Metric-Free Individual Fairness in Online Learning

Yahav Bechavod* Christopher Jung† Zhiwei Steven Wu‡

February 14, 2020

Abstract

We study an online learning problem subject to the constraint of individual fairness, which requires that similar individuals are treated similarly. Unlike prior work on individual fairness, we do not assume the similarity measure among individuals is known, nor do we assume that such measure takes a certain parametric form. Instead, we leverage the existence of an *auditor* who detects fairness violations without enunciating the quantitative measure. In each round, the auditor examines the learner's decisions and attempts to identify a pair of individuals that are treated unfairly by the learner. We provide a general reduction framework that reduces online classification in our model to standard online classification, which allows us to leverage existing online learning algorithms to achieve sub-linear regret and number of fairness violations. Surprisingly, in the stochastic setting where the data are drawn independently from a distribution, we are also able to establish PAC-style fairness and accuracy generalization guarantees (Yona and Rothblum [2018]), despite only having access to a very restricted form of fairness feedback. Our fairness generalization bound qualitatively matches the uniform convergence bound of Yona and Rothblum [2018], while also providing a meaningful accuracy generalization guarantee. Our results resolve an open question by Gillen et al. [2018] by showing that online learning under an unknown individual fairness constraint is possible even without assuming a strong parametric form of the underlying similarity measure.

*School of Computer Science and Engineering, The Hebrew University. Email: yahav.bechavod@cs.huji.ac.il.
†Departmenet of Computer and Information Science, University of Pennsylvania. Email: chrjung@seas.upenn.edu.
‡Department of Computer Science and Engineering, University of Minnesota. Email: zsw@umn.edu.

1 Introduction

As machine learning increasingly permeates many critical aspects of society, including education, healthcare, criminal justice, and lending, there is by now a vast literature that studies how to make machine learning algorithms fair (see, e.g., Chouldechova and Roth [2018]; Podesta et al. [2014]; Corbett-Davies and Goel [2018]). Most of the work in this literature tackles the problem by taking the *statistical group fairness* approach that first fixes a small collection of high-level groups defined by protected attributes (e.g., race or gender) and then asks for approximate parity of some statistic of the predictor, such as positive classification rate or false positive rate, across these groups (see, e.g., Hardt et al. [2016], Chouldechova [2017], Kleinberg et al. [2017], Agarwal et al. [2018]). While notions of group fairness are easy to operationalize, this approach is aggregate in nature and does not provide fairness guarantees for finer subgroups or individuals [Dwork et al., 2012, Hébert-Johnson et al., 2018, Kearns et al., 2018].

In contrast, the *individual fairness* approach aims to address this limitation by asking for explicit fairness criteria at an individual level. In particular, the compelling notion of individual fairness proposed in the seminal work of Dwork et al. [2012] requires that similar people are treated similarly. The original formulation of individual fairness assumes that the algorithm designer has access to a task-specific fairness metric that captures how similar two individuals are in the context of the specific classification task at hand. In practice, however, such a fairness metric is rarely specified, and the lack of metrics has been a major obstacle for the wide adoption of individual fairness. There has been recent work on learning the fairness metric based on different forms of human feedback. For example, Ilvento [2019] provides an algorithm for learning the metric by presenting human arbiters with queries concerning the distance between individuals, and Gillen et al. [2018] provide an online learning algorithm that can eventually learn a Mahalanobis metric based on identified fairness violations. While these results are encouraging, they are still bound by several limitations. In particular, it might be difficult for humans to enunciate a precise quantitative similarity measure between individuals. Moreover, their similarity measure across individuals may not be consistent with any metric (e.g., it may not satisfy the triangle inequality) and is unlikely to be given by a simple parametric function (e.g., the Mahalanobis metric function).

To tackle these issues, this paper studies *metric-free* online learning algorithms for individual fairness that rely on a weaker form of interactive human feedback and minimal assumptions on the similarity measure across individuals. Similar to the prior work of Gillen et al. [2018], we do not assume a pre-specified metric, but instead assume access to an *auditor*, who observes the learner's decisions over a group of individuals that show up in each round and attempts to identify a fairness violation—a pair of individuals in the group that should have been treated more similarly by the learner. Since the auditor only needs to identify such unfairly treated pairs, there is no need for them to enunciate a quantitative measure – to specify the distance between the identified pairs. Moreover, we do not impose any parametric assumption on the underlying similarity measure, nor do we assume that it is actually a metric since we do not require that similarity measure to satisfy the triangle inequality. Under this model, we provide a general reduction framework that can take any online classification algorithm (without fairness constraint) as a black-box and obtain a learning algorithm that can simultaneously minimize cumulative classification loss and the number of fairness violations. Our results in particular remove many strong assumptions in Gillen et al. [2018], including their parametric assumptions on linear rewards and Mahalanobis distances, and answer several questions left open in their work. We now provide an overview of the results.

1.1 Overview of Model and Results

We study an online classification problem: over rounds $t = 1, \dots, T$, a learner observes a small set of k individuals with their feature vectors $(x_\tau^t)_{\tau=1}^k$ in space $\mathcal{X}$. The learner tries to predict the label $y_k^t \in \{0, 1\}$ of each individual with a "soft" predictor π^t that predicts $\pi^t(x_\tau^t) \in [0, 1]$ on each x_τ^t and incurs classification loss $|\pi^t(x_\tau^t) - y_\tau^t|$. Then an auditor will investigate if the learner has violated the individual fairness constraint on any pair of individuals within this round, that is, if there exists $(\tau_1, \tau_2) \in [k]^2$ such that $|\pi^t(x_{\tau_1}^t) - \pi^t(x_{\tau_2}^t)| > d(x_{\tau_1}^t, x_{\tau_2}^t) + \alpha$, where $d\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is an unknown distance function and α denotes the auditor's tolerance. If this violation has occurred on any number of pairs, the auditor will identify one of such pairs and incur a fairness loss of 1; otherwise, the fairness loss is 0. Then the learner will update the predictive policy based on the observed labels and the received fairness feedback. Under this model, our results include:

- **A General Reduction From Metric-Free Online Classification to Standard Online Classification**

  Our reduction-based algorithm can take any no-regret online (batch) classification learner as a black-box and achieve sub-linear cumulative fairness loss and sub-linear regret on mis-classification loss compared to the most accurate policy that is fair on every round. In particular, our framework can leverage the generic exponential weights method [Freund and Schapire, 1997, Cesa-Bianchi et al., 1997, Arora et al., 2012] and also oracle-efficient methods, including variants of Follow-the-Perturbed-Leader (FTPL) (e.g., Syrgkanis et al. [2016], Suggala and Netrapalli [2019]), that further reduces online learning to standard supervised learning or optimization problems. We instantiate our framework using two online learning algorithms (exponential weights and CONTEXT-FTPL), both of which obtain a $\tilde{O}(\sqrt{T})$ on misclassification regret and cumulative fairness loss.

- **Fairness and Accuracy Generalization Guarantee of the Learned Policy**

  While our algorithmic results hold under adversarial arrivals of the individuals, in the stochastic arrivals setting we show that the uniform average policy over time is probably approximate correct and fair (PACF) [Yona and Rothblum, 2018]–that is, the policy is approximately fair on almost all random pairs drawn from the distribution and nearly matches the accuracy gurantee of the best fair policy. In particular, we show that the average policy π^{avg} with high probability satisfies

  $$\Pr_{x,x'}[|\pi^{avg}(x) - \pi^{avg}(x')| > \alpha + 1/T^{1/4}] \leq O(1/T^{1/4}),$$

  which qualitatively achieves similar PACF uniform convergence sample complexity as Yona and Rothblum [2018].[1] However, we establish our generalization guarantee through fundamentally different techniques. While their work assumes a fully specified metric and i.i.d. data, the learner in our setting can only access the similarity measure through an auditor's limited fairness violations feedback. The main challenge we need to overcome is that the fairness feedback is inherently adaptive–that is, the auditor only provides feedback for the sequence of deployed policies, which are updated adaptively over rounds. In comparison, a fully known metric allows the learner to evaluate the fairness guarantee of all policies simultaneously. As a result, we cannot rely on their uniform convergence result to bound the fairness generalization error, but instead we leverage a probabilistic argument that relates the learner's regret to the distributional fairness guarantee.

[1] Yona and Rothblum [2018] show that if a policy π is α-fair on all pairs in a i.i.d. dataset of size m, then π satisfies $\Pr_{x,x'}[|\pi(x) - \pi(x')| > \alpha + \epsilon] \leq \epsilon$, as long as $m \geq \tilde{\Omega}(1/\epsilon^4)$.

1.2 Related Work

Solving open problems in Gillen et al. [2018]. The most related work to ours is Gillen et al. [2018], which studies the linear contextual bandit problem subject to individual fairness with an unknown Mahalanobis metric. Similar to our work, they also assume an auditor who can identify fairness violations in each round and provide an online learning algorithm with sublinear regret and a bounded number of fairness violations. Our results resolve two main questions left open in their work. First, we assume a weaker auditor who only identifies a single fairness violation (as opposed to all of the fairness violations in their setting). Second, we remove the strong parametric assumption on the Mahalanobis metric and work with a broad class of similarity functions that need not be metric.

Starting with Joseph et al. [2016], there is a different line of work that studies online learning for individual fairness, but subject to a different notion called meritocratic fairness [Jabbari et al., 2017, Joseph et al., 2018, Kannan et al., 2017]. These results present algorithms that are "fair" within each round but again rely on strong realizability assumptions–their fairness guarantee depends on the assumption that the outcome variable of each individual is given by a linear function. Gupta and Kamble [2019] also studies online learning subject to individual fairness but with a known metric. They formulate a one-sided fairness constraint across time, called fairness in hindsight, and provide an algorithm with regret $O(T^{M/(M+1)})$ for some distribution-dependent constant M.

Our work is related to several others that aim to enforce individual fairness without a known metric. Ilvento [2019] studies the problem of metric learning by asking human arbiters distance queries. Unlike Ilvento [2019], our algorithm does not explicitly learn the underlying similarity measure and does not require asking auditors numeric queries. The PAC-style fairness generalization bound in our work falls under the framework of *probably approximately metric-fairness* due to Yona and Rothblum [2018]. However, their work assumes a pre-specified fairness metric and i.i.d. data from the distribution, while we establish our generalization through a sequence of adaptive fairness violations feedback over time. Kim et al. [2018] study a group-fairness relaxation of individual fairness, which requires that similar subpopulations are treated similarly. They do not assume a pre-specified metric for their offline learning problem, but they do assume a metric oracle that returns numeric distance values on random pairs of individuals. Jung et al. [2019] study an offline learning problem with subjective individual fairness, in which the algorithm tries to elicit subjective fairness feedback from human judges by asking them questions of the form "should this pair of individuals be treated similarly or not?" Their fairness generalization takes a different form, which involves taking averages over both the individuals and human judges. We aim to provide a fairness generalization guarantee that holds for almost all individuals from the population.

