{1}{\delta_k} \right\}\right) \le \delta.$$

[9] `https://harinboy.github.io/posts/FreedmansInequality/`

The proof is completed by showing that the minimum is smaller than the desired value by choosing appropriate $k = k_n$ for all n. We separately deal with the cases $T_n \geq 1$ and $0 \leq T_n < 1$.

Case 1 ($T_n \geq 1$):

Choose $k_n = \lfloor \log T_n \rfloor$. Then, it holds that

$$\sqrt{T_n/e} \leq e^{\frac{k_n}{2}} \leq \sqrt{T_n}.$$

It follows that

$$\frac{T_n}{2e^{\frac{k_n}{2}}} + e^{\frac{k_n}{2}} \leq \frac{\sqrt{eT_n}}{2} + \sqrt{T_n} \leq 2\sqrt{T_n}.$$

Therefore, it holds that

$$\begin{aligned}
&\min_{k=0,1,\ldots} \left\{ \left(\frac{T_n}{2e^{\frac{k}{2}}} + e^{\frac{k}{2}} \right) \sqrt{\log \frac{1}{\delta_k}} + \frac{1}{3} \log \frac{1}{\delta_k} \right\} \leq \left(\frac{T_n}{2e^{\frac{k_n}{2}}} + e^{\frac{k_n}{2}} \right) \sqrt{\log \frac{1}{\delta_{k_n}}} + \frac{1}{3} \log \frac{1}{\delta_{k_n}} \\
\leq & 2\sqrt{T_n \log \frac{1}{\delta_{k_n}}} + \frac{1}{3} \log \frac{1}{\delta_{k_n}} \leq 2\sqrt{T_n \log \frac{2\left(1+\log^+ T_n\right)^2}{\delta}} + 2 \log \frac{2\left(1+\log^+ T_n\right)^2}{\delta},
\end{aligned}$$

where the last inequality uses that $\frac{1}{\delta_{k_n}} = \frac{2(1+k_n)^2}{\delta} \leq \frac{2\left(1+\log^+ T_n\right)^2}{\delta}$ and $\frac{1}{3} \leq 2$.

Case 2 ($0 \leq T_n < 1$):

In this case, we choose $k_n = 0$. Since $T_n \leq 1, T_n \leq \sqrt{T_n}$ holds. Using that $\log \frac{1}{\delta_0} \geq \log 2$, it holds that $\sqrt{\log \frac{1}{\delta_0}} \leq \frac{1}{\sqrt{\log 2}} \log \frac{1}{\delta_0}$. Therefore, it holds that

$$\begin{aligned}
\left(\frac{T_n}{2e^0} + e^0 \right) \sqrt{\log \frac{1}{\delta_0}} + \frac{1}{3} \log \frac{1}{\delta_0} \leq & \sqrt{\frac{T_n}{2} \log \frac{1}{\delta_0}} + \frac{1}{\sqrt{\log 2}} \log \frac{1}{\delta_0} + \frac{1}{3} \log \frac{1}{\delta_0} \\
\leq & 2\sqrt{T_n \log \frac{1}{\delta_0}} + 2 \log \frac{1}{\delta_0}.
\end{aligned}$$

Note that in this case, $\frac{1}{\delta_0} = \frac{2}{\delta} = \frac{2\left(1+\log^+ T_n\right)^2}{\delta}$. The proof is complete.

G.9 Proof of Lemma G.9

The proof follows the argument of Lemma 1 of Neu [2015]. In this proof, write $\mathbb{E}_{j-1}[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_{j-1}]$ and, for the one-step calculation, $\mathbb{E}^*_{j-1}[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_{j-1}, \hat{\xi}_{j,1}, \ldots, \hat{\xi}_{j,m}]$. Thus the loss vector, $\mathbf{p}_j$, γ_j, the weights and $\mathbb{I}[j \in S]$ are fixed under the latter conditioning, and the conditional distribution of i_j remains $\mathbf{p}_j$. For $p_{j,i} > 0$, the IX estimator $\tilde{\xi}_{j,i}$ satisfies

$$\begin{aligned}
\tilde{\xi}_{j,i} = \frac{\hat{\xi}_{j,i}}{p_{j,i} + \gamma_j} \cdot \mathbb{I}[i_j = i] \leq & \frac{\hat{\xi}_{j,i}}{p_{j,i} + \gamma_j \hat{\xi}_{j,i}} \cdot \mathbb{I}[i_j = i] = \frac{1}{2\gamma_j} \frac{2\gamma_j \cdot \hat{\xi}_{j,i}/p_{j,i}}{1 + \gamma_j \cdot \hat{\xi}_{j,i}/p_{j,i}} \cdot \mathbb{I}[i_j = i] \\
\leq & \frac{1}{\beta_j} \log\left(1 + \beta_j \bar{\xi}_{j,i}\right),
\end{aligned} \tag{55}$$

where the first inequality is obtained by $\hat{\xi}_{j,i} \in [0,1]$ for all $i \in [m]$ and $j \in [t]$, the last step is due to the inequality $\frac{z}{1+z/2} \leq \log(1+z)$ for $z \geq 0$ and the notations $\beta_j = 2\gamma_j$ and $\bar{\xi}_{j,i} = (\hat{\xi}_{j,i}/p_{j,i}) \cdot \mathbb{I}[i_j = i]$ for $p_{j,i} > 0$ are introduced for simplicity. If $p_{j,i} = 0$, define $\bar{\xi}_{j,i} = 0$. Since $\Pr(i_j = i \mid \mathcal{F}_{j-1}, \hat{\xi}_{j,1}, \ldots, \hat{\xi}_{j,m}) = 0$ in this case, $\tilde{\xi}_{j,i} = 0$, so the final inequality in (55) also holds. In both cases, $\mathbb{E}^*_{j-1}[\bar{\xi}_{j,i}] \leq \hat{\xi}_{j,i}$.

