Fairness-aware Federated Learning: A Robust Approach

Anonymous Authors[1]

Abstract

Federated Learning is a machine learning technique where a network of clients collaborate with a server to learn a centralized model while keeping data localized. In such a setting, the training data is often statistically heterogeneous across the clients, which introduces bias in the training objective and degrades the performance of learned model. To address this issue, we propose a **R**obust **C**lient-weighted **F**ederated **L**earning (RC-FL) framework in which the goal is to minimize the worst-case weighted client losses over an uncertainty set. By deriving a variational representation, we show that RC-FL is a fairness-aware objective and can be easily optimized by solving a joint minimization problem over the model parameters and a dual variable. We then propose an optimization algorithm to solve RC-FL which can be efficiently implemented in a federated setting and provide convergence guarantees. We further prove generalization bounds for learning with this objective. Experiments on real-world datasets demonstrate the effectiveness of our framework in achieving both accuracy and fairness.

1. Introduction

Due to the emergence of unprecedented amount of data generated by mobile devices and the growing computational power of these devices, Federated Learning (FL) has becoming of increasing importance and often crucial for deployment of large-scale machine learning (Konen et al., 2016; McMahan et al., 2017). A typical Federated Learning setting consists of a network of hundreds to millions of devices (clients) which interact with each other through a central server, and its goal is to collaboratively learn a shared model while keeping the training data on the device instead of requiring the data to be uploaded and stored on the central server.

Despite its advantage of data privacy, it faces several challenges ranging from developing communication efficient algorithms to ensuring fairness (Kairouz et al., 2019). First, frequent communication is undesirable in FL as it is expensive due to unreliable and relatively slow network connection, especially when more clients are involved. To reduce communication overload, one needs to depart from the conventional distributed learning setting where the updated local models are broadcast to the central server at each iteration, and adopt more efficient communication strategies like periodic averaging (Khaled et al., 2019; Haddadpour & Mahdavi, 2019; Stich, 2019; Konečnỳ et al., 2016; Konen et al., 2016; McMahan et al., 2017).

Another major challenge in Federated Learning is the statistical heterogeneity (non-i.i.d.) of the training data, that is, the data generated or stored by each client follows a different distribution. Thus, the model which is trained on the union of these non-i.i.d. data can not generalize well to each client (Li et al., 2018; Wang et al., 2020; Reisizadeh et al., 2020). More importantly, the heterogeneity of client objectives poses a critical question. Which objective function is Federated Learning seeking to optimize? It is no longer clear that all data should be treated equally. Mohri et al. (2019) argue that the uniform distribution is not the natural target distribution and that minimizing with respect to that specific distribution is risky. To address the mismatch issue, the authors propose an agnostic federated learning framework, where the centralized model is optimized for the worst-case performance across all clients. Despite that their framework provides an approach to dealing with client heterogeneity, the proposed notion of agnostic minimax loss may be overly conservative in the sense that it focuses only on one single client with largest loss and thus causes very pessimistic performance to other clients.

Ensuring that the learned models are non-discriminatory or fair with respect to some protected groups is a topical problem in modern machine learning, and a variety of definitions of the notion of fairness have been proposed (Zafar et al., 2017; Dwork et al., 2018; Donini et al., 2018; Williamson & Menon, 2019). However in the context of federated learning, there has been little work on how to address the fairness concerns. Mohri et al. (2019) have taken a step towards

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

this goal by introducing good-intent fairness based on the maximin principle where the objective is to seek all client losses to be small. However as pointed out by Li et al. (2020a), that objective is rigid as it does not allow for flexible trade-off between fairness and accuracy. Inspired by fair resource allocation in wireless network, the authors propose a modified federated learning objective to encourage uniformity in performance across devices. Despite that their objective enables to tune the amount of fairness via a single hyper-parameter, it is not a fairness-aware objective, and also fails to take into account client heterogeneity and thus would be less effective in ensuring better fairness (Yang et al., 2020). Besides, there is no convergence guarantee for their proposed algorithm.

In this work, we propose a new framework called RC-FL, a **R**obust **C**lient-weighted **F**ederated **L**earning framework, to address the heterogeneity and fairness issues in federated learning. Instead of optimizing the model for a specific (uniform) distribution, RC-FL minimizes a Q_α-weighted loss which is a supremum of weighted aggregation of client losses over an uncertainty set Q_a of possible weights, where the parameter $\alpha := (\alpha_1, \cdots, \alpha_n)$ is personalized for each client to account for client heterogeneity. Compared to agnostic federated learning (Mohri et al., 2019), RC-FL is more flexible as the conservation level can be controlled by adjusting those parameters α_is. Agnostic loss and the standard federated learning objective can in fact be recovered from our framework using proper choice of α_i. We would like to mention that a special case (all α_is are equal) of our RC-FL has been considered independently in another work (Laguel et al., 2020) and studied numerically. In this work, we focus on developing theoretical analysis regarding a more general framework, where we propose a completely different but provably convergent algorithm and provide generalization guarantees.

RC-FL formulates the learning problem as a minimax optimization problem, which finds a global model that minimizes the worst-case weighted aggregated loss. One approach to solving this minimax problem is to employ methods from Mohri et al. (2019) which iteratively applies stochastic gradient descent ascent updates. However this approach is undesirable in federated learning setting since it requires communication at each iteration. The key advantage of RC-FL is that it enjoys a variational representation which is equivalent to a minimization problem over a dual variable. Therefore RC-FL can readily be optimized by solving a joint minimization problem with respect to the model parameter and the dual variable. We propose a simple gradient based algorithm called `rFedFair` to solve RC-FL which can be efficiently implemented in a federated setting and comes with strong theoretical guarantees.

