Fair Federated Learning via the Proportional Veto Core

Bhaskar Ray Chaudhury Aniket Murhekar Zhuowen Yuan Bo Li Ruta Mehta Ariel D. Procaccia

Abstract

Previous work on fairness in federated learning introduced the notion of *core stability*, which provides utility-based fairness guarantees to any subset of participating agents. However, these guarantees require strong assumptions on agent utilities that render them impractical. To address this shortcoming, we measure the quality of output models in terms of their ordinal *rank* instead of their cardinal utility, and use this insight to adapt the classical notion of *proportional veto core (PVC)* from social choice theory to the federated learning setting. We prove that models that are *PVC-stable* exist in very general learning paradigms, even allowing non-convex model sets, as well as non-convex and non-concave loss functions. We also design `Rank-Core-Fed`, a distributed federated learning algorithm, to train a PVC-stable model. Finally, we demonstrate that `Rank-Core-Fed` outperforms baselines in terms of fairness on different datasets.

1. Introduction

Federated learning (FL) provides an effective distributed learning paradigm whereby a group of agents holding local data samples can train a joint model without sharing their private data. The paradigm has been widely used for different applications such as autonomous vehicles (Elbir et al., 2020) and digital healthcare (Dayan et al., 2021; Xu et al., 2021).

Due to the heterogeneity in their data distributions, the benefit derived by the agents from the final model vary. Thus, a fast-growing line of work (Mohri et al., 2019; Huang et al., 2021; Li et al., 2020b; Zhang et al., 2023; Chaudhury et al., 2022) focuses on *good intent fairness* in FL, where the goal is to prevent the final model from over-fitting to particular agents at the expense of other agents.

A particularly compelling approach to fairness in FL was introduced by Chaudhury et al. (2022).[1] Their FL algorithm, `CoreFed`, guarantees outcomes that are *core stable*; this intuitively means that *every* possible coalition of agents receives utility guarantees that are stronger the larger the coalition is. In particular, core-stable outcomes are *proportional* — for each of the n agents, their utility for the output model is at least $1/n$ of their maximum utility for any model — and *Pareto optimal* — it is impossible to improve the utility of some agents without harming others. In addition, core-stable outcomes are robust to noisy outliers, as the coalition of non-noisy agents, which consists of most agents, has very strong guarantees.

However, the guarantees provided by `CoreFed` rely on the assumption that the utility functions of agents are concave. In fact, as we show in Example 1, this is unavoidable: even weaker guarantees that relax proportionality — specifically, giving each participating agent *any* fraction of their best possible utility — are unattainable for non-concave utilities. This is a major obstacle to applying the approach for more general machine learning models that are used in practice, such as deep neural networks.

Our approach. To overcome this obstacle, we seek fairness guarantees that are not functions of utilities, yet still reflect the quality of the model for the agent. Our key insight is to focus on the *rank* of models. Intuitively, we say that a model θ has rank $r \in [0, 1]$ for agent i if i weakly prefers (based on utility) θ to an r-fraction of all possible models (so more preferred models have higher rank).

This idea of rank allows us to adapt a notion of core from social choice theory that is tailored to ordinal preferences: the *proportional veto core (PVC)*. In our FL setting, we say that a model θ is *PVC-stable* if the fraction of models the coalition unanimously ranks higher than θ is at most $1 - |T|/n$; informally, each coalition can veto a fraction of models proportional to its size, hence the name of this notion. The PVC provides "rankwise" analogues of the guarantees given by the ("original") core. In particular, setting $T = \{i\}$ implies an ordinal notion of proportionality, whereby agent i ranks θ in the top $(1 - 1/n)$ fraction of their models; we call this property *rankwise proportionality*. Similarly, setting $T = \{n\}$ implies Pareto optimality. Finally, the PVC ensures that similarity in the preferences of a large subset of agents would imply high utility for these agents. As an extreme example, if 70% of agents have the same utility function, then a PVC-stable model is guaranteed to be in the

[1] See Section 1.1 for a more detailed comparison of fair federated learning algorithms

top 30% of models for these agents — , and this guarantee smoothly extends to the case of similarity instead of identity between agents (Lemma 1). Once again, the implication is that a relatively small contingent of outlier agents cannot significantly affect the guarantees of a cohesive majority.

Having introduced and justified the proportional veto core in the context of FL, our technical challenge is now obvious: *Determine under what conditions PVC-stable models are guaranteed to exist and design a practical FL algorithm that outputs PVC-stable models.*

Our results. Our main theoretical result is that PVC-stable models always exist under the essentially unrestrictive assumption that the utility functions of the agents are *Lebesgue-measurable*. Note that this assumption is much milder than the concavity assumption required for core-stability (Chaudhury et al., 2022), which means that by moving from a cardinal to ordinal notion of core, we are indeed able to significantly relax the assumptions and obtain guarantees that hold in practical machine learning settings, including deep neural networks.[2]

Next, we design an algorithm, `Rank-Core-Fed`, which outputs core-stable models that also achieve high utilitarian social welfare (sum of utilities). A challenge is that it is intractable to compute the PVC with respect to all possible models without imposing strong assumptions. Instead of optimizing over the "full" PVC, therefore, `Rank-Core-Fed` starts by identifying a promising model, computes a set of representative models $\mathcal{P}$ around the initial model, and then computes the proportional veto core with respect to $\mathcal{P}$.