2 Model and Preliminaries

We define the instance space to be $\mathcal{X}$ and its label space to be $\mathcal{Y}$. Throughout this paper, we will restrict our attention to binary labels, that is $\mathcal{Y} = \{0, 1\}$. We write $\mathcal{H} : \mathcal{X} \to \mathcal{Y}$ to denote the hypothesis class and assume that $\mathcal{H}$ contains a constant hypothesis – i.e. there exists h such that $h(x) = 0$ for all $x \in \mathcal{X}$. Also, we allow for convex combination of hypotheses for the purpose of randomizing the prediction and denote the simplex of hypotheses by $\Delta\mathcal{H}$; we call a randomized hypothesis a *policy*. Sometimes, we assume the existence of an underlying (but unknown) distribution $\mathcal{D}$ over $(\mathcal{X}, \mathcal{Y})$. For each prediction $\hat{y} \in \mathcal{Y}$ and its true label $y \in \mathcal{Y}$, there is an associated misclassification loss, $\ell(\hat{y}, y) = \mathbb{1}(\hat{y} \neq y)$. For simplicity, we overload the notation and write

$$\ell(\pi(x), y) = (1 - \pi(x)) \cdot y + \pi(x) \cdot (1 - y) = \mathop{\mathbb{E}}_{h \sim \pi}[\ell(h(x), y)].$$

2.1 Individual Fairness and Auditor

We want our deployed policy π to behave fairly in some manner, and we use the individual fairness definition from Dwork et al. [2012] that asserts that "similar individuals should be treated similarly." We assume that there is some distance function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ over the instance space $\mathcal{X}$ which captures the distance between individuals in $\mathcal{X}$, although d doesn't have to satisfy the triangle inequality. The only requirement on d is that it is always non-negative and symmetric $d(x, x') = d(x', x)$.

Definition 2.1 ((α, β)-fairness)**.** *Assume $\alpha, \beta > 0$. A policy $\pi \in \Delta\mathcal{H}$ is said to be α-fair on pair (x, x'), if*

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

We say policy π's α-fairness violation on pair (x, x') is

$$v_\alpha(\pi, (x, x')) = \max(0, |\pi(x) - \pi(x')| - d(x, x') - \alpha).$$

A policy is π is said to be (α, β)-fair on distribution $\mathcal{D}$, if

$$\Pr_{(x,x') \sim \mathcal{D}|_\mathcal{X} \times \mathcal{D}|_\mathcal{X}}[|\pi(x) - \pi(x')| > d(x, x') + \alpha] \leq \beta.$$

A policy π is said to be α-fair on set $S \subseteq \mathcal{X}$, if for all $(x, x') \in S^2$, it is α-fair.

Although individual fairness is intuitively sound, individual fairness notion requires the knowledge of the distance function d which is often hard to specify. Therefore, we rely on an auditor $\mathcal{J}$ that can detect instances of α-unfairness.

Definition 2.2 (Auditor $\mathcal{J}$)**.** *An auditor $\mathcal{J}_\alpha$ which can have its own internal state takes in a* reference set *$S \subseteq \mathcal{X}$ and a policy π. Then, it outputs ρ which is either $null$ or a pair of indices from the provided reference set to denote that there is some positive α-fairness violation for that pair. For some $S = (x_1, \ldots, x_n)$,*

$$\mathcal{J}_\alpha(S, \pi) = \begin{cases} \rho = (\rho_1, \rho_2) & \textit{if} \quad \exists \rho_1, \rho_2 \in [n]. \pi(x_{\rho_1}) - \pi(x_{\rho_2}) - d(x_{\rho_1}, x_{\rho_2}) - \alpha > 0 \\ null & \textit{otherwise} \end{cases}$$

If there exists multiple pairs with some α-violation, the auditor can choose one arbitrarily.

Remark 2.3. *Our assumptions on the auditor are much more relaxed than those of Gillen et al. [2018], which require that the auditor outputs whether the policy is 0-fair (i.e. with no slack) on all pairs S^2 exactly. Furthermore, the auditor in Gillen et al. [2018] can only handle Mahalanobis distances. Furthermore, in our setting, because of the internal state of the auditor, the auditor does not have to be a fixed function but rather can be adaptively changing in each round. Finally, we never rely on the fact the distance function d stays the same throughout rounds, meaning all our results extend to the case where the distance function governing the fairness constraints is changing every round.*

Algorithm 1: Online Fair Batch Classification FAIR-BATCH

for $t = 1, \ldots, T$ **do**
  Learner deploys π^t
  Environment chooses $(\bar{x}^t, \bar{y}^t)$
  Environment chooses the pair ρ^t
  $z^t = (\bar{x}^t, \bar{y}^t) \times \rho^t$
  Learner incurs batch misclassfication loss $\mathrm{Err}(\pi^t, z^t)$
  Learner incurs fairness loss $\mathrm{Unfair}(\pi^t, z^t)$
end

Algorithm 2: Online Batch Classification BATCH

for $t = 1, \ldots, T$ **do**
  Learner deploys π^t
  Environment chooses $z^t = (\bar{x}^t, \bar{y}^t)$
  Learner incurs misclassification loss $\mathrm{Err}(\pi^t, z^t)$
end

Figure 1: Comparison between Online Fair Batch Classification and Online Batch Classification: each setting may be summarized by the interaction between the learner and the environment: $(\Delta\mathcal{H} \times \mathcal{Z}_{\text{FAIR-BATCH}})^T$ and $(\Delta\mathcal{H} \times \mathcal{Z}_{\text{BATCH}})^T$ where $\mathcal{Z}_{\text{FAIR-BATCH}} = \mathcal{X}^k \times \mathcal{Y}^k \times ([k]^2 \cup \{null\})$ and $\mathcal{Z}_{\text{BATCH}} = \mathcal{X}^k \times \mathcal{Y}^k$.

2.2 Online Batch Classification

We now describe our online batch classification setting. In each round $t = 1, \ldots, T$, the learner deploys some model $\pi^t \in \Delta\mathcal{H}$. Upon seeing the deployed policy π^t, the environment chooses a batch of k individuals, $(x^t_\tau, y^t_\tau)_{\tau=1}^k$ and possibly, a pair of individuals from that round on which π^t will be responsible for any α-fairness violation. For simplicity, we write $\bar{x}^t = (x^t_\tau)_{\tau=1}^k$ and $\bar{y}^t = (y^t_\tau)_{\tau=1}^k$. The strategy $z^t_{\text{FAIR-BATCH}} \in \mathcal{Z}_{\text{FAIR-BATCH}}$ that the environment chooses can be described by

$$z^t_{\text{FAIR-BATCH}} = (\bar{x}^t, \bar{y}^t) \times \rho^t,$$

where $\rho^t \in [k]^2 \cup \{null\}$. Often, we will omit the subscript and simply write z^t. If $\rho^t = (\rho^t_1, \rho^t_2)$, then π^t will be responsible for the α-fairness violation on the pair $(x^t_{\rho^t_1}, x^t_{\rho^t_2})$. There are two types of losses that we are interested in: misclassification and fairness loss.

Definition 2.4 (Misclassification Loss)**.** *The (batch) misclassification loss Err*[2] *is*

$$\mathit{Err}(\pi, z^t) = \sum_{\tau=1}^{k} \ell(\pi(x^t_\tau), y^t_\tau).$$

Definition 2.5 (Fairness Loss)**.** *The α-fairness loss Unfair$_\alpha$ is*

$$\mathit{Unfair}_\alpha(\pi, z^t) = \begin{cases} \mathbb{1}\left(\pi(x^t_{\rho^t_1}) - \pi(x^t_{\rho^t_2}) - d(x^t_{\rho^t_1}, x^t_{\rho^t_2}) - \alpha > 0\right) & \text{if } \rho^t = (\rho^t_1, \rho^t_2) \\ 0 & \text{otherwise} \end{cases}$$

We want the total misclassification and fairness loss over T rounds to be as small as any $\pi^* \in Q$ for some competitor set Q, which we describe now. As said above, each round's reference set, a set of pairs for which the deployed policy will possibly be responsible in terms of α-fairness, will be defined in terms of

[2] We will overload the notation for this loss; regardless of what $\mathcal{Z}$ is, we'll assume $\mathrm{Err}(\pi, z^t)$ is well-defined as long as z^t includes $(\bar{x}^t, \bar{y}^t)$.

the instances that arrive within that round $\bar{x}^t$. The baseline Q_α that we compete against will be all policies that are α-fair on $\bar{x}^t$ for all $t \in [T]$:

$$Q_\alpha = \{\pi \in \Delta\mathcal{H} : \pi \text{ is } \alpha\text{-fair on } \bar{x}^t \text{ for all } t \in [T]\}$$

Note that because $\mathcal{H}$ contains a constant hypothesis which must be 0-fair on all instances, Q_α cannot be empty. The difference in total loss between our algorithm and a fixed π^* is called 'regret', which we formally define below.

Definition 2.6 (Algorithm $\mathcal{A}$)**.** *An algorithm $\mathcal{A} : (\Delta\mathcal{H} \times \mathcal{Z})^* \to \Delta\mathcal{H}$ takes in its past history $(\pi^\tau, z^\tau)_{\tau=1}^{t-1}$ and deploys a policy $\pi^t \in \Delta\mathcal{H}$ at every round $t \in [T]$.*

Definition 2.7 (Regret). *For some $Q \subseteq \Delta\mathcal{H}$, the regret of algorithm $\mathcal{A}$ with respect to some loss $L : \Delta\mathcal{H} \times \mathcal{Z} \to \mathbb{R}$ is denoted as $\mathbf{Regret}^L(\mathcal{A}, Q, T)$, if for any $(z_t)_{t=1}^T$,*

$$\sum_{t=1}^{T} L\left(\pi^t, z^t\right) - \inf_{\pi^* \in Q} \sum_{t=1}^{T} L\left(\pi^*, z^t\right) \leq \mathbf{Regret}^L(\mathcal{A}, Q, T),$$

where $\pi^t = \mathcal{A}((\pi^j, z^j)_{j=1}^{t-1})$. When it is not clear from the context, we will use subscript to denote the setting – e.g. $\mathbf{Regret}^L_{\text{FAIR-BATCH}}$.

We wish to develop an algorithm such that both the misclassfication and fairness loss regret is sublinear, which is often called no-regret. Note that because $\pi^* \in Q_\alpha$ is α-fair on $\bar{x}^t$ for all $t \in [T]$, we have $\text{Unfair}_\alpha(\pi^*, z^t) = 0$ for all $t \in [T]$. Hence, achieving $\mathbf{Regret}^{\text{Unfair}_\alpha}_{\text{FAIR-BATCH}}(\mathcal{A}, Q, T) = o(T)$ is equivalent to ensuring that the total number of rounds with any α-fairness violation is sublinear. Therefore, our goal is equivalent to developing an algorithm $\mathcal{A}$ so that for any $(z^t)_{t=1}^T$,

$$\mathbf{Regret}^{\text{Err}}_{\text{FAIR-BATCH}}(\mathcal{A}, Q, T) = o(T) \quad \text{and} \quad \sum_{t=1}^{T} \text{Unfair}_\alpha(\pi^t, z^t) = o(T).$$

To achieve the result above, we will reduce our setting to a setting with no fairness constraint, which we call *online batch classification* problem. Similar to the online fair batch classification setting, in each round t, the learner deploys a policy π^t, but the environment chooses only a batch of instances $(x^t_\tau, y^t_\tau)_{\tau=1}^k$. In online batch classification, we denote the strategy that the environment can take with $\mathcal{Z}_{\text{BATCH}} = \mathcal{X}^k \times \mathcal{Y}^k$. We compare the two settings in figure 1.