Our RC-FL framework also has another desirable benefit. It defines a notion of fairness, which we refer to as heterogeneous conditional value at risk (HCVaR). HCVaR is a generalization of conditional value at risk (CVaR) which is a well-studied risk-averse measure in finance and portfolio optimization (Shapiro et al., 2014; Rockafellar et al., 2000; Krokhmal et al., 2002) and has recently been used in many applications in machine learning (Chow & Ghavamzadeh, 2014; Shalev-Shwartz & Wexler, 2016; Fan et al., 2017; Curi et al., 2020; Lee et al., 2020; Soma & Yoshida, 2020; Jeong & Namkoong, 2020). In particular, Williamson & Menon (2019) propose a new definition of fairness and show that CVaR is a fairness risk measure. Compared to CVaR, HCVaR takes into account client heterogeneity by allowing different parameters α_i for each client i, which is more related to federated learning setting. The connection to HCVaR shows that RC-FL is a fairness-aware objective which involves an expectation and deviation, implying that minimizing RC-FL objective ensures that the client losses are small, and that they have low deviation (fairness). We summarize our contributions as follow.

- We present a new framework called RC-FL to address the heterogeneity and fairness issues in federated learning, which generalizes many existing federated learning objectives, including agnostic loss (Mohri et al., 2019) and standard FL objective, and naturally yields a new notion of fairness named HCVaR.

- We propose a smooth approximation to RC-FL and provide an efficient algorithm to solve it which is guaranteed to find an approximate minimizer of the original RC-FL problem for convex and smooth loss functions.

- We prove two data-dependent generalization bounds for learning with RC-FL. Our bounds show proper generalization from empirical distribution of samples to the true underlying distribution.

The rest of the paper is organized as follows. In Section 2, we establish the necessary notations and provide a brief background on federated learning. In Section 3, we give a formal definition of our RC-FL framework and describe its connection to fairness. Then in Section 4, we present an efficient federated learning algorithm for solving RC-FL. Next, we give a detailed theoretical analysis of the proposed algorithm in both full and partial device participation cases (Section 5.1), as well as generalization guarantees (Section 5.2). In Section 6, we conduct a series of experiments and compare our results with existing fair federated learning algorithms. Finally we conclude and discuss future directions. All proofs are deferred to the appendix.

2. Preliminaries

Notation: We denote by $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ a measurable instance space where $\mathcal{X}$ and $\mathcal{Y}$ represent feature and label spaces, respectively. We use $\mathcal{F} = \{f_\omega : \omega \in \mathcal{W}\}$ to denote the underlying hypothesis class of functions from $\mathcal{X}$ to $\mathcal{Y}'$ where $\mathcal{Y}'$ might differ from $\mathcal{Y}$. We are also given a loss function $l : \mathcal{Y}' \times \mathcal{Y} \to \mathbb{R}_+$, quantifying the loss incurred by a decision rule applied to a data instance $z = (x, y) \in \mathcal{Z}$, e.g., $l(f_\omega(x), y)$. Given a hypothesis $f_\omega \in \mathcal{F}$, denote the expected loss of f_ω with respect to a distribution P over $\mathcal{X} \times \mathcal{Y}$ by

$$f^P(\omega) := \mathbb{E}_{(x,y)\sim P}[l(f_\omega(x), y)].$$

Federated Learning Scenario: We consider a federated learning setting with a network of n nodes (clients) connected to a server node. Denote $[n] = \{1, \cdots, n\}$. We assume that for every $i \in [n]$ the i-th client has access to m_i training sample in $S^i = \{(x_j^i, y_j^i) \in \mathcal{X} \times \mathcal{Y} : 1 \le j \le m_i\}$ drawn i.i.d. from some unknown distribution P_i, i.e., $(x_j^i, y_j^i) \sim P_i$. In federated learning, the data on a given client is typically on the usage of the mobile device by a particular user, which might come from different environments, contexts, and applications, and hence clients can have non-i.i.d. data distributions, that is, the distributions P_i and P_j, $i \neq j$, are distinct. Let $m = \sum_{i=1}^n m_i$ and $p_i = m_i/m$. We will denote by $\hat{P}_i$ the empirical distribution associated to sample S^i. In the conventional federated learning setting, the n clients are interested in collaboratively training a single model on their joint data in a privacy-preserving way by solving the following problem

$$\min_{\omega\in\mathcal{W}} \sum_{i=1}^n p_i \frac{1}{m_i} \sum_{j=1}^{m_i} l(f_\omega(x_j^i), y_j^i)$$

with the assumption that all samples are uniformly weighted, i.e., the underlying target distribution is $\sum_{i=1}^n p_i P_i$.

However since the mixture weight of the distribution P_i, $i \in [n]$ is unknown, that assumption is rather restrictive and can lead to solutions that are harmful to the clients (Mohri et al., 2019). Moreover, the uniformly weighted aggregated loss puts less weight on clients with small number of data points during training, thus giving rise to unfairness where the learned model behaves differently across clients. To address these issues, a natural idea is to reweight the client loss. However since we do not understand precisely which weighting to pick, we propose to study a robust version of client weighted loss, which defines our new framework given in the next section.

3. Robust Client-weighted Federated Learning

In this section, we first introduce the robust client-weighted federated learning framework we consider. Then, we establish its connection to fairness.

3.1. Problem Formulation

As we stated in previous section, the conventional federated learning objective raises some issues. This motivates us to consider a **R**obust **C**lient-weighted **F**ederated **L**earning (RC-FL) framework where different weights are assigned to different clients, and the learner must learn a model that is favorable for any weighted aggregation of client losses over an uncertainty set Q of possible weights.

The benefits of RC-FL are in two folds. First, it allows us to upweight a underrepresent client to achieve better performance and thus improve model fairness. We will show later how this intimately relates to a notion of fairness. Second, RC-FL offers performance guarantees for any weighted client losses over an uncertainty set which might include the true unknown target distribution. We now formally define the robust client-weighted federated learning framework. Throughout this paper, for ease of notation, we use $\mathbb{E}[\cdot]$ to denote the expectation with respect to the randomness in selecting client i with probability p_i unless explicitly stated otherwise.

Definition 1 (RC-FL). *Let $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ be a vector where $\alpha_i \in [p_i, 1]$ for all i. Define the uncertainty set $Q_\alpha = \{q = (q_1, \cdots, q_n) : \mathbb{E}[q_i] := \sum_{i=1}^n q_i p_i = 1, q_i \in [0, \alpha_i^{-1}]\}$. Then, the Q_α-weighted loss is*

$$f_\alpha(\omega) := \sup_{q\in Q_\alpha} \mathbb{E}[q_i f^{P_i}(\omega)] = \sup_{q\in Q_\alpha} \sum_{i=1}^n q_i p_i f^i(\omega), \quad (1)$$

where we write $f^i(\omega) = f^{P_i}(\omega)$ for notational convenience.