We define the notations $\tilde{\lambda}_j = \mathbb{I}[j \in S] \cdot \sum_{i=1}^m \alpha_{j,i} \tilde{\xi}_{j,i}$ and $\lambda_j = \mathbb{I}[j \in S] \cdot \sum_{i=1}^m \alpha_{j,i} \hat{\xi}_{j,i}$. Then, we conclude that $\mathbb{E}^*_{j-1}[\exp(\tilde{\lambda}_j)] \leq \exp(\lambda_j)$ for all $j \in [t]$. This is true for $j \in [t] \setminus S$ clearly. For

$j \in S$, we have

$$\begin{aligned}\mathbb{E}^*_{j-1}\left[\exp(\tilde{\lambda}_j)\right] &= \mathbb{E}^*_{j-1}\left[\exp\left(\sum_{i=1}^m \alpha_{j,i}\tilde{\xi}_{j,i}\right)\right] \overset{(55)}{\leq} \mathbb{E}^*_{j-1}\left[\exp\left(\sum_{i=1}^m \frac{\alpha_{j,i}}{\beta_j}\log\left(1+\beta_j\bar{\xi}_{j,i}\right)\right)\right] \\ &\leq \mathbb{E}^*_{j-1}\left[\exp\left(\sum_{i=1}^m \log\left(1+\alpha_{j,i}\bar{\xi}_{j,i}\right)\right)\right] = \mathbb{E}^*_{j-1}\left[\Pi_{i=1}^m\left(1+\alpha_{j,i}\bar{\xi}_{j,i}\right)\right] \\ &= \mathbb{E}^*_{j-1}\left[1+\sum_{i=1}^m \alpha_{j,i}\bar{\xi}_{j,i}\right] \leq 1+\sum_{i=1}^m \alpha_{j,i}\hat{\xi}_{j,i} \leq \exp\left(\sum_{i=1}^m \alpha_{j,i}\hat{\xi}_{j,i}\right) = \exp(\lambda_j),\end{aligned} \tag{56}$$

where the second inequality uses $\frac{\alpha_{j,i}}{\beta_j} \leq 1$ and $x\log(1+y) \leq \log(1+xy)$ for $y \geq -1$ and $x \in [0,1]$. The remaining steps use $\bar{\xi}_{j,i}\bar{\xi}_{j,k} = 0$ for $i \neq k$, $\mathbb{E}^*_{j-1}[\bar{\xi}_{j,i}] \leq \hat{\xi}_{j,i}$, and $1+z \leq e^z$.

Multiplying the conditional bound in (56) by $\exp(-\lambda_j)$ gives

$$\mathbb{E}_{j-1}\left[\exp(\tilde{\lambda}_j - \lambda_j)\right] = \mathbb{E}_{j-1}\left[\mathbb{E}^*_{j-1}\left[\exp(\tilde{\lambda}_j - \lambda_j)\right]\right] \leq 1.$$

Consequently, $Z_n = \exp(\sum_{j=1}^n(\tilde{\lambda}_j - \lambda_j))$ is a non-negative supermartingale with $Z_0 = 1$, i.e., $\mathbb{E}_{n-1}[Z_n] \leq Z_{n-1}$. In particular, $\mathbb{E}[Z_t] \leq 1$. By Markov's inequality,

$$\Pr\left[\sum_{j=1}^t(\tilde{\lambda}_j - \lambda_j) > \epsilon\right] \leq \mathbb{E}\left[\exp\left(\sum_{j=1}^t(\tilde{\lambda}_j - \lambda_j)\right)\right]\cdot\exp(-\epsilon) \leq \exp(-\epsilon)$$

holds for any $\epsilon > 0$. By setting $\exp(-\epsilon) = \delta$ and noting that $\sum_{j=1}^t(\tilde{\lambda}_j - \lambda_j) = \sum_{j\in S}(\tilde{\lambda}_j - \lambda_j)$, we complete the proof.

G.10 Proof of Lemma G.10

Let $H_j = Q_j\mathbb{I}[j \in M_t]$ and $X_j = H_j - \mathbb{E}_{j-1}[H_j]$. Since r_j is $\mathcal{F}_{j-1}$-measurable, so is $\mathbb{I}[j \in M_t] = \mathbb{I}[r_j \geq 2]$. By construction, $\mathbb{E}_{j-1}[X_j] = 0$, so $\{X_j\}_{j\in[t]}$ is a martingale difference sequence. Moreover, by (11), we have $\eta_{q,j} \leq \sqrt{\ln m}$. On $j \in M_t$, the multi-sample estimator satisfies $\tilde{s}_{j,i} \leq m$, and Lemma G.2 gives $\mathbb{E}_{j-1}[Q_j] \leq m\eta_{q,j}/(2r_j)$. Therefore, $0 \leq H_j \leq m^2\sqrt{\ln m}/2$ and $|X_j| \leq m^2\sqrt{\ln m}/2$, and

$$\mathbb{E}_{j-1}[X_j^2] \leq \mathbb{E}_{j-1}[H_j^2] \leq \frac{m^2\sqrt{\ln m}}{2}\mathbb{E}_{j-1}[H_j] \leq \frac{m^3\sqrt{\ln m}}{4}\frac{\eta_{q,j}}{r_j}\mathbb{I}[j \in M_t].$$