3 Achieving No Regret Simultaneously

Because we wish to achieve no-regret with respect to both the misclassification and fairness loss, it is natural to consider a hybrid loss that combines them together. In fact, we define a round-based Lagrangian loss and show that the regret with respect to our Lagrangian loss also serves as the misclassification and the fairness complaint regret.

Furthermore, we show that using an auditor that can detect any fairness violation beyond certain threshold, we can still hope to achieve no-regret against an adaptive adversary.

Finally, we show how to achieve no regret with respect to the Lagrangian loss by reducing the problem to an online batch classification where there's no fairness constraint. We show that Follow-The-Perturbed-Leader style approach (CONTEXT-FTPL from Syrgkanis et al. [2016]) can achieve sublinear regret in the online batch classification setting, which allows us to achieve sublinear regret with respect to both misclassification and fairness loss in the online fair batch classification setting.

3.1 Lagrangian Formulation

Here we present a hybrid loss that we call *Lagrangian loss* that combines the misclassification loss and the magnitude of the fairness loss of round t.

Definition 3.1 (Lagrangian Loss)**.** *The (C, α)-Lagrangian loss of π is*

$$\mathcal{L}_{C,\alpha}\left(\pi, \left((\bar{x}^t, \bar{y}^t), \rho^t\right)\right) = \sum_{\tau=1}^{k} \ell\left(\pi\left(x_\tau^t\right), y_\tau^t\right) + \begin{cases} C\left(\pi(x_{\rho_1}^t) - \pi(x_{\rho_2}^t) - \alpha\right) & \rho^t = (\rho_1, \rho_2) \\ 0 & \rho^t = null \end{cases}$$

Given an auditor $\mathcal{J}_\alpha$ that can detect any α-fairness violation, we can simulate the online fair batch classification setting with an auditor $\mathcal{J}_\alpha$ by setting the pair $\rho^t_{\mathcal{J}} = \mathcal{J}_\alpha(\bar{x}^t, \pi^t)$: subscript $\mathcal{J}$ is placed on this pair to distinguish from the pair chosen by the environment. This distinction between the pair chosen by the environment and the auditor is necessary as we need to ensure that the pair used to charge the Lagrangian loss incurs constant instantaneous regret in the rounds where there is actually some fairness violation. This will be made more clear in the proof of Theorem 3.4.

Definition 3.2 (Lagrangian Regret)**.** *Algorithm $\mathcal{A}$'s $(C, \alpha, \mathcal{J}_{\alpha'})$-Lagrangian regret against Q is $\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha'}}(\mathcal{A}, Q, T)$, if for any $\left(\bar{x}^t, \bar{y}^t\right)_{t=1}^{T}$, we have*

$$\sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^t, (\bar{x}^t, \bar{y}^t), \rho^t_{\mathcal{J}}) - \min_{\pi^* \in Q} \sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^*, (\bar{x}^t, \bar{y}^t), \rho^t_{\mathcal{J}}) \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha'}}(\mathcal{A}, Q, T),$$

where $\rho^t_{\mathcal{J}} = \mathcal{J}_{\alpha'}(\bar{x}^t, \pi^t)$.

Remark 3.3. *From here on, we assume the auditor has a given sensitivity denoted by $\alpha' = \alpha + \epsilon$, where ϵ is a parameter we will fix in order to define our desired benchmark Q_α.*

Now, we show that the Lagrangian regret upper bounds the α-fairness loss regret with some slack by setting C to be appropriately big enough. Here's a high-level proof sketch. For every round t where there is any $(\alpha + \epsilon)$-fairness violation, the auditor $\mathcal{J}_{\alpha+\epsilon}$ would have found a pair with $(\alpha + \epsilon)$-fairness violation, so the fairness penalty term in the Lagrangian loss must be at least ϵ. This ensures that the instantaneous regret of that round is at least 1, and hence, we can bound the number of such rounds by the Lagrangian regret.

Theorem 3.4. *Fix some small constant $\epsilon > 0$. For any sequence of environment's strategy $(z^t)_{t=1}^T \in \mathcal{Z}^T_{\text{FAIR-BATCH}}$,*

$$\sum_{t=1}^{T} \textit{Unfair}_{\alpha+\epsilon}(\pi^t, z^t) \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}\left(\mathcal{A}, Q_\alpha, T\right),$$

where $C \geq \frac{k+1}{\epsilon}$.

Proof. Fix $(z^t)_{t=1}^T = \left((\bar{x}^t, \bar{y}^t) \times \rho^t\right)_{t=1}^T$ and any $\pi^* \in Q_\alpha$. Consider any round t where there is some $(\alpha + \epsilon)$-fairness violation with respect to $\bar{x}^t$, meaning $\exists \tau, \tau' \in [k]. v_{\alpha+\epsilon}(\pi^t, (x_\tau^t, x_{\tau'}^t)) > 0$. During those rounds, the auditor $\mathcal{J}_{\alpha+\epsilon}$ will report one of those violations $\rho^t_{\mathcal{J}} = (\rho_1^t, \rho_2^t)$. During these rounds with some $(\alpha + \epsilon)$-fairness violation, we show that the instantaneous regret with respect to the Lagrangian loss must be at least 1:

$$\begin{aligned}
&\mathcal{L}_{C,\alpha}(\pi^t,(\bar{x}^t,\bar{y}^t),\rho^t_{\mathcal{J}}) - \mathcal{L}_{C,\alpha}(\pi^*,(\bar{x}^t,\bar{y}^t),\rho^t_{\mathcal{J}}) \\
&\geq \left(\sum_{\tau=1}^{k} \ell(\pi^t(x^t_\tau),y^t_\tau) + C\left(\pi(x^t_{\rho^t_1}) - \pi(x^t_{\rho^t_2}) - \alpha\right)\right) - \left(\sum_{\tau=1}^{k} \ell(\pi^*(x^t_\tau),y^t_\tau) + Cd(x^t_{\rho^t_1},x^t_{\rho^t_2})\right) \qquad (1) \\
&= \left(\sum_{\tau=1}^{k} \ell(\pi^t(x^t_\tau),y^t_\tau) - \sum_{\tau=1}^{k} \ell(\pi^*(x^t_\tau),y^t_\tau)\right) + C\left(\pi^t(x^t_{\rho^t_1}) - \pi^t(x^t_{\rho^t_2}) - d\left(x^t_{\rho^t_1},x^t_{\rho^t_2}\right) - \alpha\right) \\
&\geq -k + C\epsilon \qquad (2) \\
&\geq 1,
\end{aligned}$$

where (1) follows from Lemma A.1, and (2) from the fact that the pair $\rho^t_{\mathcal{J}} = (\rho^t_1, \rho^t_2)$ found by the auditor $\mathcal{J}_{\alpha+\epsilon}$ must have had $(\alpha+\epsilon)$-fairness violation, so the magnitude of α-fairness violation must have been at least ϵ.

Finally, we bound the number of these rounds by the Lagrangian regret.

$$\begin{aligned}
&\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T) \\
&\geq \sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^t,(\bar{x}^t,\bar{y}^t),\rho^t_{\mathcal{J}}) - \mathcal{L}_{C,\alpha}(\pi^*,(\bar{x}^t,\bar{y}^t),\rho^t_{\mathcal{J}}) \\
&\geq \sum_{t=1}^{T} \left(\mathcal{L}_{C,\alpha}(\pi^t,(\bar{x}^t,\bar{y}^t),\rho^t_{\mathcal{J}}) - \mathcal{L}_{C,\alpha}(\pi^*,(\bar{x}^t,\bar{y}^t),\rho^t_{\mathcal{J}})\right) \cdot \mathbb{1}\left(\exists \tau, \tau' \in [k]. v_{\alpha+\epsilon}(\pi^t,(x^t_\tau, x^t_{\tau'})) > 0\right) \\
&\geq \sum_{t=1}^{T} \mathbb{1}\left(\exists \tau, \tau' \in [k]. v_{\alpha+\epsilon}(\pi^t,(x^t_\tau, x^t_{\tau'})) > 0\right) \\
&\geq \sum_{t=1}^{T} \text{Unfair}_{\alpha+\epsilon}(\pi^t, z^t)
\end{aligned}$$

Note that z^t in the last inequality contains ρ^t and not $\rho^t_{\mathcal{J}}$. $\square$

Also, we show that $(C, \alpha, \mathcal{J}_{\alpha+\epsilon})$-Lagrangian regret serves as the misclassification loss regret, too. The proof is given in Appendix A.1.

Theorem 3.5. *Fix some small constant $\epsilon > 0$. For any sequence of $(z^t)_{t=1}^T \in \mathcal{Z}^T_{\text{FAIR-BATCH}}$ and $\pi^* \in Q_\alpha$,*

$$\sum_{t=1}^{T}\sum_{\tau=1}^{k} \ell\left(\pi^t\left(x^t_\tau\right), y^t_\tau\right) - \sum_{t=1}^{T}\sum_{\tau=1}^{k} \ell(\pi^*(x^t_\tau), y^t_\tau) \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T),$$

where $C \geq \frac{k+1}{\epsilon}$. In other words, $\mathbf{Regret}^{Err}_{\text{FAIR-BATCH}}(\mathcal{A}, Q, T) \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T)$.

3.2 Reduction to Online Batch Classification

In this subsection, we will first discuss a computationally inefficient way to achieve no regret with respect to the Lagrangian loss. Then, we will show an efficient reduction to online batch classification and discuss an example of an oracle-efficient algorithm $\mathcal{A}_{\text{BATCH}}$ that achieves no-regret.

It is well known that for linear loss, exponential weights with appropriately tuned learning rate γ can achieve no regret [Freund and Schapire, 1997, Cesa-Bianchi et al., 1997, Arora et al., 2012]. Note that our Lagrangian loss

$$\mathcal{L}^t_{C,\alpha}(\pi) = \mathcal{L}_{C,\alpha}(\pi, z^t) = \sum_{\tau=1}^{k}(1 - \pi(x^t_\tau)) \cdot y^t_\tau + \pi(x^t_\tau) \cdot (1 - y^t_\tau)$$
$$+ \begin{cases} C\left(\pi(x^t_{\rho_1}) - \pi(x^t_{\rho_2}) - \alpha\right) & \rho^t = (\rho_1, \rho_2) \\ 0 & \rho^t = null \end{cases}$$

is linear in π for any z^t, and its range is $[0, C + k]$. Therefore, running exponential weights with $\gamma = \sqrt{\frac{\ln(|\mathcal{H}|)}{T}}$, we achieve the following regret with respect to the Lagrangian loss:

Corollary 3.6. *Running exponential weights with $\gamma = \sqrt{\frac{\ln(|\mathcal{H}|)}{T}}$ and $C \geq \frac{k+1}{\epsilon}$, we achieve*

$$\mathbf{Regret}^{Err}_{\text{FAIR-BATCH}}(\mathcal{A}, Q_\alpha, T) \leq (C + k)\sqrt{\ln(|\mathcal{H}|)T}$$

$$\sum_{t=1}^{T} \mathit{Unfair}_{\alpha+\epsilon}(\pi^t, z^t) \leq (C + k)\sqrt{\ln(|\mathcal{H}|)T}.$$

Nevertheless, running exponential weights is not efficient as it needs to calculate the loss for each $h \in \mathcal{H}$ every round t. To design an oracle-efficient algorithm, we reduce the online batch fair classification problem to the online batch classification problem in an efficient manner and use any online batch algorithm $\mathcal{A}_{\text{BATCH}}((\pi^j, (\bar{x}'^j, \bar{y}'^j))_{j=1}^t)$ as a black box. At a high level, our reduction involves just carefully transforming our online fair batch classification history up to t, $(\pi^j, (\bar{x}^j, \bar{y}^j, \rho^j))_{j=1}^t \in (\Delta\mathcal{H} \times \mathcal{Z}_{\text{FAIR-BATCH}})^t$ into some fake online batch classification history $(\pi^j, (\bar{x}'^j, \bar{y}'^j))_{j=1}^t \in (\Delta\mathcal{H} \times \mathcal{Z}_{\text{BATCH}})^t$ and then feeding the created artificial history to $\mathcal{A}_{\text{BATCH}}$.