Here, $\alpha_i \in [p_i, 1]$ is a tuning parameter and allowed to be different across the clients to take client heterogeneity into account. Interestingly, by setting different α, the Q-weighted loss can recover existing federated learning objectives. For example, as $\alpha_i \to 1 \forall i$, the uncertainty set Q_α would reduce to a single point, i.e., $Q_\alpha = \{(1, \cdots, 1)\}$, and f_α becomes

$$\sum_{i=1}^n p_i f^i(\omega),$$

which is the classical federated learning objective; as $\alpha_i \to p_i$ for all i, f_α reduces to the agnostic federated learning loss (AFL) (Mohri et al., 2019)

$$\max_{\lambda\in\Delta_n} \sum_{i=1}^n \lambda_i f^i(\omega),$$

where Δ_n is a simplex. Therefore, our RC-FL objective is more flexible as it can be tuned based on the conservatism level α_i of each client.

3.2. Connection to Fairness

In this section, we show that RC-FL defines a notion of fairness named heterogeneous conditional value at risk (HCVaR), which is a generalization of conditional value at risk (CVaR), a common risk measure in mathematical finance and has recently been proposed as a fairness risk measure (Williamson & Menon, 2019). We first recall the definition of CVaR. For scalar $\chi \in (0, 1]$ and random variable $f^i(\omega)$ (the randomness is w.r.t. the selection of client), the conditional value at risk is (Rockafellar et al., 2000)

$$\text{CVaR}_{1-\chi}(f^i(\omega)) = \mathbb{E}[f^i(\omega)|f^i(\omega) > \mathbb{Q}_{1-\chi}(f^i(\omega))],$$

where $\mathbb{Q}_{1-\chi}$ is the quantile at level $1-\chi$. Intuitively, CVaR measures the tail behavior of $f^i(\omega)$. Note that the good-intent fairness (AFL) is a special case of CVaR fairness risk (Mohri et al., 2019). In federated learning setting, because of client heterogeneity, we may wish to treat losses arising from different clients differently. Therefore, we consider a heterogeneous version of CVaR by allowing different weights to each client as follows.

Definition 2 (HCVaR)**.** *Given a vector $\alpha = (\alpha_1, \cdots, \alpha_n)$ satisfying $\tau_\alpha := \mathbb{E}[\alpha_i^{-1}] \geq 1$, we define heterogeneous conditional value at risk as*

$$\textit{HCVaR}_{1-\alpha}(f^i(\omega)) := \mathbb{E}_\alpha[f^i(\omega)|f^i(\omega) > \mathbb{Q}_{1-1/\tau_\alpha}(f^i(\omega))],$$

where the expectation $\mathbb{E}_\alpha[\cdot]$ is with respect to random selection of device i with probability $\frac{p_i}{\alpha_i \tau_\alpha}$.

Remark 1. If $\alpha_i = \chi$ for all i, $\text{HCVaR}_{1-\alpha}(f^i(\omega))$ would reduce to $\text{CVaR}_{1-\chi}(f^i(\omega))$.

Compare to CVaR, HCVaR measures a weighted tail-average. Therefore, we can define a notion of fairness by minimizing HCVaR which seeks that the weighted average of the largest client losses is small. This tightens the range of client losses, thus ensuring that the client losses are commensurate (fair).

With this definition in mind, we now derive a dual representation for RC-FL, reformulating the primal problem (1) over $q \in Q_\alpha$ to a dual problem over a one-dimensional variable. This dual representation shows an equivalence between Q-weighted loss and HCVaR, thus connecting RC-FL to fairness.

Lemma 1. *Denote $(\cdot)_+ := \max(\cdot, 0)$. Then,*

$$f_\alpha(\omega) = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \mathbb{E}\left[\frac{1}{\alpha_i}(f^i(\omega) - \eta)_+\right] \right\} = \textit{HCVaR}_{1-\alpha}(f^i(\omega)) \tag{2}$$

Remark 2. If the loss function is bounded, i.e., $0 \leq l(f_\omega(x), y) \leq B$ for any $z = (x, y) \in \mathcal{Z}$, the domain of η in (2) can be restricted to $\eta \in [0, B]$.

By the lemma, one may equally write $f_\alpha(\omega) = \mathbb{E}_\alpha[f^i(\omega)] + \mathcal{D}_\alpha(f^i(\omega))$ where $\mathcal{D}_\alpha(f^i(\omega)) := \text{HCVaR}_{1-\alpha}(f^i(\omega) - \mathbb{E}_a[f^i(\omega)])$ is a measure of deviation. For perfect fairness where $f^i(\omega)$ is a constant, $\mathcal{D}_\alpha(f^i(\omega)) = 0$. Therefore, the lemma shows that RC-FL is a fairness-aware objective that is an expectation plus a deviance, suggesting that minimizing RC-FL objective ensures that the client losses are small, and that they have low deviation (fairness). By changing the parameters α_is, RC-FL also allows for a flexible trade-off between average accuracy and fairness. There is another desirable side benefit. The convexity of HCVaR implies that if $\omega \to f^i(\omega)$ is convex, then so is $\omega \to f_\alpha(\omega)$. Thus, for convex l and $\mathcal{F}$, as shown in the next section, solving RC-FL (simultaneously encouraging fairness) does not pose an optimization burden.

In practice, the data-generating distribution P_i is not known to the client, and the client has only access to the finite sample S^i. Thus, for every $i \in [n]$, the expected loss can be estimated by the empirical loss $\hat{f}^i(\omega) = \frac{1}{m_i} \sum_{j=1}^{m_i} l(f_\omega(x_j^i), y_j^i)$. This leads to the definition of empirical Q_α-weighted loss,

$$\hat{f}_\alpha(\omega) := \sup_{q \in Q_\alpha} \mathbb{E}[q_i \hat{f}^i(\omega)] = \sup_{q \in Q_\alpha} \sum_{i=1}^{n} q_i p_i \hat{f}^i(\omega). \tag{3}$$

4. The Proposed Algorithm

To solve RC-FL, one may propose to directly minimize the Q_α-weighted loss, which yields a minimax optimization problem, by applying stochastic gradient descent ascent algorithm as in Mohri et al. (2019). However this approach may be undesirable in federated learning setting as it requires frequent communication. In this section, we will present a gradient optimization method for solving RC-FL problem (3) that is computationally and communication-wise efficient.