For any fixed $B > 0$, let $n_B = \max\left(\left\{j \leq t : \sum_{s=1}^j \frac{m}{r_s} \leq B\right\} \cup \{0\}\right)$. By (11) and Lemma G.3 with $\gamma = \frac{1}{2}$,

$$\begin{aligned}\sum_{j\in M_t}\frac{\eta_{q,j}}{r_j}\mathbb{I}\left[\sum_{s=1}^j \frac{m}{r_s} \leq B\right] &\leq \frac{\sqrt{\ln m}}{m}\sum_{j=1}^{n_B}\frac{\frac{m}{r_j}}{\sqrt{\sum_{s=1}^j\frac{m}{r_s}}} \\ &\overset{(24)}{\leq} \frac{2\sqrt{\ln m}}{m}\sqrt{\sum_{j=1}^{n_B}\frac{m}{r_j}} \leq \frac{2}{m}\sqrt{B\ln m}.\end{aligned}$$

Consequently,

$$\begin{aligned}\sum_{j=1}^t \mathbb{E}_{j-1}\left[\left(X_j\mathbb{I}\left[\sum_{s=1}^j\frac{m}{r_s} \leq B\right]\right)^2\right] &\leq \frac{m^3\sqrt{\ln m}}{4}\sum_{j\in M_t}\frac{\eta_{q,j}}{r_j}\mathbb{I}\left[\sum_{s=1}^j\frac{m}{r_s} \leq B\right] \\ &\leq \frac{m^3\sqrt{\ln m}}{4}\cdot\frac{2}{m}\sqrt{B\ln m} \\ &= \frac{m^2\ln m}{2}\sqrt{B} =: \Psi(B).\end{aligned}$$

Applying (25) with this choice of Ψ, we obtain, with probability at least $1-\delta$,

$$\sum_{j=1}^{t}(H_j - \mathbb{E}_{j-1}[H_j]) \le 2^{1/4}m\sqrt{\ln m\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}\ln\frac{m}{\delta}} + \frac{m^2\sqrt{\ln m}}{3}\ln\frac{m}{\delta}}. \tag{57}$$

Finally, since $\mathbb{I}[j \in M_t]$ is $\mathcal{F}_{j-1}$-measurable, $\sum_{j=1}^{t}\mathbb{E}_{j-1}[H_j] = \sum_{j\in M_t}\mathbb{E}_{j-1}[Q_j] \le \frac{m}{2}\sum_{j\in M_t}\frac{\eta_{q,j}}{r_j}$. Combining this inequality with (57) yields

$$\sum_{j\in M_t} Q_j \le 2^{1/4}m\sqrt{\ln m\sqrt{\sum_{j=1}^{t}\frac{m}{r_j}\ln\frac{m}{\delta}} + \frac{m^2\sqrt{\ln m}}{3}\ln\frac{m}{\delta} + \frac{m}{2}\sum_{j\in M_t}\frac{\eta_{q,j}}{r_j}},$$

which proves (32).

G.11 Proof of Lemma G.13

We adopt the notation introduced in Appendix G.6 and continue from (49). We utilize Lemma G.6 to get the time-uniform bound for $\texttt{term}(e_1)$ and $\texttt{term}(e_2)$. According to Hoeffding's inequality [Cesa-Bianchi and Lugosi, 2006, Lemma A.1.] and the conditions $\mathbb{E}_{t-1}[V_t] = 0$ and $|V_t| \le 1$, as well as $\mathbb{E}_{t-1}[V'_{t,i}] = 0$ and $|V'_{t,i}| \le 1$ for all $t \in \mathbb{N}^+, i \in [m]$, it follows that $\mathbb{E}_{t-1}[e^{sV_t}] \le \exp\left(s^2/2\right)$ and $\mathbb{E}_{t-1}[e^{sV'_{t,i}}] \le \exp\left(s^2/2\right)$ for any $s \in \mathbb{R}$ and for all $i \in [m]$. Therefore, the sequences $\{V_t\}_{t=1}^{\infty}$ and $\{V'_{t,i}\}_{t=1}^{\infty}$ for all $i \in [m]$ are conditionally 1-sub-Gaussian.

By Lemma G.6, with probability at least $1-\delta$, we have for all $t \in \mathbb{N}^+$,

$$\texttt{term}(\texttt{e}_1) = \sum_{j=1}^{t} V_j \le 2^{\frac{3}{4}}\sqrt{t\ln\frac{7(\ln 2t)^2}{2\delta}} \le 2\sqrt{t\left(\ln\frac{4}{\delta} + 2\ln\ln(2t)\right)}. \tag{58}$$

By Lemma G.6 and taking the union bound over $i \in [m]$, with probability at least $1-\delta$, we have for all $t \in \mathbb{N}^+$,

$$\texttt{term}(\texttt{e}_2) = \sum_{j=1}^{t} V'_{j,k^*} \le 2^{\frac{3}{4}}\sqrt{t\ln\frac{7m(\ln 2t)^2}{2\delta}} \le 2\sqrt{t\left(\ln\frac{4m}{\delta} + 2\ln\ln(2t)\right)}. \tag{59}$$

Substituting (58) and (59) into (49), and combining (48), we obtain the high-probability time-uniform bound holding for all $t \in \mathbb{N}^+$,

$$\epsilon_\phi(\bar{\mathbf{w}}_t, \bar{\mathbf{q}}_t) \le \frac{1}{t}U_w\left(t, \frac{\delta}{4}\right) + \frac{1}{t}U_q\left(t, \frac{\delta}{4}\right) + 4\sqrt{\frac{1}{t}\left(\ln\frac{16m}{\delta} + 2\ln\ln(2t)\right)}.$$

G.12 Proof of Lemma G.14

We adopt the notation introduced in Appendix G.2 and continue from (29).