Without loss of generality, we assume that $C \geq \frac{k+1}{\epsilon}$ is an integer; if it's not, then take the ceiling. Now, we describe how the transformation of the history works. For each round t, whenever $\rho^t = (\rho^t_1, \rho^t_2)$, we add C copies of $(x^t_{\rho^t_1}, 0)$ and $(x^t_{\rho^t_2}, 1)$ to the original pairs to form $\bar{x}'^t$ and $\bar{y}'^t$. Just to keep the batch size the same across each round, even if $\rho^t = null$, we add C copies of $(v, 0)$ and $(v, 1)$ where v is some arbitrary instance in $\mathcal{X}$. Setting $k' = k + 2C$, we describe this process in algorithm 3.

This reduction essentially preserves the regret, which we formally state in Lemma 3.7 and Theorem 3.8.

Lemma 3.7. *For any sequence of $(\pi^t)_{t=1}^T$, $(z^t)_{t=1}^T \in \mathcal{Z}^T_{\text{FAIR-BATCH}}$, and $\pi^* \in \Delta\mathcal{H}$,*

$$\sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^t, z^t) - \sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^*, z^t) = \sum_{t=1}^{T}\sum_{\tau=1}^{k'} \ell(\pi^t(x'^t_\tau), y'^t_\tau) - \sum_{t=1}^{T}\sum_{\tau=1}^{k'} \ell(\pi^*(x'^t_\tau), y'^t_\tau)$$

Proof. It is sufficient to show that in each round t,

$$\mathcal{L}_{C,\alpha}(\pi^t, z^t) - \mathcal{L}_{C,\alpha}(\pi^*, z^t) = \sum_{\tau=1}^{k'} \ell(\pi(x^t_\tau), y^t_\tau) - \sum_{\tau=1}^{k'} \ell(\pi(x^*_\tau), y^t_\tau)$$

First, assume $\rho^t = (\rho^t_1, \rho^t_2)$.

Algorithm 3: Reduction from Online Fair Batch Classification to Online Batch Classification

for $t = 1, \ldots, T$ **do**
  Learner deploys π^t
  Environment chooses $(\bar{x}^t, \bar{y}^t)$ and the pair ρ^t
  if $\rho^t = (\rho_1^t, \rho_2^t)$ **then**
    for $i = 1, \ldots, C$ **do**
      $x_{k+i}^t = x_{\rho_1^t}^t$ and $y_{k+i}^t = 0$
      $x_{k+C+i}^t = x_{\rho_2^t}^t$ and $y_{k+C+i}^t = 1$
    end
  end
  else
    for $i = 1, \ldots, C$ **do**
      $x_{k+i}^t = v$ and $y_{k+i}^t = 0$
      $x_{k+C+i}^t = v$ and $y_{k+C+i}^t = 1$
    end
  end
  $k' = k + 2C$
  $\bar{x}'^t = (x_\tau^t)_{\tau=1}^{k'}$ and $\bar{y}'^t = (y_\tau^t)_{\tau=1}^{k'}$
  $\pi^{t+1} = \mathcal{A}_{\text{BATCH}}\left((\pi^j, \bar{x}'^j, \bar{y}'^j)_{j=1}^t\right)$
end

$$
\begin{aligned}
&\mathcal{L}_{C,\alpha}(\pi^t, z^t) - \mathcal{L}_{C,\alpha}(\pi^*, z^t) \\
&= \left(\sum_{\tau=1}^{k} \ell(\pi^t(x_\tau^t), y_\tau^t) + C(\pi^t(x_{\rho_1^t}^t) - \pi^t(x_{\rho_2^t}^t) - \alpha)\right) - \left(\sum_{\tau=1}^{k} \ell(\pi^*(x_\tau^t), y_\tau^t) + C(\pi^*(x_{\rho_1^t}^t) - \pi^*(x_{\rho_2^t}^t) - \alpha)\right) \\
&= \left(\sum_{\tau=1}^{k} \ell(\pi^t(x_\tau^t), y_\tau^t) + \left(\sum_{\tau=1}^{C} \ell(\pi^t(x_{\rho_1^t}^t), 0) + \sum_{\tau=1}^{C} \ell(\pi^t(x_{\rho_2^t}^t), 1) - C\right)\right) \\
&- \left(\sum_{\tau=1}^{k} \ell(\pi^*(x_\tau^t), y_\tau^t) + \left(\sum_{\tau=1}^{C} \ell(\pi^*(x_{\rho_1^t}^t), 0) + \sum_{\tau=1}^{C} \ell(\pi^*(x_{\rho_2^t}^t), 1) - C\right)\right) \\
&= \sum_{\tau=1}^{k'} \ell(\pi^t(x_\tau'^t), y_\tau'^t) - \sum_{\tau=1}^{k'} \ell(\pi^*(x_\tau'), y_\tau'^t),
\end{aligned}
$$

The second equality follows from the fact that for any π and x,

$$\ell(\pi(x), 0) = \pi(x) \quad \text{and} \quad \ell(\pi(x), 1) = 1 - \pi(x).$$

If $\rho^t = null$, then the same argument applies as above; the only difference is that all the $\pi^t(v)$ will cancel with each other because the number of copies with label 0 is exactly the same as that of label 1. ☐

Theorem 3.8. *For any sequence of* $(z^t)_{t=1}^T \in \mathcal{Z}_{\text{FAIR-BATCH}}^T$ *and* $Q \subseteq \Delta\mathcal{H}$,

$$\sum_{t=1}^T \mathcal{L}_{C,\alpha}(\pi^t, z^t) - \sum_{t=1}^T \mathcal{L}_{C,\alpha}(\pi^*, z^t) \leq \mathbf{Regret}_{\text{BATCH}}^{Err}(\mathcal{A}, Q, T),$$

where $\pi^t = \mathcal{A}_{\text{BATCH}}\left((\pi^j, \bar{x}'^j, \bar{y}'^j)_{j=1}^{t-1}\right)$. *In other words,*

$$\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T) \leq \mathbf{Regret}_{\text{BATCH}}^{Err}(\mathcal{A}, Q, T).$$

One example of $\mathcal{A}_{\text{BATCH}}$ that achieves sublinear regret in online batch classification is CONTEXT-FTPL from Syrgkanis et al. [2016]. We defer the details to Appendix A.3 and present the regret guarantee here. We only focus on their small separator set setting, although their transductive setting naturally follows as well.

Theorem 3.9. *If the separator set* S *for* $\mathcal{H}$ *is of size* s*, then* CONTEXT-FTPL *achieves the following misclassification and fairness regret in the online fair batch classification setting.*

$$\mathbf{Regret}_{\text{FAIR-BATCH}}^{Err}(\mathcal{A}, Q_\alpha, T) \leq O\left(\left(\frac{sk}{\epsilon}\right)^{\frac{3}{4}} \sqrt{T \log(|\mathcal{H}|)}\right)$$

$$\sum_{t=1}^T \mathit{Unfair}_{\alpha+\epsilon}(\pi^t, z^t) \leq O\left(\left(\frac{sk}{\epsilon}\right)^{\frac{3}{4}} \sqrt{T \log(|\mathcal{H}|)}\right)$$

4 Generalization

We observe that until this point, all of our results apply to the more general setting where individuals arrive in any adversarial fashion. In order to argue about generalization, in this section, we will assume the existence of an (unknown) data distribution from which individual arrivals are drawn:

$$\{\{(x_\tau^t, y_\tau^t)\}_{\tau=1}^k\}_{t=1}^T \sim_{i.i.d.} \mathcal{D}^{Tk}$$

Despite the data are drawn i.i.d., there are two technical challenges in establishing generalization guarantee: (1) the auditor's fairness feedback at each round is limited to a single fairness violation with regards to the policy deployed in that round, and (2) both the deployed policies and the auditor are adaptive over rounds. To overcome these challenges, we will draw connection between regret guarantees in Section 3 and the learner's distributional accuracy and fairness guarantees. In particular, we will analyze the generalization bounds for the average policy over rounds.

Definition 4.1 (Average Policy). *Let* π^t *be the policy deployed by the algorithm at round* t*. The average policy* π^{avg} *is defined by:*

$$\forall x : \pi^{avg}(x) = \frac{1}{T} \sum_{t=1}^T \pi^t(x)$$

In order to be consistent with Section 3, we denote $\alpha' = \alpha + \epsilon$ in this section.

Here, we state the main results of this section:

Theorem 4.2 (Accuracy Generalization)**.** *With probabilty $1-\delta$, the misclassification loss of π^{avg} is upper bounded by*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi^{avg}(x),y)] \le \inf_{\pi\in Q_\alpha}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi(x),y)] + \frac{1}{kT}\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A},Q_\alpha,T) + \sqrt{\frac{8\ln\left(\frac{4}{\delta}\right)}{T}}$$

Theorem 4.3 (Fairness Generalization)**.** *Assume that for all t, π^t is (α,β^t)-fair ($0\le\beta^t\le 1$). With probability $1-\delta$, for any integer $q\le T$, π^{avg} is $(\alpha'+\frac{q}{T},\beta^*)$-fair where*

$$\beta^* = \frac{1}{q}\left(\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A},Q_\alpha,T) + \sqrt{2T\ln\left(\frac{2}{\delta}\right)}\right).$$

Corollary 4.4. *With probability $1-\delta$, exponential weights with learning rate $\gamma=\sqrt{\frac{\ln(|\mathcal{H}|)}{T}}$ achieves*

1. ***Accuracy:***

$$\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi^{avg}(x),y)] \le \inf_{\pi\in Q_\alpha}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi(x),y)] + O\left(\sqrt{\frac{\ln(|\mathcal{H}|)+\ln(\frac{1}{\delta})}{\epsilon^2 T}}\right).$$

2. ***Fairness:*** *π^{avg} is $(\alpha'+\lambda,\lambda)$-fair where*

$$\lambda = O\left(\left(\frac{k}{\epsilon}\right)^{\frac{1}{2}}\left(\frac{\ln\left(|\mathcal{H}|\right)+\ln\left(\frac{1}{\delta}\right)}{T}\right)^{\frac{1}{4}}\right)$$

Corollary 4.5. *Using* CONTEXT-FTPL *from Syrgkanis et al. [2016] with a separator set of size s, with probability $1-\delta$, the average policy π^{avg} has the following guarantee:*