Instead of solving the original Q_α-weighted loss (3), we aim to minimize its dual representation, which yields the following joint optimization problem

$$\min_{\omega \in \mathcal{W}} \hat{f}_\alpha(\omega) = \min_{\omega \in \mathcal{W}, \eta \in \mathbb{R}} \mathbb{E}\left[\frac{1}{\alpha_i}(\hat{f}^i(\omega) - \eta)_+ + \eta\right] \triangleq \hat{F}_\alpha(\omega, \eta), \tag{4}$$

where we rewrite $\hat{f}_\alpha$ as its dual representation given by Lemma 1. For convex $\hat{f}^i(\omega)$, Problem (4) is jointly convex in (ω, η) and thus can be solved by gradient-based optimization method. However, because of the non-smoothness and non-linearity of $(\cdot)_+$, Problem (4) is not differential, and the gradient of the clients whose loss is less than η would

Algorithm 1 `rFedFair`

Input: $\{\omega_0^i = \omega_0, \eta_0^i = \eta_0\}, \beta, \kappa, T$
for $t = 0$ **to** $T-1$ **do**
  for $i = 1$ **to** n **do**
    if t does not divide κ **then**
      $\omega_{t+1}^i = \omega_t^i - \beta \nabla_\omega \hat{f}^{\mu,i}(\omega_t^i, \eta_t^i)$
      $\eta_{t+1}^i = \eta_t^i - \beta \nabla_\eta \hat{f}^{\mu,i}(\omega_t^i, \eta_t^i)$
    else
      server chooses a set of clients Z (deterministic or random), and each of the selected client $j \in Z$ uploads to server:
      $\omega_t^j - \beta \nabla_\omega \hat{f}^{\mu,j}(\omega_t^j, \eta_t^j)$
      $\eta_t^j - \beta \nabla_\eta \hat{f}^{\mu,j}(\omega_t^j, \eta_t^j)$
      server computes the average of the received models and sends to all clients i:
      $\omega_{t+1}^i = \text{Average}(\{\omega_t^j - \beta \nabla_\omega \hat{f}^{\mu,j}(\omega_t^j, \eta_t^j)\}_{j \in Z})$
      $\eta_{t+1}^i = \text{Average}(\{\eta_t^j - \beta \nabla_\eta \hat{f}^{\mu,j}(\omega_t^j, \eta_t^j)\}_{j \in Z})$
    end if
  end for
end for
Output: $\tilde{\omega}_T := \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i=1}^n p_i \omega_t^i$

get truncated to zero, which makes gradient optimization of Problem (4) extremely challenging. To overcome this difficulty, we propose to use softmax function as a smooth approximation to the max function defined as follows.

$$\phi_\mu(x) := \mu \log(1 + e^{\frac{x}{\mu}}),$$

where μ is a predefined parameter. Note that ϕ_μ is convex and $1/\mu$-smooth. Furthermore, it smoothly approximates the max function (Bullins, 2020). This gives us a natural smooth approximation to Problem (4), namely

$$\min_{\omega \in \mathcal{W}} \hat{f}_\alpha^\mu(\omega) = \min_{\omega \in \mathcal{W}, \eta \in \mathbb{R}} \mathbb{E}\big[\hat{f}^{\mu,i}(\omega, \eta)\big] \triangleq \hat{F}_\alpha^\mu(\omega, \eta), \quad (5)$$

where $\hat{f}^{\mu,i}(\omega, \eta) := \frac{1}{\alpha_i} \phi_\mu(\hat{f}^i(\omega) - \eta) + \eta$. We can prove that $\hat{F}_\alpha(\omega, \eta)$ and $\hat{F}_\alpha^\mu(\omega, \eta)$ satisfy the following inequality.

Lemma 2. *For any ω, η,*

$$\hat{F}_\alpha(\omega, \eta) \leq \hat{F}_\alpha^\mu(\omega, \eta) \leq \hat{F}_\alpha(\omega, \eta) + \mu \tau_\alpha.$$

Lemma 2 shows that Problem (5) smoothly approximates the original non-smooth Problem (4), implying that we can solve the original Problem (4) by solving its smoothed version, which will be proven in next section. Now we propose an algorithm for solving Problem (5), called `rFedFair`. As summarized in Algorithm 1, in each iteration t of local updates, each client i updates its local model (ω_t^i, η_t^i) via a gradient descent step based on its own loss function $\hat{f}^{\mu,i}$. After κ local iterations, the server performs averaging step over the local models received from a selected set Z of clients. The averaged model is then sent back to all clients to begin the next round of local iterations with this fresh initialization. Compared to conventional federated learning algorithms like FedAvg (McMahan et al., 2017), the local update of client i using gradient descent is with respect to $\hat{f}^{\mu,i}$ instead of the empirical loss $\hat{f}^i$, and the client needs to optimize over model parameter ω and dual variable η jointly. In practice, the selection and averaging method may vary. Here, we consider the following two strategies for picking a set of clients and doing model averaging.

Full Participation: In an idealized scenario, each client participates in each round of the communication. So the server chooses $Z = [n]$, and the averaging step performs

$$\begin{aligned} \omega_{t+1}^i &= \textstyle\sum_{i=1}^n p_i(\omega_t^i - \beta \nabla_\omega \hat{f}^{\mu,i}(\omega_t^i, \eta_t^i)) \\ \eta_{t+1}^i &= \textstyle\sum_{i=1}^n p_i(\eta_t^i - \beta \nabla_\eta \hat{f}^{\mu,i}(\omega_t^i, \eta_t^i)) \end{aligned}.$$

However in practice, especially when the total number of clients is huge, the clients participating in a round of communication are expected to fail or drop out because of broken network connection or limited client availability, or there may be straggler clients, which take much longer time to send their output than other clients in the same round. Therefore, it might be unrealistic to assume that the server collects all client updates.