Finally, we conduct comprehensive experiments on different datasets with around 100 agents. Compared to baselines such as `FedAvg` and `CoreFed`, we find that `Rank-Core-Fed` has similar utilitarian social welfare and significantly better fairness guarantees.

1.1. Related work.

There is a significant body of work on fairness in FL. For instance, Huang et al. (2021) and Yang et al. (2021) impose fairness at the agent selection phase, where the server requests local model updates from the agents, while Wang et al. (2021) mitigate large differences in the magnitudes of the gradients sent to the server by each agent. Mohri et al. (2019) achieve fairness by reducing the final model's worst performance on any client (a.k.a. egalitarian fairness). Donahue & Kleinberg (2021) discuss the impacts of egalitarian fairness and also introduce the concept of *proportional fairness*, where the error rates of the agents are proportional to the size of their local data.[3] This notion is more appealing than egalitarian fairness in scenarios where all agents have equal opportunity to collect and contribute data and therefore rewarding agents with larger contributions (larger datasets) seems fair. Following fairness objectives used in telecommunications (Mo & Walrand, 2000), Li et al. (2020b) use the q-mean welfare (the weighted average of the q^{th} power of the utilities) to quantify the overall performance and fairness of a model and determine the final model by maximizing this objective. Li et al. (2020a) introduce a novel framework through a subtle modification of the ERM paradigm that helps establish fairness, robustness, reduction in the influence of outliers, and a few other desirable properties. For a detailed overview of all these notions, we refer the reader to the survey by Shi et al. (2021).

Note that a significant part of the literature chooses the final model by optimizing an objective that reflects the fairness and efficiency of the model (Mohri et al., 2019; Li et al., 2020b;a). However, little is known about the *fairness guarantees* for individual agents; in other words, those algorithms do not make any promises to the agents about their benefit from the final model. Such guarantees provided by the solution are often desirable as they help *explain* the fairness of the final model to the agents (Procaccia, 2019). In FL, such guarantees can also incentivize participation.

Methods that do provide explainable fairness guarantees to the agents include those of Chaudhury et al. (2022) and Zhang et al. (2023). Zhang et al. (2023) guarantee proportionality with respect to the utilities, and, as discussed earlier, Chaudhury et al. (2022) ensure that the final model is core-stable. However, both papers require very strong assumptions (like concavity of the utility functions) for their guarantees. In this paper, we achieve similar guarantees through the proportional veto core in far more general settings.

2. Theoretical Guarantees

We start by formally presenting our setting. A federated learning (FL) instance has n participating agents, denoted by $[n] = \{1, \dots, n\}$, who want to jointly training a model $\theta \in P \subseteq \mathbb{R}^N$.

To be consistent with the standard literature in social choice theory as well as prior work on fairness in FL (Chaudhury et al., 2022), we denote the utility of agent i for model θ by $u_i(\theta)$. The utility $u_i(\theta)$ measures the accuracy of θ on the data distribution of agent i. For convenience, we assume that the utility functions are normalized, i.e., for each agent i, we have $u_i(\theta) \in [0, 1]$; this assumption is without loss of

[2] Standard applications of deep neural networks result in continuous utility functions (Sze et al., 2017; Zhang & Sabuncu, 2018), which are Lebesgue-measurable (see Proposition 2 in Appendix A).

[3] Note that this notion of proportionality is very different from the notion of proportionality used in our paper, which is inspired by the exact same notion in social choice theory

generality as our fairness guarantees are scale invariant.

2.1. Inapproximability of Proportionality

As discussed in Section 1, our approach is motivated by the observation that core-stability, in the sense of Chaudhury et al. (2022), is inapproximable when agent utility functions are non-concave, e.g., in deep neural networks. In fact, even approximate proportionality is infeasible, in the sense that it is impossible to guarantee agents *any* fraction of their maximum utility for models in P. This is true even for smooth utility functions, as demonstrated by the following example.

Example 1. Let $P = [0, 1]$ and $n = 2$. We define the utility functions of the two agents as follows:

$$u_1(\theta) = \begin{cases} L \cdot (\theta - \frac{1}{\sqrt{L}})^2 & \theta \leq \frac{1}{\sqrt{L}} \\ 0 & \theta > \frac{1}{\sqrt{L}} \end{cases}$$

$$u_2(\theta) = u_1(1 - \theta)$$

We first note that both utility functions are $2L$-smooth: Consider $u_1(\cdot)$ and two points θ_1 and $\theta_2 \in [0, 1]$. If both $\theta_1, \theta_2 \in [0, 1/\sqrt{L}]$, then

$$||\nabla u_1(\theta_1) - \nabla u_1(\theta_2)||_2 = 2L||\theta_2 - \theta_1||_2.$$