1. ***Accuracy:***

$$\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi^{avg}(x),y)] \le \inf_{\pi\in Q_\alpha}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi(x),y)] + O\left(\frac{1}{k^{\frac{1}{4}}}\left(\frac{s}{\epsilon}\right)^{\frac{3}{4}}\sqrt{\frac{\ln(|\mathcal{H}|)+\ln\left(\frac{1}{\delta}\right)}{T}}\right).$$

2. ***Fairness:*** *π^{avg} is $(\alpha'+\lambda,\lambda)$-fair where*

$$\lambda = O\left(\left(\frac{sk}{\epsilon}\right)^{\frac{3}{4}}\left(\frac{\ln\left(|\mathcal{H}|\right)+\ln\left(\frac{1}{\delta}\right)}{T}\right)^{\frac{1}{4}}\right)$$

Remark 4.6. *Recall that the sensitivity of the auditor α' is fixed, and the learner chooses the parameter $\epsilon\in(0,\alpha')$, which in return determines $\alpha=\alpha'-\epsilon$ and the set of policy Q_α the learner is competing against. In the case where $\alpha'=\Omega(1)$, the learner can choose ϵ in the order of $\Omega(1)$ and guarantee that π^{avg} is $(\alpha'+\lambda,\lambda)$-fair with $\lambda=\tilde{O}(T^{-1/4})$. In this regime, corollary 4.4 and 4.5 imply that policy π^{avg} has a non-trivial accuracy guarantee and a fairness generalization bound that qualitatively matches the uniform convergence bound of Yona and Rothblum [2018].*

The accuracy generalization bound of Theorem 4.2 is attained by applying Azuma's inequality on the left hand side of the inequality in Theorem 3.5 and then leveraging the fact that our classification loss function is linear with respect to the base classifiers over which it is defined. The full proof is given in Appendix B.

A more challenging task is, however, to argue about the fairness generalization guarantee of the average policy (Theorem 4.3). To provide some intuition for why this is the case, let us first attempt to upper bound the probability of running into an α'-fairness violation by the average policy on a randomly selected pair of individuals:

Observation 4.7. *Suppose for all t, π^t is (α', β^t)-fair. Then, π^{avg} is $\left(\alpha', \sum_{t=1}^{T} \beta^t\right)$-fair.*

This bound is very dissatisfying, as the statement is vacuous when $\sum_{t=1}^{T} \beta^t \geq 1$. The reason for such a weak guarantee is that by aiming to upper bound the unfairness probability for the original fairness violation threshold α', we are subject to worst-case compositional guarantees[3] in the sense that the average policy may result to have an α'-fairness violation on any fraction of the distribution (over pairs) where one or more of the deployed policies induces an α'-fairness violation. This bound is tight, and we refer the reader to Appendix B for the full example.

Interpolating α and β To circumvent the above setback, our strategy will be to relax the target violation threshold of the desired fairness guarantee of the average policy to $\alpha'' > \alpha'$. How big should we set α''? A good intuition may arrive from considering the following thought experiment: assume worst-case compositional guarantees, and then, select a pair of individuals x, x' on which the average policy has an α''-fairness violation. We aim to lower bound the number of policies from $\{\pi^1, \ldots, \pi^T\}$ that have an α'-fairness violation on this pair. As we will see, setting α'' to be sufficiently larger will force the number of these policies required to produce an α''-fairness violation of the average policy on x, x' to be high, resulting in the following improved bound:

Lemma 4.8. *Assume that for all t, π^t is (α', β^t)-fair $(0 \leq \beta^t \leq 1)$. For any integer $q \leq T$, π^{avg} is $\left(\alpha' + \frac{q}{T}, \frac{1}{q}\sum_{t=1}^{T} \beta^t\right)$-fair.*

High-Level Proof Idea Setting $\alpha'' = \alpha' + \frac{q}{T}$ has the following implication: for any pair of individuals (x, x'), in order for π^{avg} to have an α''-fairness violation on x, x', at least q of the policies in $\{\pi^1, \ldots, \pi^T\}$ must have an α'-fairness violation on x, x'. We will then say a subset $A \subseteq \mathcal{X} \times \mathcal{X}$ is α'-covered by a policy π, if π has an α'-violation on every element in A. We will denote by $A_q^{\alpha'} \subseteq \mathcal{X} \times \mathcal{X}$ the subset of pairs of elements from $\mathcal{X}$ that are α'-covered by at least q policies in $\{\pi^1, \ldots, \pi^T\}$. Next, consider the probability space $\mathcal{D}|_\mathcal{X} \times \mathcal{D}|_\mathcal{X}$ over pairs of individuals. The lemma then follows from observing that for any $q \leq T$, $\Pr(A_q^{\alpha'}) \leq \frac{1}{q}\Pr(A_1^{\alpha'})$, as this will allow us to upper bound the probability of an α''-fairness violation by the stated bound.

[3]This is the case where, for every pair of individuals x, x' on which there exists a policy in $\{\pi^1, \ldots, \pi^T\}$ that has an α'-fairness violation: (1) every policy $\pi \in \{\pi^1, \ldots, \pi^T\}$ that has an α-fairness violation on x, x' has a maximal violation (of value 1), (2) all non-violating policies (in $\{\pi^1, \ldots, \pi^T\}$) on x, x' are arbitrarily close to the violation threshold α' on this pair, and (3) all of the directions of the policies' predictions' differences on x, x' are correlated (no cancellations averaging over policies).

Proof. Fix $\{\pi^t\}_{t=1}^T$, and assume that $\forall t : \pi^t$ is (α', β^t)-fair. If we set $q \leq T$, then we know that

$$
\begin{aligned}
&\Pr_{x,x'}\left[\left|\pi^{avg}(x) - \pi^{avg}(x')\right| - d(x,x') > \alpha' + \frac{q}{T}\right] \\
&= \Pr_{x,x'}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\left[\pi^t(x) - \pi^t(x')\right]\right| - d(x,x') > \alpha' + \frac{q}{T}\right] \\
&\leq \Pr_{x,x'}\left[\left[\frac{1}{T}\sum_{t=1}^{T}\left|\pi^t(x) - \pi^t(x')\right|\right] - d(x,x') > \alpha' + \frac{q}{T}\right] \\
&\leq \Pr_{x,x'}\left[\exists\{i_1,\ldots,i_q\} \subseteq [T], \forall j, |\pi^{i_j}(x) - \pi^{i_j}(x')| - d(x,x') > \alpha'\right] && (3) \\
&\leq \frac{1}{q}\sum_{t=1}^{T}\Pr_{x,x'}\left[|\pi^t(x) - \pi^t(x')| - d(x,x') > \alpha'\right] && (4) \\
&\leq \frac{1}{q}\sum_{t=1}^{T}\beta^t
\end{aligned}
$$

Transition 3 is given by the following observation: fix any x, x' and assume

$$\left|\{\pi^t : t \in [T], |\pi^t(x) - \pi^t(x')| - d(x,x') > \alpha'\}\right| \leq q$$

Then, we have

$$|\pi^{avg}(x) - \pi^{avg}(x')| - d(x,x') \leq \frac{q + (T-q)\alpha'}{T} = \alpha' + \frac{q}{T} - \frac{\alpha' q}{T} < \alpha' + \frac{q}{T}.$$

Transition 4 stems from the following argument: for any x, x', denote by

$$V_{x,x'}^{\alpha'} := \{\pi^t : t \in [T], |\pi^t(x) - \pi^t(x')| - d(x,x') > \alpha'\}$$

the subset of deployed policies that have an α'-fairness violation on x, x'. We know that

$$
\begin{aligned}
&\frac{1}{q}\sum_{t=1}^{T}\Pr_{x,x'}\left[|\pi^t(x) - \pi^t(x')| - d(x,x') > \alpha'\right] \\
&= \frac{1}{q}\sum_{t=1}^{T}\int_x\int_{x'}\mathbb{P}(x,x')\cdot\mathbb{1}[|\pi^t(x) - \pi^t(x')| - d(x,x') > \alpha']dx'dx \\
&= \frac{1}{q}\int_x\int_{x'}\sum_{t=1}^{T}\mathbb{P}(x,x')\cdot\mathbb{1}[|\pi^t(x) - \pi^t(x')| - d(x,x') > \alpha']dx'dx \\
&= \frac{1}{q}\int_x\int_{x'}\sum_{\pi^t\in V_{x,x'}^{\alpha'} : \left|V_{x,x'}^{\alpha'}\right| \geq q}\mathbb{P}(x,x') + \sum_{\pi^t\in V_{x,x'}^{\alpha'} : \left|V_{x,x'}^{\alpha'}\right| < q}\mathbb{P}(x,x')dx'dx \\
&\geq \int_x\int_{x'}\mathbb{P}(x,x')\cdot\mathbb{1}\left[|V_{x,x'}^{\alpha'}| \geq q\right]dx'dx \\
&= \Pr_{x,x'}\left[\exists\{i_1,\ldots,i_q\} \subseteq [T], \forall j, |\pi^{i_j}(x) - \pi^{i_j}(x')| - d(x,x') > \alpha'\right],
\end{aligned}
$$

where $\mathbb{P}(x,x')$ denotes the probability measure of x, x' defined by $\mathcal{D}|_{\mathcal{X}} \times \mathcal{D}|_{\mathcal{X}}$. This concludes the proof. $\square$

Equipped with Lemma 4.8, we next note that the stochastic assumption allows us to link the distributional fairness guarantees of the deployed policies to the algorithm's regret. The implication of this, which we state next, is that the case of deploying too many highly unfair policies along the interaction must result in a contradiction to the algorithm's proven regret guarantee. Hence, we will be able to bound the sum of the distributional α'-fairness guarantees of all the policies deployed by the algorithm.

Lemma 4.9. *With probability* $1-\delta$*, we have*

$$\sum_{t=1}^{T} \beta^t \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T) + \sqrt{2T \ln\left(\frac{2}{\delta}\right)}$$

Proof. Fix the sequence $(z^t)_{t=1}^T$ and also the random coins of the algorithm, meaning $\mathcal{A}((\pi^j, z^j)_{j=1}^{t-1})$ is deterministic. Then, $(\pi^t)_{t=1}^T$ is fixed as well. By Theorem 3.4, for any sequence of environment's strategy $(z^t)_{t=1}^T$,

$$\sum_{t=1}^{T} \mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}^t, \pi^t\right) \neq \emptyset\right) = \sum_{t=1}^{T} \text{Unfair}_{\alpha+\epsilon}(\pi^t, z^t) \leq \mathbf{Regret}^{\mathcal{L}_{C,\alpha},\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T),$$

where $C = \frac{k+1}{\epsilon}$.

Lemma 4.10.

$$\Pr_{\substack{(\bar{x}^t,\bar{y}^t)_{t=1}^T \\ \pi^t=\mathcal{A}(\cdots)}}\left[\left|\sum_{t=1}^{T} \mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}'^t, \pi^t\right) \neq \emptyset\right) - \mathop{\mathbb{E}}_{(\bar{x}'^t,\bar{y}'^t)_{t=1}^T}\left[\sum_{t=1}^{T} \mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}'^t, \pi^t\right) \neq \emptyset\right)\right]\right| \geq \gamma\right] \leq 2\exp\left(-\frac{\gamma^2}{2T}\right)$$

Proof. Consider the following sequence $(B^t)_{t=1}^T$:

$$B^t = \sum_{j=1}^{t} \mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}^j, \pi^j\right) \neq \emptyset\right) - \mathop{\mathbb{E}}_{(\bar{x}'^j,\bar{x}'^j)_{j=1}^t}\left[\sum_{j=1}^{t} \mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}'^j, \pi^j\right) \neq \emptyset\right)\right].$$

Note that this is a martingale: $\mathbb{E}[B^t | B^1, \ldots, B^{t-1}] = B^{t-1}$ because conditioning upon previous rounds, the algorithm π^t is deterministically chosen.