Partial Participation: A more practical strategy is to sample a subset of clients. To pick a subset of clients at communication step, we use the sampling scheme (Li et al., 2020b) where server chooses a subset of clients $Z \subseteq [n]$ with size $K < n$ uniformly at random with replacement according to the sampling probabilities $p_1, p_2, \cdots, p_n$. Then, the server performs averaging step as follows.

$$\begin{aligned} \omega_{t+1}^i &= \frac{1}{K} \textstyle\sum_{j \in Z}(\omega_t^j - \beta \nabla_\omega \hat{f}^{\mu,j}(\omega_t^j, \eta_t^j)) \\ \eta_{t+1}^i &= \frac{1}{K} \textstyle\sum_{j \in Z}(\eta_t^j - \beta \nabla_\eta \hat{f}^{\mu,j}(\omega_t^j, \eta_t^j)) \end{aligned}.$$

Note that our algorithm significantly reduces the number of communications as the local model of clients are aggregated periodically.

5. Theoretical Results

In this section, we establish our main theoretical results. We first show that Algorithm 1 converges to the global minimum of the original non-smooth problem (4) for convex and smooth losses in both full and partial participation cases. Next, we prove that the returned solution will properly generalize from training data to unseen test samples.

5.1. Convergence Analysis

We first provide convergence guarantees for full participation and then extend the result to partial participation.

Before introducing the convergence result, we define a few notations. Let $\nabla_\omega f(\omega)$ be the derivative of a function $f(\omega)$ with respect to ω. The dot product between two vectors ω and ω' is denoted by $<\omega, \omega'>$, and the norm of a vector is represented by $||\cdot||$. We also define $\hat{f}^*_\alpha := \min_{\omega\in\mathcal{W}} \hat{f}_\alpha(\omega)$ and $(\omega^*, \eta^*) := \arg\min_{\omega,\eta} \hat{F}^\mu_\alpha(\omega, \eta)$. We make the following standard assumptions on the loss functions.

Assumption 1. *For every client $i \in [n]$, the empirical loss $\hat{f}^i(\omega)$ is L_1-smooth and convex. That is, for any ω, ω', we have*

$$\hat{f}^i(\omega') \leq \hat{f}^i(\omega) + <\nabla_\omega \hat{f}^i(\omega), \omega' - \omega> + \frac{L_1}{2}||\omega' - \omega||^2$$
$$\hat{f}^i(\omega') \geq \hat{f}^i(\omega) + <\nabla_\omega \hat{f}^i(\omega), \omega' - \omega>$$

Assumption 2. *The empirical loss $\hat{f}^i(\omega)$ is L_2-Lipschitz, i.e., for any ω, ω', we have $|\hat{f}^i(\omega) - \hat{f}^i(\omega')| \leq L_2||\omega - \omega'||$.*

Note that Assumption 2 is only used to prove the smoothness of $\hat{f}^{\mu,i}(\omega, \eta)$, and we will not use it to quantify the degree of client heterogeneity. Instead, we consider the following notion of dissimilarity of client data distribution introduced by Khaled et al. (2019).

Quantifying the degree of client heterogeneity. We use $\rho^2 := \sum_{i=1}^n p_i ||\nabla_{\omega,\eta} \hat{f}^{\mu,i}(\omega^*, \eta^*)||^2$ for measuring the degree of client heterogeneity. Note that ρ is always finite and in case that the client data is actually i.i.d. and all α_is are equal, $\rho = 0$ as it is expected.

5.1.1. Full Participation

We now present the convergence of `rFedFair` with full participation.

Theorem 1. *Under the assumptions, if we choose the learning rate β such that $\beta \leq \frac{1}{40L}$ and $6L^2\beta^2(\kappa-1)^2 \leq 1$. Then Algorithm 1 with full participation satisfies,*

$$\hat{f}_\alpha(\tilde{\omega}_T) - \hat{f}^*_\alpha \leq \frac{2||\bar{\theta}_0 - \theta^*||^2}{\beta T} + 26L\beta^2(\kappa-1)^2\rho^2 + \mu\tau_\alpha,$$

where $\tilde{\omega}_T := \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^n p_i \omega_t^i$ and $L := (L_1 + \frac{L_2^2+1}{\mu}) \max_i \frac{1}{\alpha_i}$.

Theorem 1 shows that `rFedFair` converges at rate $\mathcal{O}(1/T)$, which is the classical result for convex optimization. Note that there are two additional error terms in our bound. The second term is due to the client heterogeneity and would reduce to 0 when $\rho = 0$ or $\kappa = 1$, which is consistent with existing results. The last term is introduced by the smooth approximation of the original Problem (4). We can make it small by choosing a small μ. For instance, for small ϵ, by picking $\mu = \frac{\epsilon}{2\tau_\alpha}$, the result in Theorem 1 shows that `rFedFair` requires $\frac{T}{\kappa} = \mathcal{O}(1/\epsilon^2)$ rounds of communication between clients and server to achieve a ϵ-approximate solution.

5.1.2. Partial Participation

As we discussed in Section 4, in the practice of federated learning where the number of clients is very large, it is more desirable to perform averaging over a random subset of clients. We now shift our attention to the case and provide convergence guarantees for `rFedFair` with partial participation.

Theorem 2. *Under the assumptions, if we choose the learning rate β such that $L\beta(3\kappa/K + 2) \leq \frac{1}{20}$ and $6L^2\beta^2(\kappa-1)^2 \leq 1$. Then Algorithm 1 with partial participation satisfies,*

$$\mathbb{E}(\hat{f}_\alpha(\tilde{\omega}_T) - \hat{f}^*_\alpha) \leq \frac{2||\bar{\theta}_0 - \theta^*||^2}{\beta T} + 26L\beta^2(\kappa-1)^2\rho^2 + \frac{12}{K}\beta\kappa\rho^2 + \mu\tau_\alpha,$$

where the expectation is with respect to randomness in selecting the clients.

