Monotone Individual Fairness

Yahav Bechavod [1]

Abstract

We revisit the problem of online learning with individual fairness, where an online learner strives to maximize predictive accuracy while ensuring that similar individuals are treated similarly. We first extend the frameworks of Gillen et al. (2018); Bechavod et al. (2020), which rely on feedback from human auditors regarding fairness violations, to allow for auditing schemes that can aggregate feedback from any number of auditors, using a rich class we term *monotone aggregation functions*, for which we also prove a useful characterization. Using our generalized framework, we present an oracle-efficient algorithm guaranteeing a bound of $\mathcal{O}(T^{\frac{3}{4}})$ simultaneously for regret and number of fairness violations. We then study an online classification setting where label feedback is available for positively-predicted individuals only, and present an algorithm guaranteeing a bound of $\mathcal{O}(T^{\frac{5}{6}})$ simultaneously for regret and number of fairness violations. In both settings, our algorithms improve on the best known bounds for oracle-efficient algorithms. Furthermore, our algorithms offer significant improvements in computational efficiency, greatly reducing the number of required calls to an (offline) optimization oracle, as opposed to previous algorithms which required T such calls every round.

1. Introduction

As algorithms are increasingly ubiquitous in variety of domains where decisions are highly consequential to human lives — including lending, hiring, education, and healthcare — there is by now a vast body of research aimed at formalizing, exploring, and analyzing different notions of fairness, and suggesting new algorithms capable of obtaining them in conjunction with high predictive accuracy. The majority of this body of work takes a *statistical group fairness* approach, where a collection of groups in the population is defined (often according to "protected attributes"), and the aim is to then approximately equalize a chosen statistic of the predictor (such as overall error rate, false positive rate, etc.) across them. From the perspective of the *individual*, however, group fairness notions fail to deliver meaningful guarantees, as they are *aggregate* in nature, only binding over averages over many people. This was also pointed out by Dwork et al. (2012) original "catalog of evils".

Furthermore, the majority of the work in algorithmic fairness follows *statistical data generation* assumptions, where data points are assumed to arrive in i.i.d. fashion from a distribution, in either a batch setting, an online setting, or a bandit setting. Many domains where fairness is a concern, however, may not (and often do not) follow such assumptions, due to, for instance: (1) strategic effects (e.g. individuals attempting to modify their features to "better fit" a specific policy in hopes of receiving more favorable outcomes, or individuals who decide whether to even apply based on the policy which was deployed) (see, e.g., Dranove et al. (2003); Dee et al. (2019); Gonzalez-Lira & Mobarak (2019); Greenstone et al. (2020)), (2) distribution shifts over time (e.g. the ability to repay a loan may be affected by changes to the economy or recent events), (3) adaptivity to previous decisions (e.g. if an individual receives a loan, that may affect the ability to repay future loans by this individual or her vicinity), (4) one-sided label feedback (a college can only track the academic performance of students who have been admitted in the first place).

The seminal work of Dwork et al. (2012) advocates for a different view, approaching fairness from the perspective of the individual. In the core of their formulation is the assertion that "similar individuals should be treated similarly". Formally, they require that a (randomized) predictor obey a Lipschitz condition, where similar predictions are made on individuals deemed similar, according to a task specific *metric*. As Dwork et al. (2012) acknowledge, however, the availability of such metrics is one of the most challenging aspects in their framework. In many domains, it seems, it is not clear how such metrics can be elicited or learned.

A recent line of work, starting with Gillen et al. (2018), suggests an elegant framework aiming at the above two issues precisely, as they study an *adversarial* online learning prob-

[1] Department of Computer and Information Sciences, University of Pennsylvania. Correspondence to: Yahav Bechavod <yahav@seas.upenn.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

lem, where the learner receives additional feedback from an *auditor*. Specifically, the auditor is tasked with identifying *fairness violations* (pairs of individuals who she deems similar, and were given very different assessments) made by the learner, and reporting them in real time. In their framework, they assume the metric according to which the auditor reports her perceived violations is *unknown* to the learner. They assert that while in many cases, enunciating the *exact* metric might be a difficult task for the auditor, she is likely to "know unfairness when she sees it". More generally, Gillen et al. (2018) operate in a linear contextual bandit setting, and the goal in their setting is to achieve low regret while also minimizing the number of fairness violations made by the learner. Importantly, they assume that the metric takes a specific parametric form (Mahalanobis distance), and that the auditor must identify *all* existing violations.

Their framework has since been extended by Bechavod et al. (2020), who studied the problem absent a linear payoff structure, dispensed with the need to make parametric assumptions on the metric (in fact, their formulation even allows for a similarity function which does not take *metric* form), allowed for different auditors at different timesteps, and only required any auditor to report a *single* violation, in case one or more exist. Finally, Bechavod & Roth (2023) extended the framework by exploring majority-based auditing schemes, capable of incorporating feedback from multiple auditors, with potentially conflicting opinions, and studying the problem under partial information.

In this work, we make progress on both the conceptual and technical fronts of learning with individual fairness. We first introduce a novel framework for auditing for unfairness, which generalizes upon the ones in previous works (Gillen et al. (2018); Bechavod et al. (2020); Bechavod & Roth (2023)), and is based on detecting violations by applying a rich class of aggregation functions on feedback from multiple auditors. In particular, our framework will allow for a *different* number and identity of auditors at each timestep, and different aggregation functions. Using our framework, we present new oracle-efficient algorithms for both the full information and partial information settings of online learning with individual fairness. Our algorithms are based on carefully combining the objectives of accuracy and fairness in a Lagrangian formulation, which allows us to improve over the best known bounds in both settings (Bechavod et al., 2020; Bechavod & Roth, 2023). Importantly, our algorithms greatly reduce the computational complexity of previous approaches, as we present a new approach and analysis based on distinguishing between the tasks of constraint elicitation and objective minimization.

1.1. Overview of Results

We provide an overview of our results and a roadmap for the paper. We identify a natural class of auditing schemes we term *monotone auditing schemes*, which is capable of leveraging feedback from any number of auditors regarding fairness violations, and aggregate it using a broad class of aggregation functions. We formalize and prove a useful characterization for such auditing schemes (Section 2).

We define an online learning framework with individual fairness feedback from monotone auditing schemes, generalizing the ones in Gillen et al. (2018); Bechavod et al. (2020); Bechavod & Roth (2023) (Section 3). We then define a regularized Lagrangian loss function, which is able, on every timestep, to carefully combine the objectives of accuracy and fairness (Section 3.3).

Using our Lagrangian formulation, we present an oracle-efficient algorithm, based on a reduction to Context-FTPL (Syrgkanis et al., 2016), guaranteeing a bound of at most $O(T^{\frac{3}{4}})$ for each of regret and number of fairness violations. Importantly, our construction will only require making $\tilde{O}(\epsilon^{-2})$ calls to an optimization oracle on every round. Thus improving on the best known bounds and oracle complexity by Bechavod et al. (2020). (Section 4).

We then consider a more challenging setting where label feedback is available for positively-predicted individuals only. We present an oracle-efficient algorithm, leveraging our Lagrangian formulation along with a reduction to Context-Semi-Bandit-FTPL (Syrgkanis et al., 2016), guaranteeing a bound of $O(T^{\frac{5}{6}})$ for each of regret and number of fairness violations, while only requiring $\tilde{O}(\epsilon^{-2} + k^2 T^{\frac{1}{3}})$ calls to an optimization oracle on every round. Thus improving on the best known bounds and oracle complexity by Bechavod & Roth (2023). (Section 5).

We conclude with a discussion and directions for future research (Section 6).

1.2. Related Work

Our work is primarily related to two strands of research: individual fairness, and online learning with long-term constraints. The seminal work of Dwork et al. (2012) introduced the notion of individual fairness. They leave open the question of the similarity metric. Rothblum & Yona (2018) study an offline setting where the metric is assumed to be *known*, and suggest algorithms for learning predictors that give PAC-style accuracy and individual fairness guarantees. Kim et al. (2018) study a group-relaxation of individual fairness in a batch setting with access to a (noisy) oracle specifying distances between groups. Ilvento (2020) suggests learning the metric using a combination of comparison and distance queries to auditors. Our framework will not require querying numerical distance queries. Jung et al. (2021)

study a batch setting, eliciting similarity constraints from a set of "stakeholders", and prove generalization bounds for both accuracy and fairness. Finally, as elaborated on in the introduction, our work is closely related to Gillen et al. (2018); Bechavod et al. (2020); Bechavod & Roth (2023).

For the problem of online convex optimization with a static, known ahead of time, set of constraints, Zinkevich (2003) first proposed (projection-based) online gradient descent. In addition to requiring perfect knowledge of the constraints (rather than only reported *violations*), online gradient descent entails a projection step on each round, which may be computationally demanding if the set of constraints is complex. The problem of online learning with long-term constraints, hence, offers a *relaxation* with respect to constraint violation — the learner's goal is to minimize its regret, while being allowed to violate the constraints *at a vanishing rate*. Works in this field consider three main scenarios: constraints that are static and are known ahead of time (Mahdavi et al., 2012; Jenatton et al., 2016; Yuan & Lamperski, 2018; Yu & Neely, 2020), arrive stochastically in real time (Yu et al., 2017; Wei et al., 2020), or arrive adversarially (Mannor et al., 2009; Sun et al., 2017; Chen et al., 2017; Liakopoulos et al., 2019; Chen & Giannakis, 2019; Cao & Liu, 2019; Yi et al., 2020).

In our setting, however, the learner *will not know* the set of constraints at any round (as they will be held *implicitly* by the auditors), but rather has weaker access, only through reported fairness *violations*. Additionally, the literature on online learning with long-term constraints primarily pertains to online convex optimization. When instantiated over the simplex over a set of experts (as will be our case, with a hypothesis class $\mathcal{H}$), the proposed algorithms in this literature generally require maintaining and updating on each round the set of weights on $\mathcal{H}$ *explicitly*, which can be computationally prohibitive for large hypothesis classes. We hence strive to develop *oracle-efficient* algorithms, which, given access to an (offline) optimization oracle, will dispense us of the need to explicitly maintain these weights. We refer the reader to Appendix A for an extended related work section.