Now, we apply Azuma's inequality. Since $|B^t - B^{t-1}| \leq 1$, we have

$$\Pr\left[|B^T - B^1| \geq \gamma\right] \leq 2\exp\left(\frac{-\gamma^2}{2T}\right).$$

$\square$

Lemma 4.11. *Fix the sequence of policies* $(\pi^t)_{t=1}^T$*. Say that each* π^t *is* (α, β^t)*-fair. Then:*

$$\sum_{t=1}^{T} \mathop{\mathbb{E}}_{(\bar{x}'^t,\bar{y}'^t)_{t=1}^T}\left[\mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}'^t, \pi^t\right) \neq \emptyset\right)\right] \geq \sum_{t=1}^{T} \beta^t.$$

Proof. We will lower bound the probability of having an α'-fairness violation on a pair of individuals among those who have arrived in a single round. Because all possible pairs within a single batch of independently arrived individuals are not independent, we resort to a weaker lower bound by dividing all of the arrivals into independent pairs. As we go through the batch, we take every two individuals to form a pair, and note that these pairs must be independent.

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{(\bar{x}'^t,\bar{y}'^t)_{t=1}^T}\left[\mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}'^t,\pi^t\right)\neq\emptyset\right)\right]\\
&=\mathop{\Pr}_{(\bar{x}'^t,\bar{y}'^t)_{t=1}^T}\left[\exists\tau,\tau'\in[k]:|\pi^t(x'^t_\tau)-\pi^t(x'^t_{\tau'})|-d(x'^t_\tau,x'^t_{\tau'})>\alpha'\right]\\
&\geq\Pr\left[\exists i\in\left[\left\lfloor\frac{k}{2}\right\rfloor\right]:|\pi^t(x'^t_{2i-1})-\pi^t(x'^t_{2i})|-d(x'^t_{2i-1},x'^t_{2i})>\alpha'\right]\\
&=1-\Pr\left[\forall i\in\left[\left\lfloor\frac{k}{2}\right\rfloor\right]:|\pi^t(x'^t_{2i-1})-\pi^t(x'^t_{2i})|-d(x'^t_{2i-1},x'^t_{2i})\leq\alpha'\right]\\
&=1-\prod_{i=1}^{\lfloor\frac{k}{2}\rfloor}(1-\beta^t)\\
&=1-(1-\beta^t)^{\lfloor\frac{k}{2}\rfloor}\\
&\geq\beta^t
\end{aligned}
$$

□

Next, we combine Lemma 4.10 and Lemma 4.11. With probability $1-\delta$,

$$
\begin{aligned}
&\mathbf{Regret}^{\mathcal{L}_{C,\alpha},\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A},Q_\alpha,T)\\
&\geq\sum_{t=1}^T\mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}'^t,\pi^t\right)\neq\emptyset\right)\\
&\geq\sum_{t=1}^T\mathop{\mathbb{E}}_{(\bar{x}'^t,\bar{y}'^t)_{t=1}^T}\left[\mathbb{1}\left(\mathcal{J}_{\alpha+\epsilon}\left(\bar{x}'^t,\pi^t\right)\neq\emptyset\right)\right]-\sqrt{2T\ln\left(\frac{\delta}{2}\right)}\\
&\geq\sum_{t=1}^T\beta^t-\sqrt{2T\ln\left(\frac{\delta}{2}\right)}
\end{aligned}
$$

□

Proof of theorem 4.3 The theorem follows immediately from combining Lemma 4.8 and Lemma 4.9.

5 Conclusion and Future Directions

In this work, we have removed several binding restrictions in the context of learning with individual fairness. We hope that relieving the metric assumption as well as the assumption regarding full access to the similarity measure, and only requiring the auditor to detect a single violation for every time interval, will be helpful in making individual fairness more achievable and easier to implement in practice. As for future work - first of all, it would be interesting to explore the interaction with different models of feedback (one natural variant being one-sided feedback). Second, thinking about a model where the auditor only has access to binary decisions may be helpful in further closing the gap to practical use. Third, as most of the literature on individual fairness (including this work) is decoupling the similarity measure from the distribution over the target variable, it would be desirable to try to explore and quantify the compatibility of the two in specific instances.

6 Acknowledgments

We thank Sampath Kannan, Akshay Krishnamurthy, Katrina Ligett, and Aaron Roth for helpful conversations at an early stage of this work. Part of this work was done while YB, CJ, and ZSW were visiting the Simons Institute for the Theory of Computing. YB is supported in part by Israel Science Foundation (ISF) grant #1044/16, the United States Air Force and DARPA under contracts FA8750-16-C-0022 and FA8750-19-2-0222, and the Federmann Cyber Security Center in conjunction with the Israel national cyber directorate. CJ is supported in part by NSF grant AF-1763307. ZSW is supported in part by the NSF FAI Award #1939606, an Amazon Research Award, a Google Faculty Research Award, a J.P. Morgan Faculty Award, a Facebook Research Award, and a Mozilla Research Grant. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force and DARPA.

References

Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford, and Hanna M. Wallach. 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*, pages 60–69, 2018. URL `http://proceedings.mlr.press/v80/agarwal18a.html`.

J. Angwin, J. Larson, S. Mattu, and L. Kirchner. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. *ProPublica*, 2016.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *J. ACM*, 44(3):427485, May 1997. ISSN 0004-5411. doi: 10.1145/258128.258179. URL `https://doi.org/10.1145/258128.258179`.

Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, Special Issue on Social and Technical Trade-Offs, 2017.

Alexandra Chouldechova and Aaron Roth. The frontiers of fairness in machine learning. *CoRR*, abs/1810.08810, 2018. URL `http://arxiv.org/abs/1810.08810`.

Sam Corbett-Davies and Sharad Goel. The measure and mismeasure of fairness: A critical review of fair machine learning. arXiv, 2018.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. Fairness through awareness. In *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, pages 214–226, 2012. doi: 10.1145/2090236.2090255. URL https://doi.org/10.1145/2090236.2090255.

Anthony Flores, Christopher Lowenkamp, and Kristin Bechtel. False positives, false negatives, and false analyses: A rejoinder to machine bias: Theres software used across the country to predict future criminals. and its biased against blacks.. *Technical report, Crime & Justice Institute*, 2016.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997. doi: 10.1006/jcss.1997.1504. URL https://doi.org/10.1006/jcss.1997.1504.

Stephen Gillen, Christopher Jung, Michael J. Kearns, and Aaron Roth. 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.*, pages 2605–2614, 2018. URL http://papers.nips.cc/paper/7526-online-learning-with-an-unknown-fairness-metric.

Swati Gupta and Vijay Kamble. Individual fairness in hindsight. In *Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019, Phoenix, AZ, USA, June 24-28, 2019*, pages 805–806, 2019. doi: 10.1145/3328526.3329605. URL https://doi.org/10.1145/3328526.3329605.

Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. In *Neural Information Processing Systems (NIPS)*, 2016.

Úrsula Hébert-Johnson, Michael P. Kim, Omer Reingold, and Guy N. Rothblum. Multicalibration: Calibration for the (computationally-identifiable) masses. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 1944–1953, 2018. URL http://proceedings.mlr.press/v80/hebert-johnson18a.html.

Christina Ilvento. Metric learning for individual fairness. *CoRR*, abs/1906.00250, 2019. URL http://arxiv.org/abs/1906.00250.

Shahin Jabbari, Matthew Joseph, Michael J. Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 1617–1626, 2017. URL http://proceedings.mlr.press/v70/jabbari17a.html.

Matthew Joseph, Michael J. Kearns, Jamie H. Morgenstern, and Aaron Roth. 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*, pages 325–333, 2016. URL http://papers.nips.cc/paper/6355-fairness-in-learning-classic-and-contextual-ban

Matthew Joseph, Michael J. Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. 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*, pages 158–163, 2018. doi: 10.1145/3278721.3278764. URL https://doi.org/10.1145/3278721.3278764.

Christopher Jung, Michael J. Kearns, Seth Neel, Aaron Roth, Logan Stapleton, and Zhiwei Steven Wu. Eliciting and enforcing subjective individual fairness. *CoRR*, abs/1905.10660, 2019. URL http://arxiv.org/abs/1905.10660.

Sampath Kannan, Michael J. Kearns, Jamie Morgenstern, Mallesh M. Pai, Aaron Roth, Rakesh V. Vohra, and Zhiwei Steven Wu. Fairness incentives for myopic agents. In *Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17, Cambridge, MA, USA, June 26-30, 2017*, pages 369–386, 2017. doi: 10.1145/3033274.3085154. URL https://doi.org/10.1145/3033274.3085154.

Michael J. Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 2569–2577, 2018. URL http://proceedings.mlr.press/v80/kearns18a.html.

Michael P. Kim, Omer Reingold, and Guy N. Rothblum. 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*, pages 4847–4857, 2018. URL http://papers.nips.cc/paper/7733-fairness-through-computationally-bounded-awarene

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. In *Proceedings of the 8th Innovations in Theoretical Computer Science Conference*, 2017.

Northpointe. Practitioners Guide to COMPAS. 2012.

John Podesta, Penny Pritzker, Ernest J. Moniz, John Holdren, and Jefrey Zients. Big data: Seizing opportunities and preserving values. 2014.

Arun Sai Suggala and Praneeth Netrapalli. Online non-convex learning: Following the perturbed leader is optimal. *CoRR*, abs/1903.08110, 2019. URL http://arxiv.org/abs/1903.08110.

Vasilis Syrgkanis, Akshay Krishnamurthy, and Robert E. Schapire. 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*, pages 2159–2168, 2016. URL http://proceedings.mlr.press/v48/syrgkanis16.html.

Gal Yona and Guy N. Rothblum. Probably approximately metric-fair learning. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 5666–5674, 2018. URL http://proceedings.mlr.press/v80/yona18a.html.