The result in Theorem 2 is similar to that of Theorem 1 except the learning rate condition and an additional term $\frac{12}{K}\beta\kappa\rho^2$ which measures the difference between random selection of clients and full participation. We note that despite that the convergence rate depends on the sampling size K, that influence might be limited because of the presence of other error terms, i.e., $26L\beta^2(\kappa-1)^2\rho^2$. Thus, in practice, one may choose a small set of clients to overcome the problem of dropouts without severely harming the training process. This result might be extended to other sampling schemes, and we leave it to future work.

5.2. Generalization Bounds

In previous sections, we propose an algorithm to minimize the empirical RC-FL probelm (3) which is guaranteed to find an approximate solution. Now we provide learning guarantees for generalization to the true Q_α-weighted loss (1).

To simplify notation, we denote a function class $\mathcal{H}$ by compositing the functions in $\mathcal{F}$ with the loss function $l(\cdot,\cdot)$, i.e., $\mathcal{H} = \{(x, y) \to l(f_\omega(x), y) : \omega \in \mathcal{W}\}$. The Rademacher complexity of the function space $\mathcal{H}$ given training sample $S^i = \{(x^i_j, y^i_j) \in \mathcal{X} \times \mathcal{Y} : 1 \leq j \leq m_i\}$ drawn i.i.d. from some distribution P_i is defined as

$$\mathfrak{R}^i_{m_i}(\mathcal{H}) = \mathbb{E}_{\sigma, S^i \sim P_i^{m_i}}\left[\sup_{\omega\in\mathcal{W}} \frac{1}{m_i}\sum_{j=1}^{m_i} \sigma_j l(f_\omega(x^i_j), y^i_j)\right],$$

where $\{\sigma_j\}_{j=1}^{m_i}$ are independent Rademacher random variables, i.e., $\mathbb{P}[\sigma_j = +1] = \mathbb{P}[\sigma_j = -1] = \frac{1}{2}$. In federated learning setting, each client has its own data from a different distribution. Therefore, we define a weighted Rademacher complexity for function space $\mathcal{H}$ with respect to the joint

data $S = (S^1, \cdots, S^n)$

$$\mathfrak{R}_m(\mathcal{H}) = \mathbb{E}_{\sigma,S}\left[\sup_{\omega\in\mathcal{W}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{p_i}{\alpha_i m_i}\sigma_{ij} l(f_\omega(x_j^i), y_j^i)\right].$$

With these definitions at hand, we can state our first result characterizing uniform convergence properties of Q_α-weighted loss in terms of weighted Rademacher complexity.

Theorem 3. *Suppose that the function space $\mathcal{H}$ is bounded, i.e., there exists some $B > 0$ such that $l(f_\omega(x), y) \le B$ holds for all $\omega \in \mathcal{W}$ and $(x, y) \in \mathcal{Z}$. Fix $\alpha = (\alpha_1, \cdots, \alpha_n)$ and $m = (m_1, \cdots, m_n)$. Then, for any $\delta > 0$, with probability at least $1 - \delta$ over the draw of samples $S^i \sim P_i^{m_i}$, for all $\omega \in \mathcal{W}$*

$$f_\alpha(\omega) \le \tau_\alpha \hat{f}_\alpha(\omega) + 2\mathfrak{R}_m(\mathcal{H}) + B\sqrt{\sum_{i=1}^{n} \frac{p_i^2 \log(\frac{1}{\delta})}{2\alpha_i^2 m_i}}. \quad (6)$$

Remark 3. If the loss function l takes values in $\{+1, -1\}$ and the function space $\mathcal{H}$ admits VC-dimension d, the data-dependent weighted Rademacher complexity $\mathfrak{R}_m(\mathcal{H})$ can be upper bounded by $\sqrt{\sum_{i=1}^{n} \frac{2dp_i^2}{\alpha_i^2 m_i} \log(\frac{em}{d})}$.

Theorem 3 recovers the usual uniform convergence bound for expected loss if letting $a_i \to 1$ for all i. We note that Mohri et al. (2019) also derive a bound using weighted Rademacher complexity. Compared to (6), their bound has an additional non-vanishing term $B\iota$, and the last term of the bound is multiplied by an extra factor of $\sqrt{n \log 1/\iota}$; roughly, this is due to their proof technique relying on a use of union bound over a ι-cover of the simplex Δ_n. Theorem 3, on the other hand, exploits the relation between Q_α-weighted losses and mean of client losses to arrive at a bound, which can avoid invoking a covering, but at the expense of constant-factor τ_α to $\hat{f}_\alpha(\omega)$. One may expect learning guarantees, not scaling with this Lipschitz constant τ_α, and thus we present an alternative bound which removes the constant factor at a price of weaker last two terms.

Theorem 4. *Let $\alpha = (\alpha_1, \cdots, \alpha_n)$ and $m = (m_1, \cdots, m_n)$ be fixed, and let the function space $\mathcal{H}$ be bounded, i.e., there exists some $B > 0$ such that $\sup_{z\in\mathcal{Z}} l(f_\omega(x), y) \le B$ holds for all $\omega \in \mathcal{W}$. Then, for any $\delta > 0$, with probability at least $1 - \delta$ over the draw of samples $S^i \sim P_i^{m_i}$, for all $\omega \in \mathcal{W}$*

$$f_\alpha(\omega) \le \hat{f}_\alpha(\omega) + 2\sum_{i=1}^{n} \frac{p_i}{\alpha_i}\mathfrak{R}_{m_i}^i(\mathcal{H}) + \sum_{i=1}^{n} \frac{p_i}{\alpha_i} B\sqrt{\frac{\log\frac{2n}{\delta}}{2m_i}}. \quad (7)$$

Similar to typical uniform convergence guarantees for empirical risk, the bound (7) vanishes to zero at the rate $1/\sqrt{m_i}$ for standard hypothesis space whose Rademacher complexity could be bounded from above by $\tilde{\mathcal{O}}(1/\sqrt{m_i})$ term.