If $\theta_1, \theta_2 > 1/\sqrt{L}$, then

$$||\nabla u_1(\theta_1) - \nabla u_1(\theta_2)||_2 = 0.$$

Lastly, if $\theta_1 \leq 1/\sqrt{L}$, and $\theta_2 > 1/\sqrt{L}$, then

$$\nabla u_1(\theta_2) = \nabla u_1(1/\sqrt{L}) = 0,$$

and therefore

$$\begin{aligned} ||\nabla u_1(\theta_1) - \nabla u_1(\theta_2)||_2 &= ||\nabla u_1(\theta_1) - \nabla u_1(1/\sqrt{L})||_2 \\ &= 2L||\theta_1 - 1/\sqrt{L}||_2 \\ &\leq 2L||\theta_1 - \theta_2||_2. \end{aligned}$$

The highest possible utility for both agents is 1: for agent 1, it is realized when $\theta = 0$, and for agent 2, it is realized when $\theta = 1$. However, for any θ, one of the agents will have a utility of 0: if $\theta \geq 1/\sqrt{L}$ then the utility of agent 1 is 0, while if $\theta < 1/\sqrt{L}$ then the the utility of agent 2 is 0. Thus, at least one of the agents will have a utility of zero and cannot be guaranteed any approximation of the proportionality.

2.2. Definition of the Proportional Veto Core

The above example motivates us to define a fairness notion that is feasible much more broadly. With the aim of achieving core-stability like guarantees akin to those of Chaudhury et al. (2022) in general settings, we build on the notion of proportional veto core (Moulin, 1981).

In the classical social choice setting, n voters each express their preferences as a ranking of the m alternatives. The proportional veto core principle requires that if a is a winning alternative, then for any coalition T of voters, the number of alternatives unanimously preferred by agents in T over a should be at most $m(1 - |T|/n)$. In other words, if any coalition T prefers a sufficiently large number of candidates ($\geq (1 - |T|/n)$ fraction) over a specific alternative a then c is "vetoed" by the coalition.

To adapt this notion to our FL setting, we rely on ideas from measure theory. For a set $D \subseteq \mathbb{R}^N$, the *Lebesgue measure* $\lambda(D)$ measures the high-dimensional volume of D. We make the assumption that the set of models P is Lebesgue-measurable; this is a very mild assumption as, for example, any closed set is measurable. We also assume that the agent utility functions $u_i(\cdot)$ are Lebesgue-measurable [4]. This assumption is weaker than continuity: for completeness, we show in Proposition 2 in Appendix A that any *piece-wise continuous* function is Lebesgue-measurable.

Definition 1 (ε-Proportional Veto Core)**.** For $\varepsilon \in (0, 1/n)$, let $v(T) := \frac{|T|}{n} - \varepsilon$ be the *ε-proportional veto* function. A model $\theta \in P$ is said to be *blocked* by a coalition $\emptyset \neq T \subseteq [n]$ of agents if there exists a Lebesgue-measurable set $B \subseteq P$ such that for all $i \in T$ and $\theta' \in B$, $u_i(\theta') \geq u_i(\theta)$ with at least one inequality strict, and

$$v(T) \geq 1 - \frac{\lambda(B)}{\lambda(P)}.$$

A model θ is said to be in the *ε-proportional veto core (PVC)* of P — or equivalently, θ is *ε-PVC-stable* — if it is not blocked by any coalition of agents.

Put another way, a model θ is in the proportional veto core of P if for any subset T of agents, the fraction of models that are a Pareto-improvement over θ for agents in T is strictly smaller than than $1 - |T|/n + \varepsilon$.

2.3. Properties of PVC-Stable Models

We now show that an ε-PVC-stable model satisfies the following desirable properties.

(i) *(Fairness)* An ε-PVC-stable model θ is *ε-rankwise proportional*. This means every agent prefers θ to a $(1/n - \varepsilon)$ fraction of all models.

(ii) *(Efficiency)* An ε-PVC-stable model is ε-Pareto-optimal. This means that the set of models which are

[4] A function is measurable iff its lower level sets are measurable; see Appendix A

weakly-preferred over θ by all agents and strictly preferred by some is at most an ε fraction of all models.

(iii) *(Improved guarantees under aligned preferences)* In instances where a large set T of $\alpha \cdot n$ agents have *well-aligned* preferences, an ε-PVC-stable model is in the top $(\alpha - 2\varepsilon)$ fraction of models for each agent in T.

Below we formally define these properties and prove that an ε-PVC-stable model satisfies them.

PVC-stable models are rankwise proportional. We define the rank of a model θ for an agent i as follows. Let $P_i(\theta) = \{\theta' \mid u_i(\theta) \geq u_i(\theta')\}$.[5] We define

$$\mathsf{rank}_i(\theta) = \lambda(P_i(\theta))/\lambda(P).$$

Since the Lebesgue-measure of the set D is the high-dimensional volume of D, $\mathsf{rank}_i(\theta)$ intuitively measures the fraction of models that agent i prefers at most as much as the model θ.

We now define our fairness metric. Given any FL instance, we say that θ is *rankwise proportional* if for every agent, θ is at least as good as a $1/n$ fraction of the entire set of models. Formally, for $\varepsilon > 0$:

Definition 2 (ε-rankwise proportionality)**.** A model θ is said to be *ε-rankwise proportional* if $\mathsf{rank}_i(\theta) \geq 1/n - \varepsilon$ for all agents i.