2. Individual Fairness and Monotone Auditing Schemes

We begin by defining notation we will use throughout this work. We denote a feature space by $\mathcal{X}$, and a label space by $\mathcal{Y}$, where we will focus on the case where $\mathcal{X} = \mathbb{R}^d$, and $\mathcal{Y} = \{0, 1\}$. We denote by $\mathcal{H} : \mathcal{X} \to \mathcal{Y}$ a hypothesis class of binary predictors, and assume that $\mathcal{H}$ contains a constant classifier. For the purpose of achieving more favorable trade-offs between accuracy and fairness, we will allow a learner to deploy *randomized* predictors from $\Delta\mathcal{H} : \mathcal{X} \to [0, 1]$. In the settings we will focus on, $\mathcal{X}$ will generally consist of features pertaining to human individuals (e.g. income, repayment history, debt), and $\mathcal{Y}$ will encode a target variable a learner wishes to predict correctly (e.g. defaulting on payments). From here on, we will denote k-tuples (corresponding to k individuals) of features and labels by $\bar{x} = (\bar{x}^1, \ldots, \bar{x}^k) \in \mathcal{X}^k$, $\bar{y} = (\bar{y}^1, \ldots, \bar{y}^k) \in \mathcal{Y}^k$.

Next, we define a fairness violation, following the notion of individual fairness by Dwork et al. (2012).

Definition 2.1 (Fairness violation)**.** Let $\alpha \geq 0$ and let $d : \mathcal{X} \times \mathcal{X} \to [0, 1]$.[1] We say that a policy $\pi \in \Delta\mathcal{H}$ has an α-fairness violation (or simply "α-violation") on $(x, x') \in \mathcal{X}^2$ with respect to d if

$$\pi(x) - \pi(x') > d(x, x') + \alpha.$$

where $\pi(x) = \Pr_{h \sim \pi}[h(x) = 1]$.

Note that Definition 2.1 also encodes the *direction* of the violation (which individual received the higher prediction), as this will be important in our construction.[2]

We next define a fairness auditor, having access to a set of individuals and their assigned predictions, tasked with reporting its perceived violations.

Definition 2.2 (Auditor)**.** We define a fairness auditor $j : \Delta\mathcal{H} \times \mathcal{X}^k \times \mathbb{R}^+ \to \{0, 1\}^{k \times k}$ as

$$\Big[j(\pi, \bar{x}, \alpha)\Big]_{l,r} := \begin{cases} 1 & \pi(\bar{x}^l) - \pi(\bar{x}^r) > d^j(\bar{x}^l, \bar{x}^r) + \alpha \\ 0 & \text{otherwise} \end{cases},$$

where $d^j : \mathcal{X} \times \mathcal{X} \to [0, 1]$ is auditor j's (implicit) distance function. if $j(\pi, \bar{x}, \alpha) = 0^{k \times k}$, we define $\vec{j}(\pi, \bar{x}, \alpha) :=$ Null. Otherwise, we define $\vec{j}(\pi, \bar{x}, \alpha) := (\bar{x}^l, \bar{x}^r)$, where $(l, r) \in [k]^2$ are (arbitrarily) selected such that $[j(\pi, \bar{x}, \alpha)]_{l,r} = 1$. We denote the space of all such auditors by $\mathcal{J}$.

Remark 2.3. In its most general form, an auditor returns a k-by-k matrix encoding its objections with respect to a specific policy on a set of individuals. We will later discuss notable cases where there is no requirement for the auditor to actually enunciate the entire matrix, but rather only detect the existence of a *single* violation, in case one or more exist.

[1] d represents a function specifying the auditor's judgement of the "similarity" between individuals in a specific context. We do not require that d be a metric: only that it be non-negative and symmetric.

[2] Technically speaking, since the learner will know the predictions $\pi(x), \pi(x')$, the auditor only has to report the (unordered) pair $\{x, x'\}$ in case he perceives a violation has occurred on it — the direction of the violation can then be inferred by the learner, since she knows which of x, x' was given a higher prediction under π. It will nevertheless be convenient in our construction to explicitly incorporate the direction in the definition of a fairness violation.

2.1. Monotone individual fairness auditing schemes

So far, our formulation of individual fairness in auditing follows the one in Gillen et al. (2018); Bechavod et al. (2020). In this work, we suggest a more general approach to auditing for unfairness, that will allow us to aggregate over the preferences of multiple auditors, with different (and even conflicting) opinions. For this purpose, we will consider aggregation functions $f : \left(\{0,1\}^{k\times k}\right)^m \to \{0,1\}^{k\times k}$ that map the outputs of multiple auditors $\bar{j} = (j^1, \dots, j^m) \in \mathcal{J}^m$ into a single output matrix. We denote the space of all such functions by $\mathcal{F}$. We proceed to define an auditing scheme, which takes as input the judgements of a panel of auditors, and decides on which pairs a fairness violation has occurred, according to a predefined aggregation function.

Definition 2.4 (Auditing scheme)**.** Let $m \in \mathbb{N} \setminus \{0\}$. We define an auditing scheme $\mathcal{S} : \Delta\mathcal{H} \times \mathcal{X}^k \times \mathbb{R}^+ \times \mathcal{F} \times \mathcal{J}^m \to \{0,1\}^{k\times k}$ as

$$\mathcal{S}(\pi, \bar{x}, \alpha, f, \bar{j}) := f(j^1(\pi, \bar{x}, \alpha), \dots, j^m(\pi, \bar{x}, \alpha)).$$

If $\mathcal{S}(\pi, \bar{x}, \alpha, f, \bar{j}) = 0^{k\times k}$, we define $\vec{\mathcal{S}}(\pi, \bar{x}, \alpha, f, \bar{j}) :=$ Null. Otherwise, we define $\vec{\mathcal{S}}(\pi, \bar{x}, \alpha, f, \bar{j}) := (\bar{x}^l, \bar{x}^r)$, where $(l, r) \in [k]^2$ are (arbitrarily) selected such that $[\mathcal{S}(\pi, \bar{x}, \alpha, f, \bar{j})]_{l,r} = 1$. We denote the space of all such auditing schemes by $\bar{\mathcal{S}}$.

As we are particularly interested in individual fairness auditing schemes, we will henceforth restrict our attention to a subclass of $\mathcal{F}$, where the value of each entry in the aggregate matrix is only affected by the corresponding entries in all of the input matrices, and aggregation of individual auditors outputs is done in a similar manner, regardless of individuals' position in $\bar{x}$.

Definition 2.5 (Independent aggregation functions)**.** We define the class $\mathcal{F}^{Ind} \subseteq \mathcal{F}$ of independent aggregation functions as functions of the form

$$\forall (l,r) \in [k]^2 : \left[f(A^1, \dots, A^m)\right]_{l,r} = \bar{f}(A^1_{l,r}, \dots, A^m_{l,r}),$$

where $A^1, \dots, A^m \in \{0,1\}^{k\times k}$, $\bar{f} : \{0,1\}^m \to \{0,1\}$.

We will next consider the case where $A^1, \dots, A^m$ are the output matrices of auditors $j^1, \dots, j^m$, respectively.

Restricting our attention to $\mathcal{F}^{Ind}$, however, still seems insufficient. In particular, it still permits selecting $f \in \mathcal{F}^{Ind}$ under which having additional auditors object the predictions made on a particular pair can actually result in changing an aggregate decision of reporting a violation to one where no violation is reported (for example, consider f that is defined such that $\bar{f} = 1$ if and only if *exactly* one of the m auditors objects). To remedy this, we will focus on independent aggregation functions that are *monotone* — which we next formally define. We begin by defining an ordering over the space of all possible objection profiles by a set of m auditors on a fixed pair of individuals (x^l, x^r).

Definition 2.6 (Aggregation order)**.** Let $v, v' \in \{0,1\}^m$. We say that v' constitutes a stronger objection profile than v, and denote $v \preccurlyeq v'$, if $\forall i \in [m], v_i \leq v'_i$.

Intuitively, $v \preccurlyeq v'$ if and only if every auditor who objected to the predictions of π on (x^l, x^r) with sensitivity level α resulting in v, still objects the predictions of a policy π' on (x^l, x^r) with sensitivity level α' resulting in v'. This will be the case in two important scenarios: when we fix π and decrease the auditors' sensitivity α for reporting violations, or alternatively when we fix α and consider π' such that $\pi'(x^l) - \pi'(x^r) \geq \pi(x^l) - \pi(x^r)$.

Next, we define the classes of monotone aggregation functions (and schemes) in line with the discussion above.

Definition 2.7 (Monotone aggregation functions)**.** We define the class $\mathcal{F}^{Mon} \subseteq \mathcal{F}^{Ind}$ of monotone aggregation functions, as functions $f \in \mathcal{F}^{Ind}$ such that

$$\forall v, v' \in \{0,1\}^m : v \preccurlyeq v' \implies \bar{f}(v) \leq \bar{f}(v').$$

Definition 2.8 (Monotone auditing scheme)**.** We say an auditing scheme $\mathcal{S}$ is monotone if it uses an aggregation function $f \in \mathcal{F}^{Mon}$.[3,4]

2.2. Characterizing monotone individual fairness auditing schemes

In what follows, we prove a characterization of monotone auditing schemes, when auditors are queried for individual fairness violations. As we will see, querying specifically for such violations, in combination with an aggregation scheme that is monotone, will imply that for every pair of individuals (x^l, x^r), the aggregate decision will always be equivalent to the decision of the same *single* auditor, regardless of the deployed policy and selected sensitivity parameter.In what follows, we will use j^0, j^{m+1} to denote "dummy" auditors, with respective distance functions: $\forall x, x' \in \mathcal{X}^2$, $d^{j^0}(x, x') = 0$, $d^{j^{m+1}}(x, x') = 1$. j^0 hence objects to any non-identical predictions made on any two individuals, while j^{m+1} never objects to any predictions.

Lemma 2.9. *Let $\mathcal{S}$ (fixing $f \in \mathcal{F}^{Mon}$) be a monotone individual fairness auditing scheme, and fix a panel of auditors $\bar{j} = (j^1, \dots, j^m)$. Then, for any pair $(x^l, x^r) \in \mathcal{X}^2$, there exist $i^* = i(f, \bar{j}, (x^l, x^r)) \in \{0\} \cup [m+1]$ such that*

[3] Apart from being a natural and desirable quality for auditing schemes, we will later see how monotonicity will also be important in the analysis of our algorithms (in particular, see Lemma 2.9 and Lemma C.1).

[4] Monotone aggregation schemes have also been studied in social choice theory (see, e.g. (Woodall, 1997; Ornstein & Norman, 2013)) in the context of voting rules.

$\forall \pi \in \Delta\mathcal{H}, \alpha \in \mathbb{R}^+$,

$$\mathcal{S}(\pi, (x^l, x^r), \alpha, f, \bar{j}) = j^{i^*}(\pi, (x^l, x^r), \alpha).$$

2.3. On the complexity of auditing

Remark 2.10. Monotone auditing schemes are far more expressive than simply considering majority-based schemes (Bechavod & Roth, 2023). As a simple example, consider an auditing scheme over five auditors $(j^1, \ldots, j^5)$, where an objection on a pair is reported if either j^1 objects or in case a majority of 4/5 objections is reached. It is straightforward to see that this aggregation scheme is in fact monotone, as adding objections to any objection profile $v = (v^1, \ldots, v^5) \in \{0,1\}^5$ by the auditors with respect to a pair (x^l, x^r) can only result in an unchanged decision, or in changing the decision to reporting a violation.