Compared to the generalization bound of Theorem 3, the bound (7) does not involve a constant factor τ_α. But the last two terms are less favorable that of (6). This can be observed as follows. By the sub-additivity of sup and the linearity of expectation, we can write

$$\begin{aligned}
&\mathfrak{R}_m(\mathcal{H}) \\
&= \mathbb{E}_{\sigma,S}\left[\sup_{\omega\in\mathcal{W}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{p_i}{\alpha_i m_i}\sigma_{ij} l(f_\omega(x_j^i), y_j^i)\right] \\
&\le \mathbb{E}_{\sigma,S}\left[\sum_{i=1}^{n} \frac{p_i}{\alpha_i} \sup_{\omega\in\mathcal{W}} \sum_{j=1}^{m_i} \frac{1}{m_i}\sigma_{ij} l(f_\omega(x_j^i), y_j^i)\right] \\
&= \sum_{i=1}^{n} \tfrac{p_i}{\alpha_i}\mathfrak{R}_{m_i}^i(\mathcal{H})
\end{aligned}.$$

Analogously, the last term satisfies the inequality $\sqrt{\sum_{i=1}^{n} \frac{p_i^2}{\alpha_i^2 m_i}} \le \sum_{i=1}^{n} \frac{p_i}{\alpha_i}\sqrt{\frac{1}{m_i}}$ by subadditivity of $\sqrt{\cdot}$.

6. Experiments

In this section, we numerically evaluate the performance of the proposed RC-FL framework and `rFedFair` algorithm in terms of accuracy and fairness on real-world datasets. We experiment with four federated datasets considered in prior work using both convex and non-convex models, including Fashion MNIST (Xiao et al., 2017) with a logistic regression model, a Vehicle dataset collected from a distributed sensor network (Duarte & Hu, 2004) with a linear SVM, tweet data curated from Sentiment140 (Go et al., 2009) with a LSTM classifier for text sentiment analysis. Despite that the convergence guarantees for our algorithm only hold for convex loss functions, we empirically show that it behaves well in non-convex models.

We implement `rFedFair` in the Tensorflow (Abadi et al., 2016) platform and simulate a federated learning scenario with one server and n clients, where n is the total number of clients in the datasets. To construct client data, we partition each dataset in the following way, and for each client, we randomly split data into 80% training set, 10% validation set and 10% test set.

Fashion MNIST. The Fashion MNIST (Xiao et al., 2017) dataset is an MNIST-like dataset where images are classified into 10 categories of clothing. We follow the same procedure as that in Mohri et al. (2019) to construct three client datasets for a subset of data labelled with categories `t-shirt/top`, `pullover`, and `shirt`, each consisting of a class of clothing.

Vehicle. The Vehicle dataset consists of sensor data collected from a distributed network of 23 sensors (Duarte & Hu, 2004). Each data has a 100-dimension feature and a binary label. We model each sensor as a client. This produces a dataset with 23 clients. We then train a linear SVM to predict whether a vehicle is AAV-type or DW-type.

Sent140. The dataset is a collection of tweets from $1,101$ accounts from Sentiment140 (Go et al., 2009) where each

Table 1. Test accuracy distribution for models trained with different objectives.

DATASET	METHODS	AVERAGE (%)	WORST 10% (%)	BEST 10% (%)	VARIANCE
VEHICLE	FEDAVG	87.3±0.5	43.0±1.0	95.7±1.0	291±18
	AFL	84.3±0.4	49.3±1.6	93.4±0.7	239±14
	q-FFL	87.7±0.7	69.9±0.6	94.0±0.9	48±5
	RC-FL	87.6±0.3	**73.4**±2.8	**94.3**±0.9	**39**±11
SENT140	FEDAVG	65.1±4.8	15.9±4.9	100.0±0.0	697±132
	AFL	61.4±0.6	12.9±1.3	100.0±0.0	689±39
	q-FFL	66.5±0.2	23.0±1.4	100.0±0.0	509±30
	RC-FL	**70.2**±0.8	**29.0**±0.6	100.0±0.0	**486**±12

account is associated with a client. We train a model consisted of two LSTM layers followed by one fully-connected layer for binary sentiment classification which takes a 25-word sequence as input and embeds each of these into a 300-dimensional space using pretrained Glove (Pennington et al., 2014).

In all our experiments, we compare RC-FL with the model trained with standard federated learning objective (FedAvg) (McMahan et al., 2017), agnostic loss (AFL) (Mohri et al., 2019) and q-FFL (Li et al., 2020a), where the latter two aim to address fairness issues in federated learning. We use `rFedFair` with full participation on Fashion MNIST and Vehicle datasets as the number of clients is small, and partial participation on Sentiment140 where we sample 10 clients at each communication round, i.e., $K = 10$. Our framework is flexible in that it allows each client to select different α_i. For Fashion MNIST, we choose different α_i for each client. For Vehicle, we choose α_1 for a client and let other clients share the same α_2. For Sentiment140, since the number of clients is very large, for simplicity, we just choose the same α for all clients. The number of local updates is fixed to $\kappa = 5$ for all the experiments. All results are averaged over 5 independent trials.

In Figure 1, we compare the test accuracy across the three clients from Fashion MNIST dataset. We observe that our RC-FL model achieves fairer (almost identical) test accuracy across the clients while maintaining roughly the same average accuracy. We further report the worst and best 10% test accuracy and the variance of test accuracy distribution for Vehicle and Sent140 datasets in Table 1. Again, RC-FL achieves lower variance and higher test accuracy on the clients with worse 10% performance for Vehicle dataset despite slightly reduction in average accuracy. Finally, for Sent140, our model performs significantly better than other baselines in terms of both average accuracy and fairness.

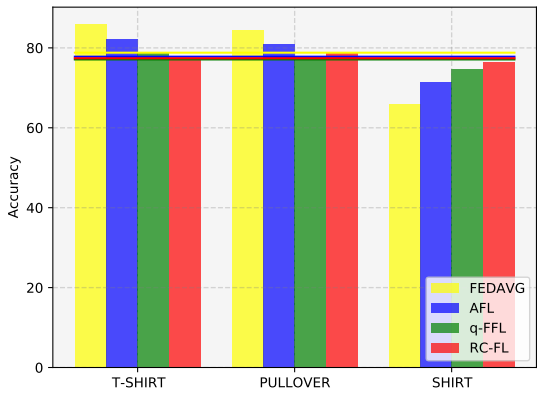

Figure 1. Fashion MNIST dataset: test accuracy across clients for models trained with different objectives.