To see why an ε-PVC-stable model θ is ε-rankwise proportional, consider $T = \{i\}$ for any fixed agent i. By Definition 1, i strictly prefers at most a $(1 - |\{i\}|/n + \varepsilon)$ fraction of models over θ. Equivalently, we have $\mathsf{rank}_i(\theta) \geq 1/n - \varepsilon$, implying that θ is ε-rankwise proportional.

Finally, we remark that this guarantee is nearly tight, as seen from the following example.

Example 2. Let $P = [0, 1]$. Consider two agents with $u_1(\theta) = \theta$, and $u_2(\theta) = 1 - \theta$. Then

$$\mathsf{rank}_1(\theta) = \frac{\lambda([0, \theta])}{\lambda([0, 1])} = \theta,$$

and

$$\mathsf{rank}_2(\theta) = \frac{\lambda([\theta, 1])}{\lambda([0, 1])} = 1 - \theta.$$

Note that $\theta^* = 1/2$ is a PVC-stable model such that $\mathsf{rank}_1(\theta^*) = \mathsf{rank}_2(\theta^*) = 1/n$, since $n = 2$. However there is no model $\theta \in P$ such that $\mathsf{rank}_1(\theta) > 1/2$ and $\mathsf{rank}_2(\theta) > 1/2$.

[5] We show that the sets $P_i(\theta)$ are Lebesgue measurable in Appendix A.

PVC-stable models are Pareto-optimal. To measure the economic efficiency of a model, we use the standard notion of Pareto-optimality (PO).

Definition 3. (Pareto optimality) A model $\theta \in P$ *Pareto-dominates* a model $\theta' \in P$ if all agents prefer θ at least as much as θ', and at least one agent prefers θ strictly more than θ'. Thus, $\forall i \in [n] : u_i(\theta) \geq u_i(\theta')$, with at least one inequality strict. A model θ is said to be *Pareto optimal* (PO) if no model θ' Pareto-dominates θ. A model θ is ε-PO if the set B of models which Pareto-dominate θ satisfies $\lambda(B) \leq \varepsilon \cdot \lambda(P)$.[6]

To see why an ε-PVC-stable model θ is ε-PO, consider $T = [n]$ in Definition 1. Then the fraction of models $\theta' \in P$ that Pareto-dominate θ is strictly smaller than ε, implying ε-Pareto optimality.

PVC-stable models give improved guarantees under aligned preferences. Here we show how a PVC-stable model gives better guarantees to large groups of agents that have aligned preferences. This is particularly useful in instances that contain noisy agents (agents with vastly different preferences from the majority), as such a property ensures that a small group of noisy agents would not be able to significantly impact the fairness guarantees of a large group of non-noisy agents.

We first remark that this property is not always satisfied by some of the existing fairness notions in FL. The following example supports this claim.

Example 3. Consider n agents and the set of models $P = [1, m]$, where n divides m. Let $u_i(\theta) = 1 - (\theta - 1)/m$ for all $i \in [n-1]$. Let $u_n(\theta) = \frac{\theta}{m(n-1)}$ for $\theta \leq m(1 - 1/n)$, and

$$u_n(\theta) = \frac{1}{n} + (n-1) \cdot \left(\frac{\theta}{m} - \left(1 - \frac{1}{n}\right)\right)$$

for all $\theta > m(1 - 1/n)$.

Note that the maximum utility that any agent can achieve is 1, and all agents have the same utility of $1/n$ for $\theta = m(1 - 1/n)$. Also observe that a federated learning algorithm that incorporates egalitarian fairness, weighted equity-based fairness, proportionality, or core-stability will choose the final model $\theta = m(1 - 1/n)$, giving every agent a utility of $1/n$. However, the same algorithm, if run in the absence of agent n, will choose $\theta = 1$ as the final model, giving every agent in $[n-1]$ a utility of 1. Thus, under all the foregoing fairness notions, a single agent can drastically change the best achievable utility for all other agents.

By contrast, notice that an ϵ-PVC-stable model θ requires that the agents in $[n-1]$ unanimously prefer at most a

[6] Note that Lemma 2 in Appendix A shows that B is a Lebesgue-measurable set.

$1-(n-1)/n+\varepsilon = 1/n+\varepsilon$ fraction of models in P over θ. This will ensure that $\theta \leq m(1/n+\varepsilon)$, guaranteeing each agent in $[n-1]$ a model in their top $(1-1/n-\varepsilon)$ percent and a utility of $(1-1/n-\varepsilon)$ of their best achievable utility.

We now argue that one can expect similar guarantees more generally. Consider a group of T of agents ($|T| = n\alpha$) such that for all $\theta \in P$, we have $|\mathsf{rank}_i(\theta) - \mathsf{rank}_{i'}(\theta)| \leq \delta$ for all $i, i' \in T$, i.e., the preferences of agents in T are δ-*aligned*. For simplicity, we assume that the level sets of the utility functions of the agents have zero measure (instance is non-degenerate). Intuitively, this means that the fraction of models with rank at least β is equal to $1-\beta$, for $\beta \in [0,1]$. We then prove:

Proposition 1. *Let T be a group of agents of size $|T| = n\alpha$ with δ-aligned preferences. If $\theta \in P$ is ε-PVC-stable then for all $i \in T$, $\mathsf{rank}_i(\theta) \geq \alpha - 2\varepsilon - 3\delta$.*

Proof. Assume otherwise. Let $i \in T$, and $\mathsf{rank}_i(\theta) < \alpha - 2\varepsilon - 3\delta$. Then $\mathsf{rank}_{i'}(\theta) < \alpha - 2\varepsilon - 2\delta$ for all $i' \in T$.