For $\texttt{term}(\texttt{d})$, we extend Lemma G.9 to the time-uniform version as follows.

Lemma G.17. *Assume the sampling and measurability conditions of Lemma G.9 hold for all $j \in \mathbb{N}^+$: $\{\gamma_j\}_{j\ge1}$ is a positive, non-increasing process, and the non-negative weights satisfy $\alpha_{j,i} \le 2\gamma_j$. Let $\{r_t\}_{t\ge1}$ be a process generated according to Definition 3.1, and define $S_t = \{j \in [t] \mid r_j = 1\}$. Then, with probability at least $1-\delta$, for all t, we have*

$$\sum_{j\in S_t}\sum_{i=1}^{m}\alpha_{j,i}(\tilde{\xi}_{j,i} - \hat{\xi}_{j,i}) \le \ln\frac{4(\ln 2t)^2}{\delta}. \tag{60}$$

Similar to (31) and by Lemma G.17, we have, with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$\begin{aligned}\sum_{j\in S_t}\left(\frac{\eta_{q,j}}{2}+\gamma_j\right)\sum_{i=1}^m \tilde{s}_{j,i} &\overset{(14)}{=} \sum_{j\in S_t}\eta_{q,j}\sum_{i=1}^m \tilde{s}_{j,i} \overset{(60)}{\le} \sum_{j\in S_t}\eta_{q,j}\sum_{i=1}^m \hat{s}_{j,i} + \ln\frac{4(\ln 2t)^2}{\delta}\\ &\overset{(8)}{\le} \sum_{j\in S_t}\eta_{q,j}m + \ln\frac{4(\ln 2t)^2}{\delta} \overset{(11)}{=} \sum_{j\in S_t}\sqrt{\frac{m\ln m}{\sum_{s=1}^j \frac{1}{r_s}}} + \ln\frac{4(\ln 2t)^2}{\delta}\\ &\le \sum_{j\in S_t}\sqrt{\frac{m\ln m}{|[j]\cap S_t|}} + \ln\frac{4(\ln 2t)^2}{\delta} \le 2\sqrt{|S_t| m\ln m} + \ln\frac{4(\ln 2t)^2}{\delta}.\end{aligned} \tag{61}$$

For `term(e)`, we extend Lemma G.10 to the time-uniform version as follows.

Lemma G.18. *Let $\{r_t\}_{t\ge 1}$ follow Definition 3.1, so r_t is $\mathcal{F}_{t-1}$-measurable, and define $M_t = \{j \in [t] \mid r_j \ge 2\}$. Let $\{\eta_t\}_{t\ge 1}$ be positive and non-increasing, with η_t and $\mathbf{q}_t \in \Delta_m$ being $\mathcal{F}_{t-1}$-measurable for every $t \in \mathbb{N}^+$. For each round with $r_t \ge 2$, define*

$$\tilde{s}_{t,i} = \frac{\hat{s}_{t,i}}{q_{t,i} + (1-q_{t,i})\frac{r_t-1}{m-1}}\mathbb{I}[i \in C_t], \qquad Q_t = \frac{\eta_t}{2}\sum_{i=1}^m q_{t,i}\tilde{s}_{t,i}^2,$$

where $|\hat{s}_{t,i}| \le 1$ and the random set $C_t \subseteq [m]$ satisfies

$$\Pr[i \in C_t \mid \mathcal{F}_{t-1}] = q_{t,i} + (1-q_{t,i})\frac{r_t-1}{m-1}, \qquad i \in [m].$$

Set $Q_t = 0$ when $r_t = 1$. Then, with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$\sum_{j\in M_t} Q_j \le \eta_1 m^2 \ln\left(\frac{C_t}{\delta}\right) + m\sqrt{\eta_1 m \sum_{j\in M_t}\frac{\eta_j}{r_j}\ln\left(\frac{C_t}{\delta}\right)} + \frac{m}{2}\sum_{j\in M_t}\frac{\eta_j}{r_j},$$

where $C_t = 2\left(1+\ln^+\left(\frac{1}{\eta_1 m}\sum_{j\in M_t}\frac{\eta_j}{r_j}\right)\right)^2$.