7. Conclusion

In this paper, we propose RC-FL, a new federated learning framework in which the centralized model is optimized with respect to a robust client weighted loss. We define a notion of fairness named HCVaR which generalizes the conditional value at risk by taking into account client heterogeneity and show an equivalence between this formulation and HCVaR, implying that RC-FL is a fairness-aware objective. We then present an efficient algorithm to solve this objective and provide theoretical guarantees regarding both convergence and generalization. Experimental results show that the solution obtained by RC-FL can lead to significant benefits in practice in terms of both accuracy and fairness. There remains many avenues for future investigation. Our framework requires that the weight q_i lies in an interval $[0, \alpha_i^{-1}]$ and therefore focuses on clients with large losses. However, in some scenarios, one may be concerned with more structured uncertainty, e.g., a small subset $[a_i, b_i] \subset [0, \alpha_i^{-1}]$. An interesting question is whether our results can be generalized to that general case. Moreover, the convergence guarantee in Theorem 2 only applies to sampling with replacement, and extending the result to other sampling schemes (e.g., sampling without replacement) might be an interesting topic for future research.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pp. 265–283, 2016.

Bullins, B. Highly smooth minimization of non-smooth problems. In *Conference on Learning Theory*, pp. 988–1030. PMLR, 2020.

Chow, Y. and Ghavamzadeh, M. Algorithms for cvar optimization in mdps. *Advances in neural information processing systems*, 27:3509–3517, 2014.

Curi, S., Levy, K., Jegelka, S., Krause, A., et al. Adaptive sampling for stochastic risk-averse learning. *Advances in Neural Information Processing Systems*, 33, 2020.

Donini, M., Oneto, L., Ben-David, S., Shawe-Taylor, J. S., and Pontil, M. Empirical risk minimization under fairness constraints. In *Advances in Neural Information Processing Systems*, pp. 2791–2801, 2018.

Duarte, M. F. and Hu, Y. H. Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing*, 64(7):826–838, 2004.

Dwork, C., Immorlica, N., Kalai, A. T., and Leiserson, M. Decoupled classifiers for group-fair and efficient machine learning. In *Conference on Fairness, Accountability and Transparency*, pp. 119–133, 2018.

Fan, Y., Lyu, S., Ying, Y., and Hu, B. Learning with average top-k loss. In *Advances in neural information processing systems*, pp. 497–505, 2017.

Go, A., Bhayani, R., and Huang, L. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009, 2009.

Haddadpour, F. and Mahdavi, M. On the convergence of local descent methods in federated learning. *arXiv preprint arXiv:1910.14425*, 2019.

Jeong, S. and Namkoong, H. Robust causal inference under covariate shift via worst-case subpopulation treatment effects. In *Conference on Learning Theory*, pp. 2079–2084. PMLR, 2020.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

Khaled, A., Mishchenko, K., and Richtárik, P. First analysis of local gd on heterogeneous data. *arXiv preprint arXiv:1909.04715*, 2019.

Konečnỳ, J., McMahan, H. B., Ramage, D., and Richtárik, P. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*, 2016.

Konen, J., McMahan, H. B., Yu, F. X., Richtarik, P., Suresh, A. T., and Bacon, D. Federated learning: Strategies for improving communication efficiency. In *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.

Krokhmal, P., Palmquist, J., and Uryasev, S. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of risk*, 4:43–68, 2002.

Laguel, Y., Pillutla, K., Malick, J., and Harchaoui, Z. Device heterogeneity in federated learning: A superquantile approach. *arXiv preprint arXiv:2002.11223*, 2020.

Lee, J., Park, S., and Shin, J. Learning bounds for risk-sensitive learning. *Advances in Neural Information Processing Systems*, 33, 2020.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. Federated optimization in heterogeneous networks. *arXiv preprint arXiv:1812.06127*, 2018.

Li, T., Sanjabi, M., Beirami, A., and Smith, V. Fair resource allocation in federated learning. In *International Conference on Learning Representations*, 2020a. URL https://openreview.net/forum?id=ByexElSYDr.

Li, X., Huang, K., Yang, W., Wang, S., and Zhang, Z. On the convergence of fedavg on non-iid data. In *International Conference on Learning Representations*, 2020b. URL https://openreview.net/forum?id=HJxNAnVtDS.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pp. 1273–1282. PMLR, 2017.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. *Foundations of machine learning*. MIT press, 2018.

Mohri, M., Sivek, G., and Suresh, A. T. Agnostic federated learning. In *International Conference on Machine Learning*, pp. 4615–4625, 2019.

Neumann, J. v. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.

Pennington, J., Socher, R., and Manning, C. D. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

Reisizadeh, A., Farnia, F., Pedarsani, R., and Jadbabaie, A. Robust federated learning: The case of affine distribution shifts. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Rockafellar, R. T., Uryasev, S., et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Shalev-Shwartz, S. and Wexler, Y. Minimizing the maximal loss: How and why. In *ICML*, pp. 793–801, 2016.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. *Lectures on stochastic programming: modeling and theory*. SIAM, 2014.

Soma, T. and Yoshida, Y. Statistical learning with conditional value at risk. *arXiv preprint arXiv:2002.05826*, 2020.

Stich, S. U. Local SGD converges fast and communicates little. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=S1g2JnRcFX.

Wang, J., Liu, Q., Liang, H., Joshi, G., and Poor, H. V. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in Neural Information Processing Systems*, 33, 2020.

Williamson, R. and Menon, A. Fairness risk measures. In *International Conference on Machine Learning*, pp. 6786–6797, 2019.

Xiao, H., Rasul, K., and Vollgraf, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Yang, C., Wang, Q., Xu, M., Wang, S., Bian, K., and Liu, X. Heterogeneity-aware federated learning. *arXiv preprint arXiv:2006.06983*, 2020.

Zafar, M. B., Valera, I., Gomez Rodriguez, M., and Gummadi, K. P. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th international conference on world wide web*, pp. 1171–1180, 2017.