Next, define $B = \{\theta' \in P \mid \mathsf{rank}_i(\theta') \geq \alpha - 2\varepsilon\}$. Since our instance is non-degenerate, we have $\lambda(B)/\lambda(P) = 1-\alpha+2\varepsilon$. Furthermore, note that for all $\theta' \in B$ and all $i' \in T$, we have

$$\mathsf{rank}_{i'}(\theta') \geq \alpha - 2\varepsilon - \delta \geq \alpha - 2\varepsilon - 2\delta > \mathsf{rank}_{i'}(\theta).$$

Therefore, models in B Pareto-dominate θ for agents in T. Now, observe that

$$\lambda(B)/\lambda(P) = 1-\alpha+2\varepsilon > 1-|T|/n+\varepsilon,$$

which contradicts that θ is ε-PVC-stable. □

Concretely, if 90% of agents have δ-aligned preferences, then the above claim shows that all agents in T have a rank of at least $0.9 - 2\varepsilon - 3\delta$ for an ε-PVC-stable model θ. By contrast, the individual fairness guarantee of Theorem 1 is rank at least $1/n - \varepsilon$. This shows that under aligned preferences, a significant improvement in the individual fairness guarantee is possible.

2.4. Existence of PVC-Stable Models

Having shown that a model in the proportional veto core satisfies desirable core-like fairness and efficiency properties, we now turn to feasibility. We show that under mild assumptions, the PVC is always non-empty. The following theorem is our main theoretical result.

Theorem 1. *In any FL instance where the set P of feasible models is non-empty and Lebesgue-measurable and agents have Lebesgue-measurable utility functions, the ε-proportional veto core is non-empty for any $\varepsilon \in (0, 1/n)$.*

As mentioned earlier, the assumption that P is Lebesgue-measurable is very mild. Similarly, the assumption that the functions $u_i(\cdot)$ are Lebesgue-measurable is also very mild. In particular, it is weaker than continuity: for completeness, we show in Proposition 2 in Appendix A that any *piece-wise continuous* function is Lebesgue-measurable. Note that this includes standard applications of deep neural networks (Sze et al., 2017; Zhang & Sabuncu, 2018).

To prove the theorem, we require a technical lemma. The lemma asserts that for an agent with a measurable utility function u over a measurable set B, one can find for any given $\alpha \in (0,1)$ a set $A \subseteq B$ of volume α times of that of B such that the agent prefers every model in $B \setminus A$ over A according to u.

Lemma 1. *Let B be a measurable set with $\lambda(B) > 0$, and let $u : B \to \mathbb{R}$ be a measurable function. Then for any $\alpha \in (0,1)$ there exists a set $A \subseteq B$ s.t. $\lambda(A) = \alpha \cdot \lambda(B)$, and for all $\theta_1 \in B \setminus A$ and $\theta_2 \in A$, $u(\theta_1) \geq u(\theta_2)$.*

Proof. For sets $C, D \subseteq B$, let $C \geq_u D$ (resp. $C >_u D$) denote the statement that for all $\theta_1 \in C, \theta_2 \in D$, $u(\theta_1) \geq u(\theta_2)$ (resp. $u(\theta_1) > u(\theta_2)$).

Let $A_\ell = \{\theta \in B | u(\theta) \leq \ell\}$ be the lower-level set of $u(\cdot)$ for $\ell \in \mathbb{R}$. If there exists an $\ell \in \mathbb{R}$ s.t. $\lambda(A_\ell) = \alpha\lambda(B)$ then we are done, as $B \setminus A_\ell \geq_u A_\ell$ by definition of A_ℓ.

Therefore suppose $\lambda(A_\ell) \neq \alpha\lambda(B)$ for every $\ell \in \mathbb{R}$. Let $\ell_1 = \sup\{\ell \mid \lambda(A_\ell) < \alpha\lambda(B)\}$ and $\ell_2 = \inf\{\ell \mid \lambda(A_\ell) > \alpha\lambda(B)\}$. Let $\alpha_1 = \lambda(A_{\ell_1})/\lambda(B)$ and $\alpha_2 = \lambda(A_{\ell_2})/\lambda(B)$. We clearly have $\ell_1 < \ell_2$ and $\alpha_1 < \alpha < \alpha_2$. Note that $A_{\ell_1} \subseteq A_{\ell_2}$.

Define the set $C := A_{\ell_2} \setminus A_{\ell_1}$. Since A_{ℓ_1} and A_{ℓ_2} are measurable by standard properties (see Def. 4 in Appendix A), it follows that $C = A_{\ell_2} \setminus A_{\ell_1}$ is measurable as well. By definition of ℓ_1 and ℓ_2, for every $\theta \in C$, $u(\theta) = \ell_2$. We have

$$\lambda(C) = \lambda(A_{\ell_2}) - \lambda(A_{\ell_1}) = (\alpha_2 - \alpha_1)\lambda(B) > 0.$$

Now we find[7] a set $D \subseteq C$ s.t. $\lambda(D) = \left(\frac{\alpha-\alpha_1}{\alpha_2-\alpha_1}\right) \cdot \lambda(C)$.