Remark 2.11. Note that for schemes where certain auditors with a veto right — these members *are never required to fully enunciate their objection matrix*, but rather just report a *single* pair where they deem a violation to exist or that there are no violations. In particular, employing a single auditor is a special case of a member with veto right, making the task of auditing much simpler. For general auditing schemes, however, (non-veto having) panel members are required to report an objection matrix, as otherwise, one might run into a case of Condorcet's paradox (Condorcet, 2014) — for example, when each auditor reports a different pair out of multiple objections, and while a pair on which an objection profile resulting in a violation is formed, it is never detected.

Remark 2.12. Varying the size of the sensitivity parameter $\alpha \in [0,1]$ corresponds to more stringent constraints (for smaller values of α), or less stringent ones (for larger values), hence offering a natural "lever" for the learner to explore different points on the resulting accuracy-fairness frontier.

3. Online Learning with Individual Fairness

Here, we formally define our problem setting. We begin by defining the two types of losses we wish to minimize: misclassification loss and unfairness loss.

Definition 3.1 (Misclassification loss)**.** We define the misclassification loss as, for all $\pi \in \Delta\mathcal{H}$, $\bar{x} \in \mathcal{X}^k$, $\bar{y} \in \{0,1\}^k$:

$$\text{Error}(\pi, \bar{x}, \bar{y}) := \mathop{\mathbb{E}}_{h\sim\pi}[\ell^{0-1}(h, \bar{x}, \bar{y})].$$

Where for all $h \in \mathcal{H}$, $\ell^{0-1}(h, \bar{x}, \bar{y}) := \sum_{i=1}^{k} \ell^{0-1}(h, (\bar{x}^i, \bar{y}^i))$, and $\forall i \in [k] : \ell^{0-1}(h, (\bar{x}^i, \bar{y}^i)) = \mathbb{1}[h(\bar{x}^i) \neq \bar{y}^i]$.[5]

In particular, the misclassification loss is linear in π. We define the unfairness loss, to reflect the existence of one or more fairness violations according to an auditing scheme.

Definition 3.2 (Unfairness loss)**.** We define the unfairness loss as, for all $\pi \in \Delta\mathcal{H}$, $\bar{x} \in \mathcal{X}^k$, $\mathcal{S} \in \bar{\mathcal{S}}$, $\alpha \in \mathbb{R}^+$,

$$\text{Unfair}(\pi, \bar{x}, \mathcal{S}, \alpha) := \begin{cases} 1 & \vec{\mathcal{S}}(\pi, \bar{x}, \alpha) = (\bar{x}^l, \bar{x}^r) \\ 0 & \vec{\mathcal{S}}(\pi, \bar{x}, \alpha) = \text{Null} \end{cases}.$$

There is, however, an issue with working directly with the unfairness loss: as we will see in Section 4, we will only have access to realizations $h \sim \pi$, rather than the actual probabilities. Taking the expectation in this case will not be helpful either, as it is easy to construct cases where $\text{Unfair}(\pi, \bar{x}, \mathcal{S}, \alpha) = 0$, yet $\mathbb{E}_{h\sim\pi}[\text{Unfair}(\pi, \bar{x}, \mathcal{S}, \alpha)] = 1$ (we refer the reader to Lemma 4.11 in (Bechavod & Roth, 2023)). We will hence rely on resampling $h \sim \pi$ multiple times to form $\tilde{\pi}$, an empirical approximation of π, and use it to elicit fairness violations from the auditing scheme. We hence next introduce an unfairness proxy loss:

Definition 3.3 (Unfairness proxy loss)**.** We define the unfairness proxy loss as, for all $\pi, \tilde{\pi} \in \Delta\mathcal{H}$, $\bar{x} \in \mathcal{X}^k$, $\mathcal{S} \in \bar{\mathcal{S}}$, $\alpha \in \mathbb{R}^+$, $\beta \in \mathbb{R}$,

$$\overline{\text{Unfair}}(\pi, \tilde{\pi}, \bar{x}, \mathcal{S}, \alpha, \beta) := \begin{cases} \left[\pi(\bar{x}^l) - \pi(\bar{x}^r)\right] - \\ \left[\tilde{\pi}(\bar{x}^l) - \tilde{\pi}(\bar{x}^r)\right] + \beta & \vec{\mathcal{S}}(\tilde{\pi}, \bar{x}, \alpha) = (\bar{x}^l, \bar{x}^r) \\ 0 & \vec{\mathcal{S}}(\tilde{\pi}, \bar{x}, \alpha) = Null \end{cases}.$$

Importantly, the role of $\pi, \tilde{\pi}$ will be very different; As we will see in Section 4, we will only have sampling access to π. Hence, we will have $\tilde{\pi}$ be an empirical approximation of π, and use it to elicit fairness violations from the auditing scheme. Note, additionally, that when fixing $\tilde{\pi}$, the unfairness proxy loss is *linear* with respect to π. Finally, the β parameter will be used to offset the result to a desired range.

In the following lemma, we argue that if $\tilde{\pi}$ is in fact a good enough approximation of π, the unfairness proxy loss provides a meaningful upper bound to the unfairness loss.

Lemma 3.4. *Let* $\pi, \tilde{\pi} \in \Delta\mathcal{H}$, $\bar{x} \in \mathcal{X}^k$, $\mathcal{S} \in \bar{\mathcal{S}}$, $\alpha \in (0,1]$, $\epsilon' \in (0, \alpha]$. *If* $\forall i \in [k] : \left|\pi(\bar{x}^i) - \tilde{\pi}(\bar{x}^i)\right| \leq \frac{\epsilon'}{4}$, *then* $\textit{Unfair}(\pi, \bar{x}, \mathcal{S}, \alpha) \leq \frac{2}{\epsilon'}\overline{\textit{Unfair}}(\pi, \tilde{\pi}, \bar{x}, \mathcal{S}, \alpha - \epsilon', \epsilon')$.

3.1. Online learning setting

Our setting is formally described in Algorithm 1, where we denote a Learner by $\mathbf{L}$, and an Adversary by $\mathbf{A}$.[6]

[5]For simplicity, we define our misclassification loss as the expectation (over $h \sim \pi$) of the 0-1 loss. However, one can consider different base loss functions as well.

[6]In the setting described in Algorithm 1, we assume that the number of incoming individuals on every round is constant — k. It is however possible to consider a more general scenario, where this number changes between rounds. In this more general case, our bounds will simply scale with $max_{t\in[T]}k_t$ instead of k.

Algorithm 1 Online Learning with Individual Fairness

Input: Number of rounds T, hypothesis class $\mathcal{H}$, violation size $\alpha \in (0, 1]$
for $t = 1, \dots, T$ **do**
 L deploys $\pi^t \in \Delta\mathcal{H}$;
 A selects $(\bar{x}^t, \bar{y}^t) \in \mathcal{X}^k \times \mathcal{Y}^k$;
 A selects auditing scheme $\mathcal{S}^t$ (fixing $\bar{j}^t, f^t$);
 L suffers misclassification loss $\text{Error}(\pi^t, \bar{x}^t, \bar{y}^t)$;
 L suffers unfairness loss $\text{Unfair}(\pi^t, \bar{x}^t, \mathcal{S}^t, \alpha)$;
 L observes $(\bar{x}^t, \bar{y}^t), \rho^t = \vec{\mathcal{S}}^t(\pi^t, \bar{x}^t, \alpha, f^t, \bar{j}^t)$;
end for

To build intuition, consider the following motivating example of loan approvals: a government-based financial institution wishes to predict incoming loan applications in a manner that is simultaneously accurate (highly predictive of future repayment), and fair (similar applicants receive similar assessments). To obtain fairness feedback, the institution periodically hires panels of auditors (financial experts, ethicists, etc.) who report assessments they deem unfair.

In the notation of Algorithm 1, π^t is a lending policy deployed at time t. For each applicant i of the k arriving loan applicants at round t, $\bar{x}^{t,i} \in \mathcal{X}$ are relevant features (income, repayment history, debt, etc.), and $\bar{y}^{t,i} \in \{0, 1\}$ indicates if the applicant will repay the loan if approved. The auditing scheme $\mathcal{S}^t$ aggregates the reports of a panel of auditors $\bar{j}^t = (j^{t,1}, \dots, j^{t,m_t})$ with respect to the predictions made by π^t on applicants $\bar{x}^t = (\bar{x}^{t,1}, \dots, \bar{x}^{t,k})$ according to aggregation function f^t, and reports back in case a violation was found. Finally, the deployed lending policy is measured by whether it predicted repayment accurately, and whether it treated similar applicants (in the eyes of the panel) similarly.

In what follows, we adopt the following notation, $\forall t \in [T]$:

$$\text{Error}^t(\pi) := \text{Error}(\pi, \bar{x}^t, \bar{y}^t),$$
$$\text{Unfair}^t_\alpha(\pi) := \text{Unfair}(\pi, \bar{x}^t, \mathcal{S}^t, \alpha),$$
$$\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha,\beta}(\pi) := \overline{\text{Unfair}}(\pi, \tilde{\pi}^t, \bar{x}^t, \mathcal{S}^t, \alpha, \beta).$$

3.2. Learning objectives

Next, we formally define our learning objectives. Ideally, a learner could wish to refrain completely from having any fairness violations, by restricting, on every round, the set of active policies to only ones that obey the active fairness constraints. There are k^2 such constraints every round — corresponding to all pairs of individuals in $\bar{x}^t$. However, these constraints are *implicit* — they are decided by the (internal) preferences of the auditors in $\bar{j}^t$, along with the aggregation function f^t. Making these constraints *explicit* would require strictly stronger access to the auditors than assumed in our framework, querying for *exact distances* between all pairs in $\bar{x}^t$. In our framework, however, auditors are only required to report fairness *violations*, and are not even required to specify the size of those violations.[7]

We will hence adopt a slightly more relaxed objective, where we allow the learner to violate the constraints, but only for a *sub-linear* number of times. This is the approach also taken, in the context of learning with individual fairness, by Gillen et al. (2018); Bechavod et al. (2020); Bechavod & Roth (2023), and more generally in the literature on online learning with long-term constraints (e.g. Mahdavi et al. (2012); Jenatton et al. (2016); Sun et al. (2017); Castiglioni et al. (2022)). We next define the class of policies we wish to compete with — policies that refrain from violations of slightly smaller sensitivity of $\alpha - \epsilon$, for $\epsilon \in (0, \alpha]$.[8]

Definition 3.5. [Fair-In-Hindsight Policies] Denote the realized sequence of individuals, labels, auditors, and aggregation functions by the adversary until round $t \in [T]$ by

$$\Psi^t = \left((\bar{x}^1, \bar{y}^1, \bar{j}^1, f^1), \dots, (\bar{x}^t, \bar{y}^t, \bar{j}^t, f^t)\right).$$

We define the comparator class of $(\alpha - \epsilon)$-fair policies as[9,10]

$$\Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t) := \{\pi \in \Delta\mathcal{H} : \forall t \in [T],\ \text{Unfair}^t_{\alpha-\epsilon}(\pi) = 0\}.$$

Finally, let $\pi^* \in \text{argmin}_{\pi \in \Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t)} \sum_{t=1}^T \text{Error}^t(\pi)$.