We denote $Q_t = \frac{\eta_{q,t}}{2}\sum_{i=1}^m q_{t,i}(\tilde{s}_{t,i})^2$ for $t \in \mathbb{N}^+$. By Lemma G.18, applied with $\eta_t = \eta_{q,t}$, and $\sqrt{\frac{\ln m}{m}} \le \eta_{q,1} \le \sqrt{\ln m}$, we have, with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$\begin{aligned}\sum_{j\in M_t} Q_j &\le m\sqrt{\sqrt{\ln m}\, m \sum_{j\in M_t}\frac{\eta_{q,j}}{r_j}\ln\left(\frac{C_{1,t}}{\delta}\right)} + m^2\sqrt{\ln m}\ln\left(\frac{C_{1,t}}{\delta}\right) + \frac{m}{2}\sum_{j\in M_t}\frac{\eta_{q,j}}{r_j}\\ &\le m\sqrt{2\ln m\sqrt{\sum_{j\in M_t}\frac{m}{r_j}\ln\left(\frac{C_{1,t}}{\delta}\right)}} + m^2\sqrt{\ln m}\ln\left(\frac{C_{1,t}}{\delta}\right) + \sqrt{\sum_{j\in M_t}\frac{m}{r_j}\ln m},\end{aligned} \tag{62}$$

where the last inequality is obtained by (33) and $C_{1,t} = 2\left(1+\ln^+\left(\frac{2}{m}\sqrt{\sum_{j\in M_t}\frac{1}{r_j}}\right)\right)^2$.

For `term(f)`, we define $V_t = \langle \mathbf{q}_t, \hat{\mathbf{s}}_t\rangle - \hat{s}_{t,c_t}$ when $r_t = 1$ and $V_t = \langle \mathbf{q}_t, \hat{\mathbf{s}}_t\rangle - \langle \mathbf{q}_t, \tilde{\mathbf{s}}_t\rangle$ when $r_t \ge 2$. Since r_t and $\mathbf{q}_t$ are $\mathcal{F}_{t-1}$-measurable, the same conditional-expectation calculation as for fixed-time term (f) gives $\mathbb{E}_{t-1}[V_t] = 0$ for all $t \in \mathbb{N}^+$. By Lemma G.2 and (8), we know $|V_t| \le \frac{m}{2}$ for all $t \in \mathbb{N}^+$. Let $X_t = \frac{2V_t}{m}$, we have $|X_t| \le 1$ for all $t \in \mathbb{N}^+$. Moreover, for $r_t = 1$, we have

$$\mathbb{E}_{t-1}\left[X_t^2\right] = \frac{4}{m^2}\mathbb{E}_{t-1}\left[\left(\langle \mathbf{q}_t, \hat{\mathbf{s}}_t\rangle - \hat{s}_{t,c_t}\right)^2\right] \le \frac{4}{m^2},$$

and for $r_t \ge 2$, we have

$$\begin{aligned}\mathbb{E}_{t-1}\left[X_t^2\right] &= \frac{4}{m^2}\mathbb{E}_{t-1}\left[\left(\langle \mathbf{q}_t, \hat{\mathbf{s}}_t\rangle - \langle \mathbf{q}_t, \tilde{\mathbf{s}}_t\rangle\right)^2\right] = \frac{4}{m^2}\left(\mathbb{E}_{t-1}\left[\left(\langle \mathbf{q}_t, \tilde{\mathbf{s}}_t\rangle\right)^2\right] - \mathbb{E}_{t-1}\left[\left(\langle \mathbf{q}_t, \hat{\mathbf{s}}_t\rangle\right)^2\right]\right)\\ &\le \frac{4}{m^2}\mathbb{E}_{t-1}\left[\left(\langle \mathbf{q}_t, \tilde{\mathbf{s}}_t\rangle\right)^2\right] \overset{(23)}{\le} \frac{4}{mr_t}\mathbb{E}_{t-1}\left[\langle \mathbf{q}_t, \tilde{\mathbf{s}}_t\rangle\right] \le \frac{4}{mr_t}.\end{aligned}$$

Then, by Lemma G.7, we have, with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$\begin{aligned}\sum_{j=1}^{t} V_j =& \frac{m}{2}\sum_{j=1}^{t} X_j \le m \ln \frac{C_{2,t}}{\delta} + m\sqrt{\left(\frac{4}{m^2}\left(\sum_{j\in M_t}\frac{m}{r_j} + |S_t|\right)\right)\ln\frac{C_{2,t}}{\delta}} \\ \le& \sqrt{4\sum_{j=1}^{t}\frac{m}{r_j}\ln\frac{C_{2,t}}{\delta}} + m\ln\frac{C_{2,t}}{\delta}.\end{aligned} \tag{63}$$

where $C_{2,t} = 2\left(1+\ln^+\left(\frac{4}{m^2}\left(\sum_{j\in M_t}\frac{m}{r_j}+|S_t|\right)\right)\right)^2$.

For $\texttt{term(g)}$, we extend the fixed-weight construction preceding (36) to all times using Ville's inequality. For $n \in \mathbb{N}^+$ and $i \in [m]$, define $D_{n,i} := \sum_{j\in S_n}(\tilde{s}_{j,i} - \hat{s}_{j,i})$. For a fixed $B > 0$, let $\lambda_B = \sqrt{\ln(m)/B}$. The proof of Lemma G.9, applied with the weights $\alpha_{j,\ell}^{(B,i)} = \lambda_B \mathbb{I}\left[\sum_{s=1}^{j}\frac{m}{r_s} \le B\right]\mathbb{I}[\ell = i]$, shows that

$$Z_{n,i}^{(B)} = \exp\left(\lambda_B \sum_{j\in S_n}\mathbb{I}\left[\sum_{s=1}^{j}\frac{m}{r_s}\le B\right](\tilde{s}_{j,i}-\hat{s}_{j,i})\right)$$

is a non-negative supermartingale with $Z_{0,i}^{(B)} = 1$.