A Omitted Details from Section 3

A.1 Omitted Details from Subsection 3.1

Lemma A.1. *Fix the sequence of the environment's strategies* $(z^t)_{t=1}^T$ *and hence* Q_α. *For any* $t \in [T]$, $\rho = (\rho_1, \rho_2)$, *and* $\pi^* \in Q_\alpha$,

$$\mathcal{L}_{C,\alpha}(\pi^*, (\bar{x}^t, \bar{y}^t), \rho) \leq \sum_{\tau=1}^{k} \ell\left(\pi^*\left(x_\tau^t\right), y_\tau^t\right) + Cd(x_{\rho_1}^t, x_{\rho_2}^t).$$

Proof. We know that $\pi^* \in Q_\alpha$ must be α-fair on $\bar{x}^t$ for any $t \in [T]$, meaning

$$\pi^*(x^t_{\rho_1}) - \pi^*(x^t_{\rho_2}) - d(x^t_{\rho_1}, x^t_{\rho_2}) - \alpha \leq 0$$

Thus, we have

$$\begin{aligned}\mathcal{L}_{C,\alpha}(\pi^*, (\bar{x}^t, \bar{y}^t), (\rho_1, \rho_2)) &= \sum_{\tau=1}^{k} \ell\left(\pi^*\left(x^t_\tau\right), y^t_\tau\right) + C\left(\pi^*(x^t_{\rho_1}) - \pi(x^t_{\rho_2}) - \alpha\right) \\ &\leq \sum_{\tau=1}^{k} \ell\left(\pi^*\left(x^t_\tau\right), y^t_\tau\right) + Cd(x^t_{\rho_1}, x^t_{\rho_2})\end{aligned}$$

□

Theorem 3.5. *Fix some small constant $\epsilon > 0$. For any sequence of $(z^t)_{t=1}^T \in \mathcal{Z}^T_{\text{FAIR-BATCH}}$ and $\pi^* \in Q_\alpha$,*

$$\sum_{t=1}^{T}\sum_{\tau=1}^{k} \ell\left(\pi^t\left(x^t_\tau\right), y^t_\tau\right) - \sum_{t=1}^{T}\sum_{\tau=1}^{k} \ell(\pi^*(x^t_\tau), y^t_\tau) \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}\left(\mathcal{A}, Q_\alpha, T\right),$$

where $C \geq \frac{k+1}{\epsilon}$. In other words, $\mathbf{Regret}^{Err}_{\text{FAIR-BATCH}}(\mathcal{A}, Q, T) \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T)$.

Proof. Fix $(z^t)_{t=1}^T = \left((\bar{x}^t, \bar{y}^t) \times \rho^t\right)_{t=1}^T$ and $\pi^* \in Q_\alpha$.

It is sufficient to show that for any round $t \in [T]$,

$$\sum_{\tau=1}^{k} \ell\left(\pi^t\left(x^t_\tau\right), y^t_\tau\right) - \sum_{\tau=1}^{k} \ell(\pi^*(x^t_\tau), y^t_\tau) \leq \mathcal{L}_{C,\alpha}(\pi^t, (\bar{x}^t, \bar{y}^t), \rho^t_{\mathcal{J}}) - \mathcal{L}_{C,\alpha}(\pi^*, (\bar{x}^t, \bar{y}^t), \rho^t_{\mathcal{J}}).$$

If $\rho^t_{\mathcal{J}} = null$, then the equality holds, so assume $\rho^t_{\mathcal{J}} = (\rho_1, \rho_2)$. In that case, because the pair was found by $\mathcal{J}_{\alpha+\epsilon}$, we have

$$\pi^t(x^t_{\rho_1}) - \pi^t(x^t_{\rho_2}) - d(x^t_{\rho_1}, x^t_{\rho_2}) - \alpha > \pi^t(x^t_{\rho_1}) - \pi^t(x^t_{\rho_2}) - d(x^t_{\rho_1}, x^t_{\rho_2}) - \alpha - \epsilon > 0$$

Then, we know

$$\begin{aligned}&\sum_{\tau=1}^{k} \ell\left(\pi^t\left(x^t_\tau\right), y^t_\tau\right) - \sum_{\tau=1}^{k} \ell(\pi^*(x^t_\tau), y^t_\tau) \\ &\leq \sum_{\tau=1}^{k} \ell\left(\pi^t\left(x^t_\tau\right), y^t_\tau\right) + C\left(\pi^t(x^t_{\rho_1}) - \pi^t(x^t_{\rho_2}) - d(x^t_{\rho_1}, x^t_{\rho_2}) - \alpha\right) - \sum_{\tau=1}^{k} \ell(\pi^*(x^t_\tau), y^t_\tau) \\ &\leq \mathcal{L}_{C,\alpha}(\pi^t, (\bar{x}^t, \bar{y}^t), \rho^t_{\mathcal{J}}) - \mathcal{L}_{C,\alpha}(\pi^*, (\bar{x}^t, \bar{y}^t), \rho^t_{\mathcal{J}}) \qquad (5)\end{aligned}$$

where (5) follows from Lemma A.1. □

A.2 Omitted Details from Subsection 3.2

Theorem A.2 (Arora et al. [2012]). *For a linear loss function $L^t(\pi) = L(\pi, z^t) \in [0, M]$, exponential weights with learning rate γ has the following guarantee: for any sequence of $(L^t)_{t=1}^T$ and for any other $\pi \in \Delta\mathcal{H}$,*

$$\sum_{t=1}^{T} L^t(\pi^t) \leq \sum_{t=1}^{T} L^t(\pi) + M \cdot \left(\frac{\ln(|\mathcal{H}|)}{\gamma} + \gamma T\right),$$

where π^0 is a uniform distribution over $\mathcal{H}$ and $\pi_h^{t+1} \propto \exp(-\gamma \cdot L^t(h)) \cdot \pi_h^t$ for each $h \in \mathcal{H}$.

Lemma 3.7. *For any sequence of $(\pi^t)_{t=1}^T$, $(z^t)_{t=1}^T \in \mathcal{Z}_{\text{FAIR-BATCH}}^T$, and $\pi^* \in \Delta\mathcal{H}$,*

$$\sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^t, z^t) - \sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^*, z^t) = \sum_{t=1}^{T}\sum_{\tau=1}^{k'} \ell(\pi^t(x_\tau'^t), y_\tau'^t) - \sum_{t=1}^{T}\sum_{\tau=1}^{k'} \ell(\pi^*(x_\tau'^t), y_\tau'^t)$$

Theorem 3.8. *For any sequence of $(z^t)_{t=1}^T \in \mathcal{Z}_{\text{FAIR-BATCH}}^T$ and $Q \subseteq \Delta\mathcal{H}$,*

$$\sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^t, z^t) - \sum_{t=1}^{T} \mathcal{L}_{C,\alpha}(\pi^*, z^t) \leq \mathbf{Regret}_{\text{BATCH}}^{Err}(\mathcal{A}, Q, T),$$

where $\pi^t = \mathcal{A}_{\text{BATCH}}\left((\pi^j, \bar{x}'^j, \bar{y}'^j)_{j=1}^{t-1}\right)$. In other words,

$$\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A}, Q_\alpha, T) \leq \mathbf{Regret}_{\text{BATCH}}^{Err}(\mathcal{A}, Q, T).$$

Proof. For any sequence of $(\bar{x}^t, \bar{y}^t)_{t=1}^T$, the definition of regret gives

$$\sum_{t=1}^{T}\sum_{\tau=1}^{k} \ell(\pi^t(x_\tau^t), y_\tau^t) - \sum_{t=1}^{T}\sum_{\tau=1}^{k} \ell(\pi^*(x_\tau^t), y_\tau^t) \leq \mathbf{Regret}_{\text{BATCH}}^{\text{Err}}(\mathcal{A}, Q, T),$$

where $\pi^t = \mathcal{A}_{\text{BATCH}}\left((\pi^j, \bar{x}^j, \bar{y}^j)_{j=1}^{t-1}\right)$.

Therefore, with Lemma 3.7, it follows that even for the sequence $(\bar{x}'^t, \bar{y}'^t)_{t=1}^T$, the difference in Lagrangian loss of π^t and π^* is bounded by $\mathbf{Regret}^{\text{Err}}(\mathcal{A}, \Delta\mathcal{H}, T)$. ☐

A.3 Using CONTEXT-FTPL from Syrgkanis et al. [2016]

Syrgkanis et al. [2016] considers an adversarial contextual learning setting where in each round t, the learner randomly deploys some policy ψ^t, and the environment chooses $(\xi^t, \bar{y}^t) \in \Xi \times \{0,1\}^k$, where k indicates the number of possible actions that can be taken for the instance ξ^t. The only knowledge at round t not available to the environment is the randomness over how the learner chooses ψ^t. And there's some linear loss $L(\psi^t(\xi), \bar{y}^t)$.

They show that in the small separator setting, they can achieve sublinear regret given that they can compute a separator set prior to learning. We first define the definition of a separator set and then state their regret guarantee.

Definition A.3. *A set $S = (\xi_1, \ldots, \xi_n)$ is called a* seperator set *for $\Psi : \Xi \to \{0,1\}^k$ if for any different ψ and ψ' in Ψ, there exists $\xi \in S$ such that $\psi(\xi) \neq \psi'(\xi)$.*

Theorem A.4 (Syrgkanis et al. [2016]). *For any sequence of $(\bar{x}^t, \bar{y}^t)_{t=1}^T$,* CONTEXT-FTPL$(S, \omega)$ *initialized with a separator set S and parameter ω achieves the following regret:*

$$\mathbb{E}\left[\sum_{t=1}^{T} L(\psi^t(\xi^t), \bar{y}^t) - \sum_{t=1}^{T} L(\psi^t(\xi^t), \bar{y}^t)\right] \leq 4\omega kn \sum_{t=1}^{T} \mathbb{E}[\left\|L(\cdot, \bar{y}^t)\right\|_*^2] + \frac{10}{\omega}\sqrt{nk}\log(|\mathcal{H}|),$$

where $n = |S|$, $\left\|L(\cdot, \bar{y}^t)\right\|_* = \max_{\hat{y}\in\{0,1\}^k} L(\hat{y}, \bar{y}^t)$, and the expectation is over the algorithm CONTEXT-FTPL.

Our online batch classification setting can be easily reduced to their setting by simply considering the batch of instances $\bar{x}^t$ as one single instance, meaning we set $\Xi = \mathcal{X}^k$. And we view each policy as $\psi_h(x^t) = (h(x_1^t), \ldots, h(x_k^t))$. In other words, we can define the policy class induced by $\mathcal{H}$ as

$$\Psi_{\mathcal{H}} = \left\{\forall h \in \mathcal{H} : (x_\tau)_{\tau=1}^k \mapsto (h(x_\tau))_{\tau=1}^k\right\}.$$

Therefore, the random policy $\psi_\pi \in \Delta\Psi_{\mathcal{H}}$ can be decomposed into convex combination of ψ_h's. Finally, we construct our linear loss function as $L\left((\pi(x_\tau))_{\tau=1}^k, \bar{y}\right) = \left\langle(\pi(x_\tau))_{\tau=1}^k, 1 - 2\bar{y}\right\rangle$. Note that $\mathbb{E}_{h\sim\pi}[L(\psi_h(\xi^t), \bar{y})] = L(\psi_\pi(\xi^t), \bar{y})$ by linearity of expectation.

Furthermore, we can turn any separator set $S \subseteq \mathcal{X}$ for $\mathcal{H}$ into an equal size separator set $S' \subseteq \Xi$ for Ψ. In fact, the construction is as follows:

$$S' = \{\forall x \in S : \xi_x = (x, v, \ldots, v)\},$$

where v is some arbitrary instance in $\mathcal{X}$.

Lemma A.5. *If S is the separator set for $\mathcal{H}$, then S' is a separator set for Ψ.*

Proof. Fix any h and h'. Note that by definition of S, there exists $x \in S$ such that $h(x) \neq h'(x)$. As a result, $\psi_h(\xi_x) \neq \psi_{h'}(\xi_x)$ as $(h(x), q, \ldots, q) \neq (h(x'), q, \ldots, q)$. □

Note that the reduction preserves the loss difference any π and π'.