Finally, we formally define our learning objective. First, we formally define the regret of an online algorithm.

Definition 3.6 (Regret)**.** In the setting of Algorithm 1, we define the (external) regret of an online algorithm $\mathcal{A}$ against a comparator class $U \subseteq \Delta\mathcal{H}$ as:

$$\text{Regret}_T(U) := \sum_{t=1}^T \text{Error}(\pi^t, \bar{x}^t, \bar{y}^t) - \min_{\pi \in U} \sum_{t=1}^T \text{Error}(\pi, \bar{x}^t, \bar{y}^t).$$

For a randomized algorithm, we will consider the expected regret.

[7]Additionally, as also stated in Remark 2.11, in many notable cases, auditors will not even be required the enunciate all of their objections, but rather a single one.

[8]We adopt a slightly relaxed baseline in terms of violation sensitivity, as the adversary can always report violations of magnitude *arbitrarily* close to α.

[9]Interestingly, since in our setting the learner does not receive full information regarding the constraints, but rather very limited, "bandit"-like information on violations made by policies that were actually deployed, it is possible that the learner will not know (even in hindsight) which policies are included in the set of fair-in-hindsight policies. Nevertheless, as we will see, it will be possible to provide strong guarantees when competing against it.

[10]As we rely on the sensitivity of human auditors in reporting violations, it is reasonable to think about α, ϵ as small constants.

Equipped with Definition 3.6, we define our learning objective:

Learning objective: In the setting of Algorithm 1, obtain:

1. **Simultaneous no-regret**:
   (a) **Accuracy**: $\text{Regret}_T(\Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t)) = o(T)$.
   (b) **Fairness**: $\sum_{t=1}^T \text{Unfair}^t_{\alpha-\epsilon}(\pi^t) = o(T)$.
2. **Oracle-efficiency**: Polynomial runtime, given access to an (offline) optimization oracle.[11]

3.3. Achieving simultaneous no-regret guarantees

Obtaining each of the accuracy, fairness objectives *in isolation* is a relatively easy task — for accuracy, one can run an oracle-efficient no regret algorithm such as Context-FTPL (Syrgkanis et al., 2016) only using the misclassification loss. For fairness, one can simply predict using any constant predictor, which would ensure fairness violations never occur, regardless of the auditing scheme. However, when attempting to obtain both objectives *simultaneously*, the task becomes much more complicated. In particular, one cannot simply combine, in online fashion, the per-round outputs of said algorithms when run in isolation. The reason is that the feedback of the auditing process only pertains to the policies that have actually been deployed.[12]

Another (naive) approach is to define a joint loss function of misclassification and (linearized, proxy) unfairness $L^t(\pi) = \text{Error}^t(\pi) + \text{UnfairProx}^t(\pi)$, and run a no-regret algorithm with respect to the sequence of losses $L^1, \ldots, L^T$, in hopes of bounding each of the objectives individually. Unfortunately, this may fail. The reason is that regret may actually be *negative*[13]: $\sum_{t=1}^T \text{Error}^t(\pi^t) - \sum_{t=1}^T \text{Error}^t(\pi^*) < 0$. Hence, even if $\sum_{t=1}^T L^t(\pi^t) - \sum_{t=1}^T L^t(\pi^*) = o(T)$, the algorithm may have still violated fairness on *every round*.[14]

Bechavod et al. (2020) suggested a reductions approach to the problem, dynamically encoding fairness violations as "fake" datapoints in the incoming stream, ultimately reducing the problem to a standard (unconstrained) classification problem. They then suggested "inflating" the number of these fake datapoints, so as to, on one hand, penalize unfairness more severely, and on the other hand, not to increase the artificial dimension of the problem too sharply (since the resulting bounds deteriorate as k grows artificially larger). They then give an oracle-efficient algorithm that guaranteeing a bound of $O(T^{\frac{7}{9}})$ for each of regret and number of fairness violations. In order to circumvent the fact that in their algorithm, the learner only has sampling access to the deployed policy π^t, they suggest approximating this policy using T calls to an offline optimization oracle on every round. We next show how both the convergence rates and and oracle complexity can be improved.

3.4. Faster rates with direct Lagrangian loss

In order to obtain faster rates, we will work *directly* with the following Lagrangian formulation, combining the misclassification loss and our introduced unfairness proxy loss.

Definition 3.7 (Lagrangian loss)**.** Let $\alpha \in (0, 1]$, $\beta \in \mathbb{R}$, and fix any $\tilde{\pi} \in \Delta\mathcal{H}$. We define the $(\alpha, \beta, \tilde{\pi})$-Lagrangian loss at round $t \in [T]$ as, for all $\pi \in \Delta\mathcal{H}$, $\lambda \in \mathbb{R}^+$,

$$L^t(\pi, \lambda) := \frac{1}{k} \cdot \text{Error}^t(\pi) + \lambda \cdot \overline{\text{Unfair}}^t_{\tilde{\pi},\alpha,\beta}(\pi).$$

By doing so, we take the perspective of a saddle-point problem for our learning objective (see, e.g. Agarwal et al. (2018); Freund & Schapire (1997)) — where the primal player (who sets π) attempts to minimize loss, and the dual player (who sets λ) attempts to maximize constraint violations. Importantly, the Lagrangian loss is *linear* in $\pi \in \Delta\mathcal{H}$. This will be critical in competing against the best fair policy in $\Delta\mathcal{H}$, rather than against the more restricted class $\mathcal{H}$. In line with the approach of Bechavod et al. (2020), we will have to carefully select the value of λ, as setting λ too low would risk potentially ignoring the fairness constraints (as illustrated in the example in the second paragraph of Section 3.3), while setting λ too high would lead to worse bounds.

4. Algorithm

Equipped with the Lagrangian loss function, we remember that another central part of our learning objective is to

[11]The concept of oracle-efficiency aims to show that the online problem is not computationally harder than an offline version of the problem. Hence, when the learner has access to an optimization oracle for the offline problem (in our case, a batch ERM oracle for $\mathcal{H}$), we will be interested in algorithms that run in polynomial time, where each call to this oracle is counted as $\mathcal{O}(1)$. Algorithms such as Multiplicative Weights (Littlestone & Warmuth, 1994; Vovk, 1990; Cesa-Bianchi et al., 1997; Freund & Schapire, 1997), on the other hand, have exponential runtime and space complexity dependence on $\log|\mathcal{H}|$, as they *explicitly* maintain and update on every round a vector of probabilities over $\mathcal{H}$.

[12]For example, suppose a policy π^t was reported by $\mathcal{S}^t$ to induce a violation on individuals $(\bar{x}^{t,l}, \bar{x}^{t,r})$, when predicting, say, $\pi^t(\bar{x}^{t,l}) = 0.8$, $\pi^t(\bar{x}^{t,r}) = 0.4$. The learner would not know if $\mathcal{S}^t$ would have still reported a violation on $(\bar{x}^{t,l}, \bar{x}^{t,r})$ had he deployed a *different* policy, $\tilde{\pi}^t$, for which $0 < \tilde{\pi}^t(\bar{x}^{t,l}) - \tilde{\pi}^t(\bar{x}^{t,r}) < 0.4$.

[13]This is the case, since the algorithm has the liberty of deploying a different policy $\pi^t \in \mathcal{H}$ on every round, while competing with a *fixed* policy $\pi^* \in \Delta\mathcal{H}$.

[14]In general, having negative regret is highly desirable — it means that the algorithm performed even better than the baseline. However, in our particular case, it may actually do us a disservice — it can be used to "compensate" for fairness violations, potentially resulting in ignoring the fairness objective altogether.

provide an algorithm that is *oracle-efficient*. Our approach will be to carefully reduce our multi-objective problem to a single objective problem, using our Lagrangian formulation (Definition 3.7)). To update π^t, we will then want use Context-FTPL (Syrgkanis et al., 2016) on our generated a sequence of Lagrangian loss functions $L^1, \ldots, L^T$.

One particular difficulty is due to the fact that Contex-FTPL does not maintain π^t *explicitly*, but rather relies on access to an (offline) optimization oracle to *sample*, on each round, a single classifier $h^t \sim \pi^t$ from its *implicit* policy π^t.[15] In our setting, however, access to the exact π^t is critical, as it is used to query the auditors for fairness violations, and forming the sequence of losses $L^1, \ldots, L^T$. To circumvent this, our approach will be to distinguish between two tasks: eliciting the fairness constraints, and evaluating the error and unfairness losses. Ideally, one would like to perform both tasks using the same policy — the deployed policy π^t. Since, however, in our algorithm the learner will only have access to classifiers sampled from π^t, we will perform each task using a *different* policy. Namely, we will first form an accurate enough approximation $\tilde{\pi}^t$ of π^t, and use it to elicit the objections of the auditors. We will then use this feedback to form our Lagrangian loss L^t (as in Definition 3.7). We will then feed the Lagrangian loss to Context-FTPL, and prove accuracy, fairness guarantees for the true (implicit) policy π^t deployed by Context-FTPL.

Importantly, one must be careful when eliciting the constraints using the *approximate* policy $\tilde{\pi}^t$, as it could generate violations that would not exist if the auditing scheme was queried using π^t, or overlook other violations that should be generated. To address this, we suggest querying the auditors using $\tilde{\pi}^t$ for slightly more sensitive fairness violations, of size $\alpha - \frac{\epsilon}{2}$. We then argue that since it is sufficient for the learner to generate an approximation of π^t that is only accurate on $\bar{x}^t$ (rather than on the entire space $\mathcal{X}$), using $\tilde{\mathcal{O}}(\frac{1}{\epsilon^2})$ calls to Context-FTPL's optimization oracle would suffice to generate this approximation. This approach will allow us to upper bound a *counterfactual* quantity — the number of fairness violations that would have been reported had we used the implicit policy π^t to query the auditors. In order to run Context-FTPL (Syrgkanis et al., 2016), we assume access to a small separating set for $\mathcal{H}$, and access to an (offline) optimization oracle. The optimization oracle assumption is equivalent to access to a batch ERM oracle for $\mathcal{H}$. We next describe the small separating set assumption.