Set $B_k = m2^{k+1}$ for $k \ge 0$. Applying Ville's inequality with failure probabilities $6\delta/[m\pi^2(k+1)^2]$ and taking a union bound over all i, k, whose failure probabilities sum to δ, yields a single event of probability at least $1-\delta$ on which $Z_{t,i}^{(B_k)} \le m\pi^2(k+1)^2/(6\delta)$ for all t, i, k. Whenever $\sum_{j=1}^{t}\frac{m}{r_j} \le B_k$, all historical indicators equal one, and hence $D_{t,i} \le \lambda_{B_k}^{-1}\ln\frac{m\pi^2(k+1)^2}{6\delta}$.

If $\sum_{j=1}^{t}\frac{m}{r_j} < m$, then $S_t = \emptyset$ and $D_{t,i} = 0$. Otherwise choose $k = \lfloor \log_2\left(\sum_{j=1}^{t}\frac{1}{r_j}\right)\rfloor$, so $\sum_{j=1}^{t}\frac{m}{r_j} \le B_k \le 2\sum_{j=1}^{t}\frac{m}{r_j}$. Since $\sum_{j=1}^{t}\frac{m}{r_j} \le mt$, we have $\lambda_{B_k}^{-1} \le \sqrt{2}/(2\gamma_t)$ and $\pi^2(k+1)^2/6 \le \pi^2[\ln(2t)]^2/[6(\ln 2)^2] \le 4[\ln(2t)]^2$. Thus, on the same event, simultaneously for all $i \in [m]$ and $t \in \mathbb{N}^+$,

$$\sum_{j\in S_t}(\tilde{s}_{j,i}-\hat{s}_{j,i}) \le \frac{\sqrt{2}}{2\gamma_t}\ln\frac{4m(\ln 2t)^2}{\delta} = \sqrt{2\sum_{j=1}^{t}\frac{m}{r_j}\ln m} + \sqrt{\frac{2\sum_{j=1}^{t}\frac{m}{r_j}}{\ln m}}\ln\frac{4(\ln 2t)^2}{\delta}. \tag{64}$$

For $\texttt{term(h)}$, we define for all $i \in [m]$, $X_{t,i} = \frac{\tilde{s}_{t,i}-\hat{s}_{t,i}}{m}$ for $r_t \ge 2$ and $X_{t,i} = 0$ for $r_t = 1$. It holds that $\mathbb{E}_{t-1}[X_{t,i}] = 0$, $|X_{t,i}| \le 1$ for all $i \in [m]$ and $t \in \mathbb{N}^+$. When $r_t \ge 2$, we have

$$\begin{aligned}\mathbb{E}_{t-1}[X_{t,i}^2] =& \frac{\mathbb{E}_{t-1}[\tilde{s}_{t,i}^2] - \mathbb{E}_{t-1}[\hat{s}_{t,i}^2]}{m^2} \le \frac{\mathbb{E}_{t-1}[\tilde{s}_{t,i}^2]}{m^2} \\ \overset{(8)}{\le}& \frac{1}{m^2\left(q_{t,i}+(1-q_{t,i})\frac{r_t-1}{m-1}\right)} \le \frac{1}{m(r_t-1)} \le \frac{2}{mr_t}.\end{aligned}$$

Then by Lemma G.7 with failure probability δ/m for each coordinate and taking the union bound over $i \in [m]$, we have, with probability at least $1-\delta$, for all $i \in [m]$ and $t \in \mathbb{N}^+$,

$$\begin{aligned}\sum_{j\in M_t}(\tilde{s}_{j,i}-\hat{s}_{j,i}) =& m\sum_{j=1}^{t}X_{j,i} \le 2\sqrt{\sum_{j\in M_t}\frac{2m}{r_j}\ln\frac{mC_{3,t}}{\delta}} + 2m\ln\frac{mC_{3,t}}{\delta} \\ \le& 2\sqrt{\sum_{j\in M_t}\frac{2m}{r_j}}\left(\sqrt{\ln m} + \sqrt{\ln\frac{C_{3,t}}{\delta}}\right) + 2m\ln m + 2m\ln\frac{C_{3,t}}{\delta}\end{aligned} \tag{65}$$

where $C_{3,t} = 2\left(1+\ln^+\left(\sum_{j\in M_t}\frac{2}{mr_j}\right)\right)^2$.

Combining (61)-(65) with (29), allocating failure probability $\delta/5$ to each term, and taking the union bound, we have with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$
\begin{aligned}
&\mathrm{Regret}'_{q,t} \\
&\le \sqrt{\sum_{j=1}^{t} \frac{m}{r_j}} \left((1+\sqrt{2})\sqrt{\ln m} + 2\sqrt{\ln \frac{5C_{2,t}}{\delta}} \right) + 2\sqrt{|S_t| m \ln m} + 4\sqrt{\sum_{j \in M_t} \frac{m}{r_j} \ln m} \\
&\quad + m\sqrt{2\ln m \sqrt{\sum_{j \in M_t} \frac{m}{r_j}} \ln\left(\frac{5C_{1,t}}{\delta}\right)} + 2\sqrt{2 \sum_{j \in M_t} \frac{m}{r_j} \ln \frac{5C_{3,t}}{\delta}} + m \ln \frac{5C_{2,t}}{\delta} \\
&\quad + \left(1 + \sqrt{\frac{2\sum_{j=1}^{t} \frac{m}{r_j}}{\ln m}} \right) \ln \frac{20(\ln 2t)^2}{\delta} + m^2 \sqrt{\ln m} \ln\left(\frac{5C_{1,t}}{\delta}\right) + 2m \ln m + 2m \ln \frac{5C_{3,t}}{\delta} \\
&\overset{(38)}{\le} \sqrt{\sum_{j=1}^{t} \frac{m}{r_j}} \left(8\sqrt{\ln m} + 2\sqrt{\ln \frac{5C_{2,t}}{\delta}} \right) + m\sqrt{2\ln m \sqrt{\sum_{j \in M_t} \frac{m}{r_j}} \ln\left(\frac{5C_{1,t}}{\delta}\right)} \\
&\quad + 2\sqrt{2 \sum_{j \in M_t} \frac{m}{r_j} \ln \frac{5C_{3,t}}{\delta}} + m \ln \frac{5C_{2,t}}{\delta} + \left(1 + \sqrt{\frac{2\sum_{j=1}^{t} \frac{m}{r_j}}{\ln m}} \right) \ln \frac{20(\ln 2t)^2}{\delta} \\
&\quad + m^2 \sqrt{\ln m} \ln\left(\frac{5C_{1,t}}{\delta}\right) + 2m \ln m + 2m \ln \frac{5C_{3,t}}{\delta}.
\end{aligned}
$$