Let $A := A_{\ell_1} \cup D$. Note that $A_{\ell_1} \cap D = \emptyset$, hence

$$\begin{aligned}\lambda(A) &= \lambda(A_{\ell_1}) + \lambda(D) = \alpha_1\lambda(B) + \left(\frac{\alpha-\alpha_1}{\alpha_2-\alpha_1}\right)\lambda(B) \\ &= \alpha\lambda(B).\end{aligned}$$

Moreover, we have $B \setminus A \geq_u A$. To see why, consider $\theta_1 \in B \setminus A$ and $\theta_2 \in A$. Then $\theta_1 \notin A_{\ell_1}$ and $\theta_1 \notin D$, implying that $u(\theta_1) \geq \ell_2$. Moreover $\theta_2 \in A_{\ell_1}$ or $\theta_2 \in D$, implying that $u(\theta_2) \leq \ell_2$. Thus $u(\theta_1) \geq u(\theta_2)$.

[7] See https://math.stackexchange.com/questions/2986033/find-a-subset-with-a-specific-lebesgue-measure

We conclude that A is the required set, as $\lambda(A) = \alpha \cdot \lambda(B)$ and $B \setminus A \geq_u A$. □

With the lemma in hand, we can now prove the theorem.

Proof of Theorem 1. For sets $A, B \subseteq P$ and an agent i, let $A \geq_i B$ (resp. $A >_i B$) denote the statement that for all $\theta_1 \in A, \theta_2 \in B$, $u_i(\theta_1) \geq u_i(\theta_2)$ (resp. $u_i(\theta_1) > u_i(\theta_2)$).

We perform the following iterative procedure for n iterations. Let $r = \frac{1}{n} - \frac{\varepsilon}{n+1}$. Let $Q_0 = P$. In iteration i, we choose a set a $P_i \subseteq Q_{i-1} := P \setminus (P_1 \cup \cdots \cup P_{i-1})$ such that $Q_i := Q_{i-1} \setminus P_i \geq_i P_i$ and $\lambda(P_i) = r \cdot \lambda(P)$. Below we argue why such a set P_i can be computed.

Observe by induction that:

$$\lambda(Q_i) = \lambda(P \setminus (P_1 \cup \cdots \cup P_i)) = \lambda(P) \cdot (1 - i \cdot r).$$

Since $r < 1/n$, we have for $i \leq n-1$ that $\lambda(Q_{i-1}) > r \cdot \lambda(P)$. Thus the existence of a set $P_i \subseteq Q_{i-1}$ such that $Q_{i-1} \setminus P_i \geq_i P_i$ is guaranteed due to Lemma 1 above.

We therefore have that $Q_n = P \setminus (P_1 \cup \cdots \cup P_n)$ satisfies $\lambda(Q_n) > 0$, and therefore $Q_n \neq \emptyset$. We claim that any model $\theta \in Q_n$ lies in the proportional veto core.

For the sake of contradiction, suppose this is not the case. Let $T \subseteq [n]$ be a blocking coalition for θ and let $B \subseteq P$ be a set of models s.t. $\lambda(B) = \lambda(P) \cdot (1 - v(T))$ and models in B Pareto-dominate θ for all $i \in T$. Define $S = [k]$ for $k = \arg\max_{i \in T} i$. Observe that:

$$\lambda(Q_k) = \lambda(P \setminus \bigcup_{i \in S} P_i) = \lambda(P) \cdot (1 - |S| \cdot r). \quad (1)$$

Consider any $c \in B$. Since c Pareto-dominates θ for agents in T, there must an agent $i \in T$ s.t. $c >_i \theta$. However since $\{\theta\} \geq_i P_i$, we have that $c \notin P_i$. Hence, $B \cap \bigcup_{i \in T} P_i = \emptyset$. We use this to show that:

$$\begin{aligned}
\lambda(Q_k \cap B) &= \lambda((P \setminus \bigcup_{i \in S} P_i) \cap B) \\
&= \lambda(B \setminus \bigcup_{i \in S} (P_i \cap B)) \\
&= \lambda(B \setminus \bigcup_{i \in S \setminus T} (P_i \cap B)) \\
&= \lambda(B) - \sum_{i \in S \setminus T} \lambda(P_i \cap B) \\
&\geq \lambda(B) - \sum_{i \in S \setminus T} \lambda(P_i) \\
&= \lambda(P) \cdot \big((1 - v(T)) - |S \setminus T| \cdot r\big)
\end{aligned} \quad (2)$$

Since $\lambda(Q_k) \geq \lambda(Q_k \cap B)$, equations 1 and 2 together imply that:

$$\begin{aligned}
\lambda(P) \cdot (1 - |S| \cdot r) &= \lambda(Q_k) \\
&\geq \lambda(Q_k \cap B) \\
&\geq \lambda(P) \cdot \big((1 - v(T)) - |S \setminus T| \cdot r\big),
\end{aligned}$$

which implies $v(T) \geq r|T|$. However this is a contradiction since $r|T| = \frac{|T|}{n} - \frac{\varepsilon |T|}{n+1} > \frac{|T|}{n} - \varepsilon = v(T)$. □

3. `Rank-Core-Fed`: A PVC-Stable Algorithm

In this section, we introduce our distributed learning algorithm `Rank-Core-Fed` that trains a model that is PVC-stable and achieves high utilitarian social welfare — sum of agent utilities.