[15]Follow-The-Perturbed-Leader (FTPL)-style algorithms rely on access to an offline optimization (in our case, a batch ERM) oracle, which is invoked every round on the set of samples observed until that point, augmented by a collection of generated "fake" noisy samples. The noise distribution in this process implicitly defines, in turn, a distribution over the experts returned by the oracle. Hence calling the optimization oracle can equivalently be viewed as sampling an expert from this distribution.

Our construction is then formally described in Algorithm 2.

Definition 4.1 (Separating set)**.** We say $Q \subseteq \mathcal{X}$ is a separating set for a class $\mathcal{H} : \mathcal{X} \to \{0,1\}$, if for any two distinct hypotheses $h, h' \in \mathcal{H}$, there exists $x \in Q$ s.t. $h(x) \neq h'(x)$.

Remark 4.2. Classes for which small separating sets are known include conjunctions, disjunctions, parities, decision lists, discretized linear classifiers. Please see more elaborate discussions in Syrgkanis et al. (2016) and Neel et al. (2019).

Algorithm 2 Reduction to Context-FTPL for Online Learning with Individual Fairness

Input: Number of rounds T, hypothesis class $\mathcal{H}$, violation size $\alpha \in (0,1]$, sensitivity $\epsilon \in (0,\alpha]$, separating set $Q \subseteq \mathcal{X}$, parameters R, ω

L initializes Context-FTPL using Q, ω, history $\xi^1 = \emptyset$;

for $t = 1, \ldots, T$ **do**

  L deploys $\pi^t \in \Delta\mathcal{H}$ (implicitly by Context-FTPL(ξ^t));

  A selects $(\bar{x}^t, \bar{y}^t) \in \mathcal{X}^k \times \mathcal{Y}^k$;

  A selects panel $\bar{j}^t \in \mathcal{J}^{m_t}$, aggr. function $f^t \in \mathcal{F}$;

  for $r = 1, \ldots, R$ **do**

    L draws h^{t_r} using Context-FTPL(ξ^t);

  end for

  L sets $\tilde{\pi}^t = \mathbb{U}(h^{t_1}, \ldots, h^{t_R})$;

  L queries $\rho^t = \vec{\mathcal{S}}^t(\tilde{\pi}^t, \bar{x}^t, \alpha - \frac{\epsilon}{2}, \bar{j}^t, f^t)$;

  L updates $\xi^{t+1} = \{(L^{\tau}_{\tilde{\pi}^{\tau}, \alpha - \frac{\epsilon}{2}, \frac{\epsilon}{2}}(\cdot, \lambda^{\tau}), \bar{x}^{\tau}, \bar{y}^{\tau})\}_{\tau=1}^{t}$;

end for

Finally, we proceed to our main theorem. For the following statements, one can fix any $\alpha \in (0,1]$, $\epsilon \in (0,\alpha]$. We assume the algorithms are given access to a separating set $Q \subseteq \mathcal{X}$ for $\mathcal{H}$, of size s.

Theorem 4.3. *Algorithm 2 obtains, for any (possibly adversarial) sequence of individuals $(\bar{x}^t)_{t=1}^T$, labels $(\bar{y}^t)_{t=1}^T$, auditors $(\bar{j}^t)_{t=1}^T$, and monotone aggregation functions $(f^t)_{t=1}^T$, with probability $1-\delta$, simultaneously:*

(1) Accuracy :

$$\textit{Regret}_T(\Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t)) \leq \mathcal{O}\left(s^{\frac{3}{4}} k^{\frac{7}{4}} T^{\frac{3}{4}} \log^{\frac{1}{2}} |\mathcal{H}|\right).$$

(2) Fairness :

$$\sum_{t=1}^{T} \textit{Unfair}^t_{\alpha}(\pi^t) \leq \mathcal{O}\left(\frac{1}{\epsilon} s^{\frac{3}{4}} k^{\frac{3}{4}} T^{\frac{3}{4}} \log^{\frac{1}{2}} |\mathcal{H}|\right).$$

While only requiring $\tilde{\mathcal{O}}\left(\epsilon^{-2}\right)$ calls to a batch ERM optimization oracle every round.

Corollary 4.4. *In particular, our bounds uniformly improve on the formerly known upper bound of $O(T^{\frac{7}{9}})$ in Bechavod et al. (2020), while also reducing the per-round oracle complexity from T to $\tilde{\mathcal{O}}\left(\epsilon^{-2}\right)$.*

Remark 4.5. Having all functions $(f^t)_{t=1}^T$ be monotone will be critical in our analysis. In particular, Lemma C.1 in Appendix C is closely dependent on it, showing that $\Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t) \subseteq \{\pi : \forall t \in [T], \overline{\text{Unfair}}^t_{\bar{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi) \leq 0\}$.

5. Partial Information

In this section, we focus on the setting where the learner only observes one-sided label feedback, for individuals who have received a positive prediction. Note that such feedback structure is *extremely prevalent* in domains where fairness is a concern — a lender only observes repayment by applicants that have actually been approved for a loan to begin with, a university can only track the academic performance for candidates who have been admitted, etc. The key challenge in this setting is that the learner may not even observe *its own loss*. Note that this is *different* from a bandit setting, since feedback is available for the entire class $\mathcal{H}$ when a positive prediction is made, while no feedback (even for the deployed policy) is available for a negative prediction. In Appendix D, we formally define the setting (Algorithm 4), and present an oracle-efficient algorithm based on a reduction to Context-Semi-Bandit-FTPL (Syrgkanis et al., 2016) (Algorithm 4). We next present its guarantees.

Theorem 5.1. *Algorithm 4 obtains, in the one-sided label feedback setting, for any (possibly adversarial) sequence of individuals $(\bar{x}^t)_{t=1}^T$, labels $(\bar{y}^t)_{t=1}^T$, auditors $(\bar{j}^t)_{t=1}^T$, and monotone aggregation functions $(f^t)_{t=1}^T$, with probability $1-\delta$, simultaneously:*

(1) Accuracy :

$$Regret_T(\Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t)) \leq \mathcal{O}\left(s^{\frac{3}{4}} k^{\frac{11}{4}} T^{\frac{5}{6}} \log^{\frac{1}{2}} |\mathcal{H}|\right).$$

(2) Fairness :

$$\sum_{t=1}^{T} Unfair^t_\alpha(\pi^t) \leq \mathcal{O}\left(\frac{1}{\epsilon} s^{\frac{3}{4}} k^{\frac{7}{4}} T^{\frac{5}{6}} \log^{\frac{1}{2}} |\mathcal{H}|\right).$$

While only requiring $\tilde{\mathcal{O}}(\epsilon^{-2} + k^2 T^{\frac{1}{3}})$ calls to a batch ERM optimization oracle every round.

Corollary 5.2. *In particular, our bounds uniformly improve on the formerly known upper bound of $O(T^{\frac{41}{45}})$ in Bechavod & Roth (2023), while also reducing the per-round oracle complexity from T to $\tilde{\mathcal{O}}(\epsilon^{-2} + k^2 T^{\frac{1}{3}})$.*

6. Conclusion and Future Directions

One limitation of our approach is that it is guaranteed to run efficiently only on classes for which one can pre-compute a small separating set. However, this limitation is not unique to our setting, and is prevalent more generally in the context of adversarial online learning. Another limitation is that we can only compete with a slightly relaxed baseline (in terms of violation size). It would be is interesting to think about ways to extend our approach to obtain regret to the class where no such relaxation is required (and one can even potentially select $\alpha = 0$). Finally, proving non-trivial lower bounds in our setting is also a very interesting problem. To gain some intuition — a "trivial" policy (constant predictor) can (naively) never violate fairness, but induces linear regret. A non-constant policy, however, must risk violating fairness, as both the fairness metric and labels aren't initially known. One might then be inclined to ask, for algorithms that obtain a non-trivial regret bound $o(T)$, what level of fairness constraint violation is unavoidable?

Acknowledgements

The author wishes to thank Aaron Roth, Michael Kearns, and Georgy Noarov for useful discussions at an early stage of this work, Guy Rothblum for an observation leading to the insight in footnote 9, and Rabanus Derr, Georgy Noarov, and Mirah Shi for providing valueable feedback on the manuscript. YB is supported in part by the Israeli Council for Higher Education Postdoctoral Fellowship.

Impact Statement

The central objectives of this work are to propose and explore a new framework and algorithms that: (1) offer meaningful fairness guarantees to *individuals*, and (2) are tailored for the *specific structure* and characteristics that are often prevalent in problem domains where fairness is a concern. The first part is done by adopting and extending the notion of individual fairness by Dwork et al. (2012). We elaborate on the second part next:

Fairness as a dynamically-evolving concept Naturally, fairness is a *dynamic* concept. Society's perceptions regarding "what is fair?" are constantly evolving, with the introduction of new norms, developments, and technological advancements. In our framework, we model fairness as such — a dynamically changing concept — as we allow auditors (and aggregation functions) to change over time, potentially reflecting different and evolving perceptions.

Accounting for limitations in auditing As feedback from human auditors can be noisy or imperfect, we also consider and model the case where auditors report their preferences in a manner that *does not* obey metric form, as it is important to minimize such made assumptions. Additionally, our framework aims to make the task of auditing easier for humans in the loop, as it does not require auditors to enunciate a fairness metric, or even report exact distances between individuals. Finally, by potentially incorporating multiple auditors in each auditing scheme, we refrain from placing too much power in the hands of single auditors. In particular, our framework is fully capable of handling diverse panels of potentially disagreeing auditors with conflicting opinions.

Moving beyond statistical data generation assumptions When considering domains where decisions are made with respect to human individuals, one must take into account the specific interplay between algorithms and individuals in different contexts. One aspect we highlight in that regard, is the documented tendency of individuals to act adaptively and strategically in order to obtain more favorable outcomes, e.g. by modifying their features or deciding whether to postpone their application based on their perceived chances of being accepted. Such effects go beyond classical statistical assumptions, and in our framework we make an effort to account for them.