G.13 Proof of Lemma G.15

We adopt the notation introduced in Appendix G.3 and continue from (45). To establish a high-probability bound for all $t \in \mathbb{N}^+$, we use Lemma G.6, which provides a time-uniform bound. By Hoeffding's inequality, $\mathbb{E}_{t-1}[V_t] = 0$ and (46), we know $\mathbb{E}_{t-1}[e^{sV_t}] \le \exp\left(16D^2G^2s^2\right)$ for any $s \in \mathbb{R}$. Thus, the sequence $\{V_t\}_{t=1}^{\infty}$ is conditionally $4\sqrt{2}DG$-sub-Gaussian. By Lemma G.6, with probability at least $1-\delta$, we have for all $t \in \mathbb{N}^+$,

$$
\sum_{j=1}^{t} V_j \le 2^{\frac{3}{4}} 4\sqrt{2} DG \sqrt{t \ln \frac{7(\ln 2t)^2}{2\delta}} \le 10DG\sqrt{t\left(\ln \frac{4}{\delta} + 2\ln\ln(2t)\right)}. \tag{66}
$$

Combining (41), (45) and (66), we obtain that with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$
\sum_{j=1}^{t} \phi(\mathbf{w}_j, \mathbf{q}_j) - \min_{\mathbf{w} \in \mathcal{W}} \sum_{j=1}^{t} \phi(\mathbf{w}, \mathbf{q}_j) \le DG\sqrt{t}\left(2\sqrt{10} + 10\sqrt{\ln \frac{4}{\delta} + 2\ln\ln(2t)}\right).
$$

G.14 Proof of Lemma G.17

Under Definition 3.1, r_j and hence $\mathbb{I}[r_j = 1]$ are $\mathcal{F}_{j-1}$-measurable for every $j \in \mathbb{N}^+$. Following the notation and results in Section G.9, with $\mathbb{I}[j \in S]$ replaced by $\mathbb{I}[r_j = 1]$, we know the process $Z_t = \exp(\sum_{j=1}^{t}(\tilde{\lambda}_j - \lambda_j))$ is a supermartingale. To obtain a time-uniform bound, we follow Lee et al. [2025] by partitioning $\mathbb{N}^+$ into intervals $I_k = \{t_k, t_k+1, \ldots, t_{k+1}-1\}$ where $t_k = 2^k$ for $k \ge 0$. Applying Ville's inequality over each I_k, we have for any $\delta > 0$,

$$
\mathbb{P}\left(\exists n \in I_k : Z_n \ge \frac{\pi^2 (k+1)^2}{6\delta} \right) \le \frac{6\delta}{\pi^2 (k+1)^2}.
$$

Equivalently, we obtain

$$
\mathbb{P}\left(\exists n \in I_k : \sum_{j=1}^{n} (\tilde{\lambda}_j - \lambda_j) \ge \ln \frac{\pi^2 (k+1)^2}{6\delta} \right) \le \frac{6\delta}{\pi^2 (k+1)^2}.
$$

By $\frac{\pi^2(k+1)^2}{6} = \frac{\pi^2(\log_2 2t_k)^2}{6} = \frac{\pi^2}{6(\ln 2)^2}(\ln 2t_k)^2 \le 4(\ln 2n)^2$, we get

$$\mathbb{P}\left(\exists n \in I_k : \sum_{j=1}^{n}(\tilde{\lambda}_j - \lambda_j) \ge \ln\frac{4(\ln 2n)^2}{\delta}\right) \le \frac{6\delta}{\pi^2\left(k+1\right)^2}. \tag{67}$$

Finally, by taking the union bound over $k \ge 0$ and using the fact $\sum_{k=0}^{\infty}\frac{1}{(k+1)^2} = \frac{\pi^2}{6}$, we complete the proof.