Lemma A.6.

$$L(\psi_\pi(\bar{x}), \bar{y}) - L(\psi_{\pi'}(\bar{x}), \bar{y}) = \sum_{\tau=1}^{k} \ell(\pi(x_\tau^t), y_\tau^t) - \sum_{\tau=1}^{k} \ell(\pi'(x_\tau^t), y_\tau^t)$$

Proof.

$$\begin{aligned}
&L(\psi_\pi(\bar{x}), \bar{y}) - L(\psi_{\pi'}(\bar{x}), \bar{y}) \\
&= \left\langle(\pi(x_\tau))_{\tau=1}^k, 1 - 2\bar{y}\right\rangle - \left\langle(\pi'(x_\tau))_{\tau=1}^k, 1 - 2\bar{y}\right\rangle \\
&= \left(\sum_{\tau=1}^{k}(1 - \pi(x_\tau)) \cdot y_\tau + \pi(x_\tau) \cdot (1 - y_\tau)\right) - \left(\sum_{\tau=1}^{k}(1 - \pi'(x_\tau)) \cdot y_\tau + \pi(x_\tau) \cdot (1 - y_\tau)\right) \\
&= \sum_{\tau=1}^{k} \ell(\pi(x_\tau^t), y_\tau^t) - \sum_{\tau=1}^{k} \ell(\pi'(x_\tau^t), y_\tau^t)
\end{aligned}$$

□

Theorem A.7. *If the separator set S for $\mathcal{H}$ is of size s,* CONTEXT-FTPL *achieves the following regret in the online batch classification setting:*

$$\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell(\pi^t(x_\tau^t),y_\tau^t)-\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell(\pi^*(x_\tau^t),y_\tau^t)$$
$$\leq \mathbf{Regret}^{Err}_{\text{BATCH}}(\mathcal{A},Q,T)=O\left((ks)^{\frac{3}{4}}\sqrt{T\log(|\mathcal{H}|)}\right)$$

Proof. Since we can construct a separator set S' for Ψ (Lemma A.6), we can run CONTEXT-FTPL(S',ω) with ω appropriately set to achieve the above regret (Theorem A.4). □

Using CONTEXT-FTPL as a black box for $\mathcal{A}_{\text{BATCH}}$ to solve the online fair batch classification, we get the following regret for the misclassification and fairness loss regret.

Remark A.8. *Note that in each round t,* CONTEXT-FTPL *simply returns a base policy ψ_h such that single h will be used to classify everyone in the batch $\bar{x}^t$. However, in some cases, it might be more desirable to deploy a randomized policy π such that h is sampled from π for each individual x_τ^t in the batch. In such cases, one can either run* CONTEXT-FTPL *for each x_τ^t or run* CONTEXT-FTPL *multiple times to form a uniform mixture of h's that approximates π.*

Theorem 3.9. *If the separator set S for $\mathcal{H}$ is of size s, then* CONTEXT-FTPL *achieves the following misclassification and fairness regret in the online fair batch classification setting.*

$$\mathbf{Regret}^{Err}_{\text{FAIR-BATCH}}(\mathcal{A},Q_\alpha,T)\leq O\left(\left(\frac{sk}{\epsilon}\right)^{\frac{3}{4}}\sqrt{T\log(|\mathcal{H}|)}\right)$$

$$\sum_{t=1}^{T}\mathit{Unfair}_{\alpha+\epsilon}(\pi^t,z^t)\leq O\left(\left(\frac{sk}{\epsilon}\right)^{\frac{3}{4}}\sqrt{T\log(|\mathcal{H}|)}\right)$$

Proof. The proof follows from Theorem A.7 along with the fact that the batch size is $k+2C$, where $C\geq\frac{k+1}{\epsilon}$. □

B Omitted Details from Section 4

Tightness of Bound in Observation 4.7

Consider the following example: $\mathcal{X}=\{x_1,x_2,x_2\}$, $\mathcal{D}|_\mathcal{X}=\mathbb{U}\{\mathcal{X}\}$ (i.e. uniform distribution over $\mathcal{X}$), $\alpha'=0.1$. $\mathcal{H}=\{h_1,h_2\}$ given by:

$$h_1(x_1)=1,\ h_1(x_2)=0,\ h_1(x_3)=0$$
$$h_2(x_1)=1,\ h_2(x_2)=1,\ h_2(x_3)=0$$

Also, assume the dissimilarity measure by the auditor is:

$$d(x_1,x_2)=0,\ d(x_2,x_3)=0,\ d(x_3,x_1)=1.$$

Assume the algorithm deploys policies in the following manner: $\pi^t=\begin{cases}h_1 & t\text{ is odd}\\ h_2 & t\text{ is even}\end{cases}$

Both h_1,h_2 are exactly $(\alpha',\frac{1}{8})-fair$. If T is even, π^{avg} is exactly $(\alpha',\frac{1}{4})$-fair.

Theorem 4.2 (Accuracy Generalization)**.** *With probabilty $1-\delta$, the misclassification loss of π^{avg} is upper bounded by*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi^{avg}(x),y)] \leq \inf_{\pi\in Q_\alpha} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi(x),y)] + \frac{1}{kT}\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A},Q_\alpha,T) + \sqrt{\frac{8\ln\left(\frac{4}{\delta}\right)}{T}}$$

Proof. By Theorem 3.5, for any sequence of $(z^t)_{t=1}^T$ and $\pi^* \in Q_\alpha$,

$$\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell\left(\pi^t\left(x_\tau^t\right),y_\tau^t\right) - \ell(\pi^*(x_\tau^t),y_\tau^t) \leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A},Q_\alpha,T)$$

We select $\pi^* \in \operatorname*{argmin}_{\pi\in Q^\alpha} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(\pi(x),y)]$. Also, fix the sequence of $(z^t)_{t=1}^T$ and the random coin flips of the algorithm $\mathcal{A}$. Therefore, $(\pi^t)_{t=1}^T$ must be fixed as well.

Lemma B.1.

$$\Pr_{\substack{(\bar{x}^t,\bar{y}^t)_{t=1}^T\\ \pi^t=\mathcal{A}(\cdots)}}\left[\left|\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell\left(\pi^t\left(x_\tau^t\right),y_\tau^t\right) - \mathop{\mathbb{E}}_{(\bar{x}'^t,\bar{y}'^t)_{t=1}^T}\left[\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell\left(\pi^t\left({x'}_\tau^t\right),{y'}_\tau^t\right)\right]\right| \geq \gamma\right] \leq 2\exp\left(\frac{-\gamma^2}{2k^2T}\right)$$

Proof. Define $\left(A^t\right)_{t=1}^T$ as

$$A^t = \sum_{j=1}^{t}\sum_{\tau=1}^{k}\ell\left(\pi^j\left(x_\tau^j\right),y_\tau^j\right) - \mathop{\mathbb{E}}_{(\bar{x}'^j,\bar{y}'^j)_{j=1}^t}\left[\sum_{j=1}^{t}\sum_{\tau=1}^{k}\ell\left(\pi^j\left({x'}_\tau^j\right),{y'}_\tau^j\right)\right].$$

Note that $\left(A^t\right)_{t=1}^T$ is a martingale as $\mathbb{E}[A^t|A^1,\ldots,A^{t-1}] = A^{t-1}$ because π^t is determined deterministically in terms of the previous history.

Note that $|A^t - A^{t-1}| \leq k$. Therefore, applying Azuma's inequality yields

$$\Pr\left[|A^T - A^1| \geq \gamma\right] \leq 2\exp\left(\frac{-\gamma^2}{2k^2T}\right).$$

□

Applying the Chernoff bound, we get a similar concentration bound on π^*:

$$\Pr_{(\bar{x}^t,\bar{y}^t)_{t=1}^T}\left[\left|\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell\left(\pi^*\left(x_\tau^t\right),y_\tau^t\right) - \mathop{\mathbb{E}}_{(\bar{x}'^t,\bar{y}'^t)_{t=1}^T}\left[\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell\left(\pi^*\left({x'}_\tau^t\right),{y'}_\tau^t\right)\right]\right| \geq \gamma\right] \leq 2\exp\left(\frac{-\gamma^2}{2k^2T}\right).$$

Next, using triangle inequality, with probability $1-\delta$, it holds that

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\bar{x}^t,\bar{y}^t)_{t=1}^T}\left[\sum_{t=2}^{T}\sum_{\tau=1}^{k}\ell\left(\pi^t\left(x_\tau^t\right),y_\tau^t\right)\right] - \mathop{\mathbb{E}}_{(\bar{x}^t,\bar{y}^t)_{t=1}^T}\left[\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell(\pi^*(x_\tau^t),y_\tau^t)\right]\\ &\leq \mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}(\mathcal{A},Q_\alpha,T) + 2\sqrt{\ln\left(\frac{4}{\delta}\right)2k^2T}\end{aligned}$$

Equivalently, with probability $1-\delta$,

$$\mathop{\mathbb{E}}_{(\bar{x}^t,\bar{y}^t)_{t=1}^T}\left[\frac{1}{kT}\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell\left(\pi^t\left(x_\tau^t\right),y_\tau^t\right)\right]\leq\min_{\pi\in Q_\alpha}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\ell(\pi(x),y)\right]$$
$$+\frac{1}{kT}\mathbf{Regret}^{C,\alpha,\mathcal{J}_{\alpha+\epsilon}}\left(\mathcal{A},Q_\alpha,T\right)+2\sqrt{\frac{2\ln\left(\frac{4}{\delta}\right)}{T}}$$

For the left hand side on the inequality, observe that the following holds:

$$\begin{aligned}
\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\ell(\pi^{avg}(x),y)\right] &= \frac{1}{kT}\sum_{\tau=1}^{k}\sum_{i=1}^{T}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\mathop{\mathbb{E}}_{\pi\sim\mathbb{U}\{\pi^1,\ldots,\pi^T\}}\left[\ell(\pi(x),y)\right]\right] \qquad (6)\\
&= \frac{1}{kT}\sum_{\tau=1}^{k}\sum_{i=1}^{T}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\frac{1}{T}\sum_{t=1}^{T}\ell(\pi^t(x),y)\right]\\
&= \frac{1}{kT}\sum_{\tau=1}^{k}\sum_{t=1}^{T}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\frac{1}{T}\sum_{i=1}^{T}\ell(\pi^t(x),y)\right]\\
&= \frac{1}{kT}\sum_{\tau=1}^{k}\sum_{t=1}^{T}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\ell(\pi^t(x),y)\right]\\
&= \frac{1}{kT}\mathop{\mathbb{E}}_{(\bar{x}^t,\bar{y}^t)_{t=1}^T\sim_{i.i.d.}\mathcal{D}^{kT}}\left[\sum_{t=1}^{T}\sum_{\tau=1}^{k}\ell(\pi^t(x_\tau^t),y_\tau^t)\right]
\end{aligned}$$

Transition 6 stems from the fact that our misclassification loss ℓ is linear with respect to the base classifiers in $\mathcal{H}$. Hence, taking the uniform distribution over $\pi^1,\ldots,\pi^T$ gives

$$\ell(\pi^{avg}(x),y)=\mathop{\mathbb{E}}_{\pi\sim\mathbb{U}\{\pi^1,\ldots,\pi^T\}}\left[\ell(\pi(x),y)\right].$$

□