Avoiding feedback loops and visibility biases Many problem domains among the ones motivating our framework manifest a one-sided feedback structure, where data is available only for *accepted* (or admitted, hired, etc.) individuals. Such structure demands that we, as algorithm designers, specifically account for it, to avoid visibility biases and pernicious feedback loops. Note that such biases can also find their way to the manner in which data is collected — for example, when we study a batch setting and rely on previously collected datasets, in many cases such datasets reflect a *filtered* view of reality, as they contain individuals classified positively by the previously deployed policies (e.g. a dataset of past loan applicants), which may be inaccurate or even discriminatory. In line with this discussion, our framework focuses on online settings while refraining from making statistical data generation assumptions, while specifically accounting for partial information.

References

Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., and Wallach, H. M. A reductions approach to fair classification. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pp. 60–69, 2018.

Bechavod, Y. and Roth, A. Individually fair learning with one-sided feedback. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 1954–1977. PMLR, 2023.

Bechavod, Y., Jung, C., and Wu, Z. S. Metric-free individual fairness in online learning. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Cao, X. and Liu, K. J. R. Online convex optimization with time-varying constraints and bandit feedback. *IEEE Trans. Autom. Control.*, 64(7):2665–2680, 2019.

Castiglioni, M., Celli, A., Marchesi, A., Romano, G., and Gatti, N. A unifying framework for online optimization with long-term constraints. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

Cesa-Bianchi, N., Freund, Y., Haussler, D., Helmbold, D. P., Schapire, R. E., and Warmuth, M. K. How to use expert advice. *J. ACM*, 44(3):427–485, May 1997.

Chen, T. and Giannakis, G. B. Bandit convex optimization for scalable and dynamic iot management. *IEEE Internet Things J.*, 6(1):1276–1286, 2019.

Chen, T., Ling, Q., and Giannakis, G. B. An online convex optimization approach to proactive network resource allocation. *IEEE Trans. Signal Process.*, 65(24):6350–6364, 2017.

Condorcet, N. d. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Cambridge Library Collection - Mathematics. Cambridge University Press, 2014.

Coston, A., Rambachan, A., and Chouldechova, A. Characterizing fairness over the set of good models under selective labels. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 2144–2155. PMLR, 2021.

De-Arteaga, M., Dubrawski, A., and Chouldechova, A. Learning under selective labels in the presence of expert consistency. *CoRR*, abs/1807.00905, 2018.

Dee, T. S., Dobbie, W., Jacob, B. A., and Rockoff, J. The causes and consequences of test score manipulation: Evidence from the new york regents examinations. *American Economic Journal: Applied Economics*, 11(3):382–423, 07 2019.

Dranove, D., Kessler, D., McClellan, M., and Satterthwaite, M. Is more information better? the effects of "report cards" on health care providers. *Journal of Political Economy*, 111(3):555–588, 2003.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. S. Fairness through awareness. In Goldwasser, S. (ed.), *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, pp. 214–226. ACM, 2012.

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., and Venkatasubramanian, S. Runaway feedback loops in predictive policing. In Friedler, S. A. and Wilson, C. (eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pp. 160–171. PMLR, 23–24 Feb 2018a.

Ensign, D., Sorelle, F., Scott, N., Carlos, S., and Suresh, V. Decision making with limited feedback. In Janoos, F., Mohri, M., and Sridharan, K. (eds.), *Proceedings of Algorithmic Learning Theory*, volume 83 of *Proceedings of Machine Learning Research*, pp. 359–367. PMLR, 07–09 Apr 2018b.

Freund, Y. and Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.

Gillen, S., Jung, C., Kearns, M. J., and Roth, A. Online learning with an unknown fairness metric. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, pp. 2605–2614, 2018.

Gonzalez-Lira, A. and Mobarak, A. M. Slippery Fish: Enforcing Regulation under Subversive Adaptation. IZA Discussion Papers 12179, Institute of Labor Economics (IZA), 02 2019.

Greenstone, M., He, G., Jia, R., and Liu, T. Can technology solve the principal-agent problem? evidence from china's war on air pollution. *SSRN Electronic Journal*, 01 2020.

Gupta, S. and Kamble, V. Individual fairness in hindsight. In *Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019, Phoenix, AZ, USA, June 24-28, 2019*, pp. 805–806, 2019.

Helmbold, D. P., Littlestone, N., and Long, P. M. Apple tasting. *Inf. Comput.*, 161(2):85–139, 2000.

Ilvento, C. Metric learning for individual fairness. In Roth, A. (ed.), *1st Symposium on Foundations of Responsible Computing, FORC 2020, June 1-3, 2020, Harvard University, Cambridge, MA, USA (virtual conference)*, volume 156 of *LIPIcs*, pp. 2:1–2:11. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

Jenatton, R., Huang, J., and Archambeau, C. Adaptive algorithms for online convex optimization with long-term constraints. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 402–411, New York, New York, USA, 20–22 Jun 2016. PMLR.

Joseph, M., Kearns, M. J., Morgenstern, J. H., and Roth, A. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 325–333, 2016.

Joseph, M., Kearns, M. J., Morgenstern, J., Neel, S., and Roth, A. Meritocratic fairness for infinite and contextual bandits. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2018, New Orleans, LA, USA, February 02-03, 2018*, pp. 158–163, 2018.

Jung, C., Kearns, M., Neel, S., Roth, A., Stapleton, L., and Wu, Z. S. An algorithmic framework for fairness elicitation. In Ligett, K. and Gupta, S. (eds.), *2nd Symposium on Foundations of Responsible Computing, FORC 2021, June 9-11, 2021, Virtual Conference*, volume 192 of *LIPIcs*, pp. 2:1–2:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

Kim, M. P., Reingold, O., and Rothblum, G. N. Fairness through computationally-bounded awareness. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada*, pp. 4847–4857, 2018.

Lahoti, P., Gummadi, K. P., and Weikum, G. Operationalizing individual fairness with pairwise fair representations. *Proc. VLDB Endow.*, 13(4):506–518, 2019.

Lakkaraju, H. and Rudin, C. Learning cost-effective and interpretable treatment regimes. In Singh, A. and Zhu, X. J. (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA*, volume 54 of *Proceedings of Machine Learning Research*, pp. 166–175. PMLR, 2017.

Lakkaraju, H., Kleinberg, J. M., Leskovec, J., Ludwig, J., and Mullainathan, S. The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*, pp. 275–284. ACM, 2017.

Liakopoulos, N., Destounis, A., Paschos, G. S., Spyropoulos, T., and Mertikopoulos, P. Cautious regret minimization: Online optimization with long-term budget constraints. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3944–3952. PMLR, 2019.

Littlestone, N. and Warmuth, M. K. The weighted majority algorithm. *Inf. Comput.*, 108(2):212–261, 1994.

Mahdavi, M., Jin, R., and Yang, T. Trading regret for efficiency: Online convex optimization with long term constraints. *Journal of Machine Learning Research*, 13 (81):2503–2528, 2012.

Mannor, S., Tsitsiklis, J. N., and Yu, J. Y. Online learning with sample path constraints. *J. Mach. Learn. Res.*, 10: 569–590, 2009.

Mukherjee, D., Yurochkin, M., Banerjee, M., and Sun, Y. Two simple ways to learn individual fairness metrics from data. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7097–7107. PMLR, 2020.

Neel, S., Roth, A., and Wu, Z. S. How to use heuristics for differential privacy. In Zuckerman, D. (ed.), *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pp. 72–93. IEEE Computer Society, 2019.

Ornstein, J. and Norman, R. Frequency of monotonicity failure under instant runoff voting: Estimates based on a spatial model of elections. *Public Choice*, 161, 10 2013. doi: 10.1007/s11127-013-0118-2.

Rothblum, G. N. and Yona, G. Probably approximately metric-fair learning. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pp. 5666–5674, 2018.

Sun, W., Dey, D., and Kapoor, A. Safety-aware algorithms for adversarial contextual bandit. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3280–3288. PMLR, 06–11 Aug 2017.

Syrgkanis, V., Krishnamurthy, A., and Schapire, R. E. Efficient algorithms for adversarial contextual learning. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pp. 2159–2168, 2016.

Vargo, A., Zhang, F., Yurochkin, M., and Sun, Y. Individually fair gradient boosting. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.

Vovk, V. G. Aggregating strategies. In Fulk, M. A. and Case, J. (eds.), *Proceedings of the Third Annual Workshop on Computational Learning Theory, COLT 1990, University of Rochester, Rochester, NY, USA, August 6-8, 1990*, pp. 371–386. Morgan Kaufmann, 1990.

Wei, X., Yu, H., and Neely, M. J. Online primal-dual mirror descent under stochastic constraints. *Proc. ACM Meas. Anal. Comput. Syst.*, 4(2), jun 2020.

Woodall, D. R. Monotonicity of single-seat preferential election rules. *Discrete Applied Mathematics*, 77(1):81–98, 1997.

Yi, X., Li, X., Xie, L., and Johansson, K. H. Distributed online convex optimization with time-varying coupled inequality constraints. *IEEE Trans. Signal Process.*, 68: 731–746, 2020.

Yu, H. and Neely, M. J. A low complexity algorithm with o($\sqrt{T}$) regret and o(1) constraint violations for online convex optimization with long term constraints. *Journal of Machine Learning Research*, 21(1):1–24, 2020.

Yu, H., Neely, M. J., and Wei, X. Online convex optimization with stochastic constraints. In Guyon, I., von Luxburg, U., Bengio, S., Wallach, H. M., Fergus, R., Vishwanathan, S. V. N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 1428–1438, 2017.

Yuan, J. and Lamperski, A. Online convex optimization for cumulative constraints. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Yurochkin, M. and Sun, Y. Sensei: Sensitive set invariance for enforcing individual fairness. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.

Yurochkin, M., Bower, A., and Sun, Y. Training individually fair ML models with sensitive subspace robustness. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

Zeng, J., Ustun, B., and Rudin, C. Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society Series A*, 180(3):689–722, June 2017.

Zhang, C. Y., Cen, S. H., and Shah, D. Matrix estimation for individual fairness. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 40871–40887. PMLR, 2023.

Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In Fawcett, T. and Mishra, N. (eds.), *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, pp. 928–936. AAAI Press, 2003.