Ideally, this can be achieved by an algorithm mimicking the proof of existence in Section 2: Given a set P of feasible models such that $\lambda(P) = \lambda^*$, we iteratively consider the agents; each agent removes a set $B \subseteq P$ from P such that $\lambda(B) = (1 - \delta) \cdot \lambda^*$ and this agent prefers every remaining model in P to every model in B. When this loop terminates, we choose the model with the highest welfare from the models remaining in P.

However, such an approach is often infeasible, as computing the set B in every iteration is intractable unless we impose more structure on the utility functions of the agents and the set of feasible models P. For instance, when P is a polytope and agents have linear utilities, the set B is always a polytope and the set $P \setminus B$ is also a polytope with polynomial description complexity (see Appendix B for more details).

Therefore, it is more practical to select a *representative subset* $\mathcal{P}$ of models from P that are (i) guaranteed to be good for many agents and (ii) the ranks of the models in $\mathcal{P}$ can be estimated through computationally efficient subroutines. Our algorithm for computing the representative set $\mathcal{P}$ builds on this idea; it is described below and in Algorithm 1.

Initial Selection. Algorithm 1 starts by the group of agents jointly training a model θ_0 by using a standard algorithm such as `FedAvg`. Note that no fairness guarantees have been imposed yet.

Choosing the Representative Set. We initially define our feasible set of models as $\mathcal{P} = \{\theta^0\}$. Subsequently, each agent samples J models, uniformly at random, along the direction of the gradient of their utility function at θ_0, i.e., for all $j \in [J]$, let $\xi_j \sim U(0, 1)$,

$$\theta_{i,j} = \theta_0 + \xi_j \cdot p_i \cdot \frac{\nabla_\theta u_i(\theta_0)}{\|\nabla_\theta u_i(\theta_0)\|_2},$$

Algorithm 1 Computes a representative set of P

Parameter: Number of models to sample per iteration J, norm bound of sampled gradients p
Output: Set of models $\mathcal{P} \subseteq P$
$\theta_0 \leftarrow$ Output of `FedAvg`
Determine p_i approximately through binary search.
for $j = 1, 2, \ldots, J$ **do**
  $\xi_j \sim U(0,1)$; {sample ξ uniformly}
  for $i = 1, 2, \ldots, n$ **do**
    $\theta_{i,j} \leftarrow \theta_0 + \xi_j \cdot p_i \cdot \frac{\nabla_\theta u_i(\theta_0)}{\|\nabla_\theta u_i(\theta_0)\|_2}$;
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{\theta_{i,j}\}$
  end for
end for

Algorithm 2 `Rank-Core-Fed`: Finds a θ that belongs to the proportional veto core of $\mathcal{P}$

Input: Server models $\mathcal{P} = \{\theta_1, \theta_2, \cdots, \theta_m\}$
Output: Model weights θ
for $i = 1, \ldots, n$ **do**
  Server sends $\mathcal{P}$ to agent i
  Agent i identifies $C_i \subseteq P$,by choosing the least preferred $m/n - 1$ models from $\mathcal{P}$
  Agent i sends C_i to server
  Server sets $\mathcal{P} \leftarrow \mathcal{P} \setminus C_i$
end for
Server chooses $\theta^* \in \mathcal{P}$ such that $\sum_{i \in [n]} u_i(\theta^*)$ is maximum

where p_i is defined as the largest p such that

$$\theta_0 + p \cdot \frac{\nabla_\theta u_i(\theta_0)}{||\nabla_\theta u_i(\theta_0)||_2} \in P,$$

and $U(0,1)$ denotes a uniform distribution on the interval $[0,1]$. Note that each p_i can be estimated up to an additive approximation of ε in $\mathcal{O}(\log(1/\varepsilon))$ time through binary search.

Intuitively, for each agent i, the models $\mathcal{R}_i = \{\theta_{i,j} : j \in [J]\}$ constitute the representative set of models, as these are the models near θ_0 that can improve the utility of agent i. We define $\mathcal{P}$ to be the union of the representative models of all agents around θ_0, i.e., $\mathcal{P} = \cup_{i \in [n]} \mathcal{R}_i$. Now, we have a finite set of models around θ_0 (a model with high utilitarian welfare), and our goal is to choose a model in the proportional veto core of $\mathcal{P}$.

Selecting a Proportional Veto Core Model. We now outline our algorithm, `Rank-Core-Fed`, which selects a model in the proportional veto core of $\mathcal{P}$. This is done through iterative elimination. In iteration i, agent i sends their set of least preferred $|\mathcal{P}|/n - 1$ models in $\mathcal{P}$, denoted C_i, to the server. The server updates $\mathcal{P} \leftarrow \mathcal{P} \setminus C_i$. At the end of the iteration, $\mathcal{P}$ is still non-empty, and the model θ^* with the highest utilitarian welfare is chosen from $\mathcal{P}$. The algorithm is formally presented as Algorithm 2.