G.15 Proof of Lemma G.18

We define $H_t = Q_t \cdot \mathbb{I}[r_t \ge 2]$ and $M_t = \{j | r_j \ge 2, j \in [t]\}$. Since r_t is $\mathcal{F}_{t-1}$-measurable, so is $\mathbb{I}[r_t \ge 2]$, and $\mathbb{E}_{t-1}[H_t] = \mathbb{I}[r_t \ge 2]\mathbb{E}_{t-1}[Q_t]$. First, we note that $|\tilde{s}_{t,i}| \le m$ for all $i \in [m]$ and $t \in \mathbb{N}^+$ with $r_t \ge 2$. It follows that $0 \le Q_t \le \frac{\eta_1}{2}m^2$, and that $\mathbb{E}_{t-1}[Q_t] \le \frac{m}{2}\frac{\eta_t}{r_t}$ for all $t \in \mathbb{N}^+$ with $r_t \ge 2$ by $|\hat{s}_{t,i}| \le 1$, the conditional inclusion probabilities, and Lemma G.2. Let $X_t = \frac{2(H_t - \mathbb{E}_{t-1}[H_t])}{\eta_1 m^2}$. Since $\eta_1 > 0$ is $\mathcal{F}_0$-measurable, $\mathbb{E}_{t-1}[X_t] = 0$. It holds that $X_t \le 1$ for all $t \in \mathbb{N}^+$, and its variance when $r_t \ge 2$ is bounded as

$$\begin{aligned}
\mathbb{E}_{t-1}[X_t^2] &= \frac{4}{\eta_1^2 m^4}\mathbb{E}_{t-1}\left[\left(H_t - \mathbb{E}_{t-1}\left[H_t\right]\right)^2\right] \\
&= \frac{4}{\eta_1^2 m^4}\left(\mathbb{E}_{t-1}\left[\left(H_t\right)^2\right] - \mathbb{E}_{t-1}\left[H_t\right]^2\right) \le \frac{4}{\eta_1^2 m^4}\mathbb{E}_{t-1}\left[\left(H_t\right)^2\right] \le \frac{1}{\eta_1 m}\frac{\eta_t}{r_t}.
\end{aligned}$$

When $r_t = 1$, $H_t = X_t = 0$. Thus the conditional variance sum is bounded by $\frac{1}{\eta_1 m}\sum_{j\in M_t}\frac{\eta_j}{r_j}$. Applying Lemma G.7, we conclude that with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$\begin{aligned}
\sum_{j\in M_t}\left(Q_j - \mathbb{E}_{j-1}\left[Q_j\right]\right) &= \sum_{j=1}^{t}\left(H_j - \mathbb{E}_{j-1}\left[H_j\right]\right) = \frac{\eta_1 m^2}{2}\sum_{j=1}^{t}X_j \\
&\le \eta_1 m^2 \ln\left(\frac{C_t}{\delta}\right) + m\sqrt{\eta_1 m \sum_{j\in M_t}\frac{\eta_j}{r_j}\ln\left(\frac{C_t}{\delta}\right)}.
\end{aligned} \tag{68}$$

where $C_t = 2\left(1 + \ln^+\left(\frac{1}{\eta_1 m}\sum_{j\in M_t}\frac{\eta_j}{r_j}\right)\right)^2$.

Consequently, we have, with probability at least $1-\delta$, for all $t \in \mathbb{N}^+$,

$$\begin{aligned}
\sum_{j\in M_t} Q_j &= \sum_{j\in M_t}\left(Q_j - \mathbb{E}_{j-1}[Q_j]\right) + \sum_{j\in M_t}\mathbb{E}_{j-1}[Q_j] \\
&\overset{(68)}{\le} \eta_1 m^2 \ln\left(\frac{C_t}{\delta}\right) + m\sqrt{\eta_1 m \sum_{j\in M_t}\frac{\eta_j}{r_j}\ln\left(\frac{C_t}{\delta}\right)} + \frac{m}{2}\sum_{j\in M_t}\frac{\eta_j}{r_j}.
\end{aligned}$$

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract and introduction have clearly stated the contributions of this paper.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: Appendix D discusses the scope of the FSQ model.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

Justification: The assumptions of any theoretical result are clearly introduced in Section 2. The proofs for all theoretical results can be found in Appendix G.

Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: The setup of our experiments is clearly introduced in Section 4 and Appendix C, which is sufficient for reproducing our experimental results.

   Guidelines:

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

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

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

Answer: [No]

Justification: The code is being used in our ongoing research and has therefore not been publicly released at this stage. All datasets are publicly available, and sufficient implementation and experimental details for reproducing the results are provided in Appendix C.

Guidelines:

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

6. **Experimental setting/details**

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

Answer: [Yes]

Justification: The setup of our experiments is clearly introduced in Section 4 and Appendix C.

Guidelines:

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

7. **Experiment statistical significance**

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

Answer: [Yes]

Justification: Synthetic and diabetes risk curves show means and standard-deviation shading over five independent runs. FEMNIST and estimator-variance tables report means and standard deviations over five seeds. CivilComments results are reported for one run per method.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

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

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: We have provided the relevant information in Appendix C.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: The research conducted in the paper conforms with the NeurIPS Code of Ethics.

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: This work is primarily theoretical and aims to improve robustness across groups under resource constraints. Potential positive impacts include improved worst-group performance in applications such as federated learning, content moderation, and medical annotation. Potential negative impacts may arise if the method is deployed with

poorly specified groups or privacy-sensitive subgroup information. These risks should be addressed through careful group definition, privacy protection, and application-specific validation before deployment.

Guidelines:

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

11. **Safeguards**

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

    Answer: [N/A]

    Justification: This is primarily a theoretical paper with no such risks.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: We cite the creators of the dataset used in our experiments, and we explicitly respect the license and terms of use.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.

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

13. **New assets**

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

    Answer: [N/A]

    Justification: This work does not release new assets.

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

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

16. **Declaration of LLM usage**

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

    Answer: [N/A]

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

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.