A. Extended Related Work

In the context of individual fairness, Joseph et al. (2016; 2018), Gupta & Kamble (2019) study a time-dependent variant of individual fairness they term *fairness in hindsight*. Lahoti et al. (2019) study methods of generating individually fair representations.Yurochkin et al. (2020) suggest learning predictors that are invariant to certain perturbations of sensitive attributes. Mukherjee et al. (2020) suggest ways to learn fairness metrics from data. Yurochkin & Sun (2021) Vargo et al. (2021); Zhang et al. (2023)

B. Proofs from Section 2

Proof of Lemma 2.9. Fix $f \in \mathcal{F}^{Mon}$, a panel $\bar{j} = (j^1, \ldots, j^m) \in \mathcal{J}^m$, and a pair $(x^l, x^r) \in \mathcal{X}^2$. Consider an ordering of the panel by defining a set of indices $\{i_1, \ldots, i_m\} = [m]$ such that

$$d^{i_1}(x^l, x^r) \leq \cdots \leq d^{i_m}(x^l, x^r).$$

Denote the set of objection profiles with respect to predictions made on (x^l, x^r) which result in an aggregated decision of a violation (coordinates are according to the auditor's ordering defined above) by

$$Z = Z^{\bar{j},(x^l,x^r)} = \{z \in \{0,1\}^m : \bar{f}(z) = 1\}.$$

Remark B.1. Note that as the ordering of auditors (and hence coordinates) depends on the selection of (x^l, x^r), even a fixed aggregation function f and a fixed panel $\bar{j}$ would generate different sets $Z = Z^{\bar{j},(x^l,x^r)}$ for different selections of (x^l, x^r).

Next, consider the following index $i^* \in \{0\} \cup [m+1]$:

$$i^*(f, \bar{j}, (x^l, x^r)) = \begin{cases} m+1 & Z = \emptyset \\ 0 & (0, \ldots, 0) \in Z \\ \min\limits_{z \in Z} \max\limits_{q : z^{i_q} = 1} q & \text{otherwise} \end{cases}.$$

Since f is monotone, and given the ordering of auditors we defined, we know that the following is the set of all possible objection profiles by $j^{i_1}, \ldots, j^{i_m}$ on (x^l, x^r) which result in an aggregate decision of reporting a violation:

$$Z = \{(\overbrace{1, \ldots, 1}^{c \text{ times}}, 0, \ldots, 0) : i^* \leq c \leq m\}.$$

We hence know, $\forall \pi \in \Delta\mathcal{H}, \alpha \in \mathbb{R}^+$:

$$\mathcal{S}(\pi, (x^l, x^r), \alpha, f, \bar{j}) = j^{i^*}(\pi, (x^l, x^r), \alpha).$$

As desired.

□

C. Proofs from Section 4

We begin by stating and proving two lemmas, which, along with Lemma 3.4, will be useful for proving Theorem 4.3.

Lemma C.1. *For $\alpha \in (0, 1]$, $\epsilon \in (0, \alpha]$, and $\epsilon' = \frac{\epsilon}{2}$, it holds that $\Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t) \subseteq \{\pi : \forall t \in [T], \overline{Unfair}^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi) \leq 0\}$.*

Proof. Fix any $t \in [T]$, and let $\pi \in \Delta\mathcal{H}^{fair}_{\alpha-\epsilon}(\Psi^t)$. If $\vec{\mathcal{S}}^t(\tilde{\pi}^t, \bar{x}^t, \alpha - \epsilon') = (\bar{x}^{t,l}, \bar{x}^{t,r})$, since $\mathcal{S}^t$ is a monotone auditing scheme, using Lemma 2.9, there exists $i^* = i^*(f^t, \bar{j}^t, (\bar{x}^{t,l}, \bar{x}^{t,r})) \in \{0\} \cup [m+1]$ such that

$$\forall \pi' \in \Delta\mathcal{H}, \alpha' \in (0, 1] : \vec{\mathcal{S}}^t(\pi', (\bar{x}^{t,l}, \bar{x}^{t,r}), \alpha') = \vec{j}^{t,i^*}(\pi', (\bar{x}^{t,l}, \bar{x}^{t,r}), \alpha').$$

Hence, using Definition 3.3,

$$\begin{aligned} \overline{\text{Unfair}}^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi) &= \left[\pi(\bar{x}^{t,l}) - \pi(\bar{x}^{t,r})\right] - \left[\tilde{\pi}^t(\bar{x}^{t,l}) - \tilde{\pi}^t(\bar{x}^{t,r})\right] + \frac{\epsilon}{2} \\ &\leq \left[d^{t,i^*}(\bar{x}^{t,l}, \bar{x}^{t,r}) + \alpha - \epsilon\right] - \left[d^{t,i^*}(\bar{x}^{t,l}, \bar{x}^{t,r}) + \alpha - \frac{\epsilon}{2}\right] + \frac{\epsilon}{2} \\ &= 0. \end{aligned}$$

Where the inequality stems by combining Definition 3.5 and Lemma 2.9.

Otherwise, $\vec{\mathcal{S}}^t(\tilde{\pi}^t, \bar{x}^t, \alpha - \epsilon') = Null$, and $\overline{\text{Unfair}}^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi) = 0$.

This concludes the proof. ☐

Lemma C.2. *With probability* $1 - \delta$ *(over the draw of* $\{h^{t_r}\}_{t=1,r=1}^{t=T,r=R}$*),*

$$\forall t \in [T], i \in [k] : \left|\pi^t(\bar{x}^{t,i}) - \tilde{\pi}^t(\bar{x}^{t,i})\right| \leq \sqrt{\frac{\log\left(\frac{2kT}{\delta}\right)}{2R}}.$$

In particular, setting $R = \frac{64\log\left(\frac{2kT}{\delta}\right)}{\epsilon^2}$ *results in the right hand side being* $\frac{\epsilon}{8}$.

Proof. Fix $t \in [T], i \in [k]$. Using an additive Chernoff bound,

$$\Pr\left[\left|\pi^t(\bar{x}^{t,i}) - \tilde{\pi}^t(\bar{x}^{t,i})\right| \leq \sqrt{\frac{\log\left(\frac{2kT}{\delta}\right)}{2R}}\right] \leq \frac{\delta}{kT}.$$

The statement then follows by taking a union bound over all $t \in [T], i \in [k]$. ☐

Proof of Lemma 3.4. Assume the condition in the statement of the lemma holds. Using the condition in conjunction with the triangle inequality, we know that:

$$\vec{\mathcal{S}}(\tilde{\pi}, \bar{x}, \alpha - \epsilon') = \text{Null} \implies \vec{\mathcal{S}}(\pi, \bar{x}, \alpha) = \text{Null}.$$

In such case,

$$\text{Unfair}(\pi, \bar{x}, \mathcal{S}, \alpha) = \overline{\text{Unfair}}(\pi, \tilde{\pi}, \bar{x}, \mathcal{S}, \alpha - \epsilon', \epsilon') = 0,$$

And hence

$$\text{Unfair}(\pi, \bar{x}, \mathcal{S}, \alpha) \leq \frac{2}{\epsilon'}\overline{\text{Unfair}}(\pi, \tilde{\pi}, \bar{x}, \mathcal{S}, \alpha - \epsilon', \epsilon').$$

Otherwise, $\vec{\mathcal{S}}(\tilde{\pi}, \bar{x}, \alpha - \epsilon') = (\bar{x}^l, \bar{x}^r)$, and we know

$$\begin{aligned}
\text{Unfair}(\pi, \bar{x}, \mathcal{S}, \alpha) &\leq 1 \\
&= \frac{2}{\epsilon'}\left[\frac{-\epsilon'}{2} + \epsilon'\right] \\
&\leq \frac{2}{\epsilon'}\left[\left[\pi(\bar{x}^l) - \pi(\bar{x}^r)\right] - \left[\tilde{\pi}(\bar{x}^l) - \tilde{\pi}(\bar{x}^r)\right] + \epsilon'\right] \\
&= \frac{2}{\epsilon'}\overline{\text{Unfair}}(\pi, \tilde{\pi}, \bar{x}, \mathcal{S}, \alpha - \epsilon', \epsilon').
\end{aligned}$$

Where the first inequality stems from Definition 3.2, and the second inequality follows from the condition in the statement of this lemma, along with the triangle inequality. The claim follows. ☐

Proof of Theorem 4.3. Set $R = \frac{64\log\left(\frac{2kT}{\delta}\right)}{\epsilon^2}$, $\omega = s^{-\frac{1}{4}}k^{-\frac{1}{4}}T^{-\frac{3}{4}}\log^{-\frac{1}{2}}\mathcal{H}$, $\lambda^t = T^{\frac{1}{4}}$, and denote $\epsilon' = \frac{\epsilon}{2}$.

Using Theorem 2 from Syrgkanis et al. (2016), along with the fact that the Lagrangian loss (Definition 3.7) is linear in the first argument, we know that, for any $\pi \in \Delta\mathcal{H}$,

$$\sum_{t=1}^{T} L^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi^t, \lambda^t) - \sum_{t=1}^{T} L^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi, \lambda^t) \leq 4\omega k s \sum_{t=1}^{T} \mathbb{E}\left[\|L^t\|_*^2\right] + \frac{10}{\omega}s^{\frac{1}{2}}k^{\frac{1}{2}}\log|\mathcal{H}|, \tag{1}$$

where $\|L^t\|_* = \max_{h\in\mathcal{H}}|L^t(h, \lambda^t)|$.

Equivalently, using Definition 3.7,

$$\sum_{t=1}^{T}\frac{1}{k}\cdot\text{Error}^t(\pi^t)-\sum_{t=1}^{T}\frac{1}{k}\cdot\text{Error}^t(\pi)+\sum_{t=1}^{T}\lambda^t\cdot\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^t)-\sum_{t=1}^{T}\lambda^t\cdot\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi)$$
$$\leq 4\omega ks\sum_{t=1}^{T}\mathbb{E}\left[\|L^t\|_*^2\right]+\frac{10}{\omega}s^{\frac{1}{2}}k^{\frac{1}{2}}\log|\mathcal{H}|.$$

To upper bound the regret, we set $\pi=\pi^*$. Using Lemma C.1, we know, for all $t\in[T]$, that $\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^*)\leq 0$. Using Lemma C.2 along with the triangle inequality, we know that with probability $1-\delta$, simultaneously for all $t\in[T]$, $\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^t)\geq\frac{\epsilon}{4}$. Hence, $\sum_{t=1}^{T}\lambda^t\cdot\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^t)-\sum_{t=1}^{T}\lambda^t\cdot\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^*)\geq 0$, and we get

$$\sum_{t=1}^{T}\text{Error}^t(\pi^t)-\sum_{t=1}^{T}\text{Error}^t(\pi^*)\leq 4\omega k^2 s\sum_{t=1}^{T}\mathbb{E}\left[\|L^t\|_*^2\right]+\frac{10}{\omega}s^{\frac{1}{2}}k^{\frac{3}{2}}\log|\mathcal{H}|. \tag{2}$$

To upper bound the number of fairness violations, note that

$$\sum_{t=1}^{T}\frac{1}{k}\cdot\text{Error}^t(\pi^t)-\sum_{t=1}^{T}\frac{1}{k}\cdot\text{Error}^t(\pi)\geq-\frac{T}{k}.$$

And hence,

$$\sum_{t=1}^{T}\lambda^t\cdot\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^t)-\sum_{t=1}^{T}\lambda^t\cdot\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi)\leq 4\omega ks\sum_{t=1}^{T}\mathbb{E}\left[\|L^t\|_*^2\right]+\frac{10}{\omega}s^{\frac{1}{2}}k^{\frac{1}{2}}\log|\mathcal{H}|+\frac{T}{k}.$$

We set $\pi=\pi^*$. Using Lemma C.1, we know, for all $t\in[T]$, that $\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^*)\leq 0$.

Hence we can bound,

$$\sum_{t=1}^{T}\lambda^t\cdot\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^t)\leq 4\omega ks\sum_{t=1}^{T}\mathbb{E}\left[\|L^t\|_*^2\right]+\frac{10}{\omega}s^{\frac{1}{2}}k^{\frac{1}{2}}\log|\mathcal{H}|+\frac{T}{k}. \tag{3}$$

Next, since $\omega=s^{-\frac{1}{4}}k^{-\frac{1}{4}}T^{-\frac{3}{4}}\log^{-\frac{1}{2}}\mathcal{H}$, $\lambda^t=T^{\frac{1}{4}}$, and noting that $\|L^t\|_*\leq 4T^{\frac{1}{4}}$, we can bound the regret using Equation (2):

$$\sum_{t=1}^{T}\text{Error}^t(\pi^t)-\sum_{t=1}^{T}\text{Error}^t(\pi^*)\leq\mathcal{O}\left(s^{\frac{3}{4}}k^{\frac{7}{4}}T^{\frac{3}{4}}\log^{\frac{1}{2}}|\mathcal{H}|\right).$$

And the number of fairness violations, using Equation (3):

$$\sum_{t=1}^{T}\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}\leq\mathcal{O}\left(s^{\frac{3}{4}}k^{\frac{3}{4}}T^{\frac{3}{4}}\log^{\frac{1}{2}}|\mathcal{H}|\right).$$

We conclude, using Lemma 3.4,

$$\sum_{t=1}^{T}\text{Unfair}^t_{\alpha}(\pi^t)\leq\frac{2}{\epsilon'}\sum_{t=1}^{T}\overline{\text{Unfair}}^t_{\tilde{\pi}^t,\alpha-\epsilon',\epsilon'}(\pi^t)\leq\mathcal{O}\left(\frac{1}{\epsilon}s^{\frac{3}{4}}k^{\frac{3}{4}}T^{\frac{3}{4}}\log^{\frac{1}{2}}|\mathcal{H}|\right).$$

Which concludes the proof. □

D. Partial Information

In this section, we focus on the setting where the learner only observes one-sided label feedback, for individuals who have received a positive prediction.[16] Note that many domains where fairness is a concern naturally exhibit such feedback structure — a lender only observes repayment by applicants that have actually been approved for a loan to begin with, a university can only track the academic performance for candidates who have been admitted, etc. (also see, e.g. Lakkaraju et al. (2017); Lakkaraju & Rudin (2017); Zeng et al. (2017); De-Arteaga et al. (2018); Ensign et al. (2018a;b); Coston et al. (2021)). The key challenge in this setting is that the learner may not even observe *its own loss*. Note that this is *different* from a bandit setting, since feedback is available for the entire class $\mathcal{H}$ when a positive prediction is made, while no feedback (even for the deployed policy) is available for a negative prediction. The setting is formally described in Algorithm 3.

Algorithm 3 Online Learning with Individual Fairness and Partial Information

Input: Number of rounds T, hypothesis class $\mathcal{H}$, violation size $\alpha \in (0, 1]$
for $t = 1, \dots, T$ **do**
  L deploys $\pi^t \in \Delta\mathcal{H}$;
  A selects $(\bar{x}^t, \bar{y}^t) \in \mathcal{X}^k \times \mathcal{Y}^k$, **L** only observes $\bar{x}^t$;
  A selects auditing scheme $\mathcal{S}^t$ (fixing $\bar{j}^t, f^t$);
  L draws $h^t \sim \pi^t$, predicts $\hat{y}^{t,i} = h^t(\bar{x}^{t,i})$, $\forall i \in [k]$;
  L suffers misclassification loss Error$(h^t, \bar{x}^t, \bar{y}^t)$ (not necessarily observed by **L**);
  L suffers unfairness loss Unfair$(\pi^t, \bar{x}^t, \mathcal{S}^t, \alpha)$;
  L observes $\bar{x}^t, \bar{y}^{t,i}$ iff $\hat{y}^{t,i} = 1, \rho^t = \vec{\mathcal{S}}^t(\pi^t, \bar{x}^t, \alpha, f^t, \bar{j}^t)$;
end for

Next, we present and analyze an oracle-efficient algorithm using a reduction to Context-Semi-Bandit-FTPL (Syrgkanis et al., 2016) for the setting of individual fairness and one-sided label feedback (Algorithm 4).

Algorithm 4 Reduction to Context-Semi-Bandit-FTPL for Online Learning with Individual Fairness and Partial Information

Input: Number of rounds T, hypothesis class $\mathcal{H}$, violation size $\alpha \in (0, 1]$, sensitivity $\epsilon \in (0, \alpha]$, separating set $Q \subseteq \mathcal{X}$, parameters R, ω, M
L initializes Context-FTPL using Q, ω, M, history $\xi^1 = \emptyset$;
for $t = 1, \dots, T$ **do**
  L deploys $\pi^t \in \Delta\mathcal{H}$ (implicitly using Context-Semi-Bandit-FTPL(ξ^t));
  A selects $(\bar{x}^t, \bar{y}^t) \in \mathcal{X}^k \times \mathcal{Y}^k$, **L** only observes $\bar{x}^t$;
  A selects panel $\bar{j}^t \in \mathcal{J}^{m_t}$, aggr. function $f^t \in \mathcal{F}$;
  for $r = 1, \dots, R$ **do**
    L draws h^{t_r} using Context-Semi-Bandit-FTPL(ξ^t); {//without performing loss estimation}
  end for
  L sets $\tilde{\pi}^t = \mathbb{U}(h^{t_1}, \dots, h^{t_R})$;
  L queries $\rho^t = \vec{\mathcal{S}}^t(\tilde{\pi}^t, \bar{x}^t, \alpha - \frac{\epsilon}{2}, \bar{j}^t, f^t)$;
  L draws h^t using Context-Semi-Bandit-FTPL(ξ^t); {//with loss estimation}
  L predicts $\hat{y}^{t,i} = h^t(\bar{x}^{t,i})$, $\forall i \in [k]$, observes $\bar{\bar{y}}^t = \{y^{t,i} : \hat{y}^{t,i} = 1\}$;
  L updates history $\xi^{t+1} = \{\bar{L}^{\tau}_{\tilde{\pi}^\tau, \alpha - \frac{\epsilon}{2}, \frac{\epsilon}{2}}(\cdot, \lambda^\tau), \bar{x}^\tau, \bar{\bar{y}}^\tau\}_{\tau=1}^t$;
end for

Finally, we prove the guarantees obtained by Algorithm 4, as stated in Theorem 5.1.

[16]The one-sided label feedback setting was first introduced as the "Apple tasting" problem by Helmbold et al. (2000).

Proof of Theorem 5.1. Set $R = \frac{64 \log\left(\frac{2kT}{\delta}\right)}{\epsilon^2}$, $\omega = s^{-\frac{1}{4}} k^{-\frac{5}{4}} T^{-\frac{5}{6}} \log^{-\frac{1}{2}} \mathcal{H}$, $M = T^{\frac{1}{3}}$, $\lambda^t = T^{\frac{1}{6}}$, and denote $\epsilon' = \frac{\epsilon}{2}$.

Using Theorem 3 from Syrgkanis et al. (2016) for the semi-bandit setting, along with the fact that the Lagrangian loss (Definition 3.7) is linear in the first argument, we know that (assuming $\|L^t\|_* \leq B$), for any $\pi \in \Delta\mathcal{H}$,

$$\sum_{t=1}^{T} L^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi^t, \lambda^t) - \sum_{t=1}^{T} L^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi, \lambda^t) \leq 4B^2 \omega s k^3 MT + \frac{BkT}{eM} + \frac{10}{\omega} s^{\frac{1}{2}} k^{\frac{1}{2}} \log |\mathcal{H}|.$$

Using the same derivation as the proof of Theorem 4.3, we obtain, with probability $1 - \delta$, the following bounds:

For regret,

$$\sum_{t=1}^{T} \text{Error}^t(\pi^t) - \sum_{t=1}^{T} \text{Error}^t(\pi^*) \leq 4B^2 \omega s k^4 MT + \frac{Bk^2 T}{eM} + \frac{10}{\omega} s^{\frac{1}{2}} k^{\frac{3}{2}} \log |\mathcal{H}|.$$

For fairness violations,

$$\sum_{t=1}^{T} \lambda^t \cdot \overline{\text{Unfair}}^t_{\tilde{\pi}^t, \alpha-\epsilon', \epsilon'}(\pi^t) \leq 4B^2 \omega s k^3 MT + \frac{BkT}{eM} + \frac{10}{\omega} s^{\frac{1}{2}} k^{\frac{3}{2}} \log |\mathcal{H}| + \frac{T}{k}.$$

Next, since $\omega = s^{-\frac{1}{4}} k^{-\frac{5}{4}} T^{-\frac{5}{6}} \log^{-\frac{1}{2}} \mathcal{H}$, $M = T^{\frac{1}{3}}$, $\lambda^t = T^{\frac{1}{6}}$, and noting that $\|L^t\|_* \leq 4T^{\frac{1}{6}}$, for regret,

$$\sum_{t=1}^{T} \text{Error}^t(\pi^t) - \sum_{t=1}^{T} \text{Error}^t(\pi^*) \leq \mathcal{O}\left(s^{\frac{3}{4}} k^{\frac{11}{4}} T^{\frac{5}{6}} \log^{\frac{1}{2}} |\mathcal{H}| \right).$$

For the number of fairness violations,

$$\sum_{t=1}^{T} \text{Unfair}^t_{\alpha}(\pi^t) \leq \mathcal{O}\left(\frac{1}{\epsilon} s^{\frac{3}{4}} k^{\frac{7}{4}} T^{\frac{5}{6}} \log^{\frac{1}{2}} |\mathcal{H}| \right).$$

In closing, note that our selection of M implies, according to Theorem 3 from Syrgkanis et al. (2016) and our reduction, that the per-round number of calls to the optimization oracle is $64\epsilon^{-2} \log\left(\frac{2kT}{\delta}\right) + 16e^{-1} k^2 T^{\frac{1}{3}}$. This concludes the proof. $\square$