Next, we prove that θ^* indeed lies in the (n/m)-PVC of $\mathcal{P}$, where $m = |\mathcal{P}|$; note that for any $\epsilon > 0$, we can choose a large enough m such that $n/m < \epsilon$.

Theorem 2. *Let $\mathcal{P} = \cup_{i \in [n]} \mathcal{R}_i$ and $|\mathcal{P}| = m$. Then the output model θ^* of* `Rank-Core-Fed` *is in the (n/m)-proportional veto core of $\mathcal{P}$.*

Proof. We show that for any coalition T of agents, the total number of models that all agents in T unanimously prefer over θ^* is at most $m(1 - |T|/n + |T|/m)$. The theorem then follows by using $\lambda(P) = |P|$ in Definition 1.

The proof of this claim follows the same structure as the proof of Theorem 1. Assume otherwise. Let T be a coalition, and let all agents in T prefer all models in B over θ^*, and $|B| > m(1 - |T|/n + |T|/m)$. Let $S = \{1, 2, \ldots, k\}$, where $k = \arg\max_{i \in T} i$.

Consider the time when the last agent in S, i.e., agent k, removes their share of models from $\mathcal{P}$. Observe that none of the agents in T remove any model from B as they all prefer all models in B over θ^*, and θ^* remains unremoved until the end of the algorithm. Thus, the number of models in B that are still unremoved is at least

$$\begin{aligned}
&\geq |B| - |S \setminus T| \cdot (m/n - 1) \\
&> m \cdot \left(1 - \frac{|T|}{n} - \frac{|S \setminus T|}{n}\right) + |S \setminus T| + |T| \\
&= m \cdot \left(1 - \frac{|S|}{n}\right) + |S| \\
&= m \cdot \left(1 - \frac{|S|}{n} + \frac{|S|}{m}\right)
\end{aligned}$$

This leads to a contradiction as the total number of models that have been removed is at most $|S|(\frac{m}{n} - 1)$, implying that the total number of remaining models is at most $m \cdot (1 - \frac{|S|}{n} + \frac{|S|}{m})$. □

4. Empirical Evaluation

We evaluate our algorithm `Rank-Core-Fed` on rotated MNIST (LeCun et al., 2010) and CIFAR-10 (Krizhevsky et al., 2009) datasets. We compare our approach with three baseline algorithms: `FedAvg` (McMahan et al., 2017), `CoreFed` (Chaudhury et al., 2022) and `q-FFL` (Li et al., 2020b). We show that models trained with `Rank-Core-Fed` achieve better rankwise proportionality guarantees for fairness than the state-of-the-art baselines and comparable utilitarian social welfare.

Table 1. FL instances with different heterogeneity levels for MNIST (10 agents). For CIFAR-10 (100 agents), we repeat the rotation list of MNIST 10 times for each setting so that the number of clusters remains the same.

FL Instance	Degree of Rotation for 10 Agents
I_{high}	0, 0, 20, 20, 40, 60, 100, 120, 180, 200 (8 clusters)
I_{mid}	0, 0, 0, 0, 0, 20, 20, 20, 40, 60 (4 clusters)
I_{low}	0, 0, 0, 0, 0, 0, 20, 20, 20, 20 (2 clusters)

4.1. Experiment Setup

FL Setting. We consider FL with 10 agents for MNIST and 100 agents for CIFAR-10. We introduce heterogeneity among agents by rotating images at different degrees, following the literature (Ghosh et al., 2020). We introduce 3 FL settings with different heterogeneity levels ($I_{\text{high}}, I_{\text{mid}}, I_{\text{low}}$) by controlling the number of clusters, where the agents with the same rotation degree belong to the same cluster.

Table 1 shows the rotation list of agents for different FL Instances. For all baselines and our algorithm, we set the number of iterations of the global model update to 50.

Models. For MNIST, we use a CNN, which has two 5×5 convolution layers followed by two fully connected layers with ReLU activation. For CIFAR-10, we evaluate with a more complex network, VGG11 (Simonyan & Zisserman, 2014). In all our experiments, we define agent utility as $M - \mathcal{L}_{\text{ce}}$, where $\mathcal{L}_{\text{ce}}$ refers to the average cross entropy loss on the agent's local test data. We set M to be 1.0 in our experiments.

Comparison. Let $\theta_{FA}, \theta_{CF}, \theta_{qF}$ be the final models returned by `FedAvg`, `CoreFed`, and `q-FFL` respectively. Let $\mathcal{P}$ denote the representative set computed by `Rank-Core-Fed` prior to choosing the PVC-stable model $\theta^* \in \mathcal{P}$. In Table 2, we compare the ranks of $\theta_{FA}, \theta_{CF}, \theta_{qF}$ and θ^* in the finite set $\mathcal{P} \cup \{\theta_{FA}, \theta_{CF}, \theta_{qF}\}$.

4.2. Experiment Results

We demonstrate that our distributed algorithm `Rank-Core-Fed` achieves better fairness and similar utilitarian social welfare (sum of utilities of the agents) compared with baselines, where fairness refers to the minimum over all agents of the rank of the final model with respect to the agent; we call this β.

We show the results in Table 2. We can see that `Rank-Core-Fed` achieves a higher β in all settings. In some cases, the differences are striking; for example, for CIFAR-10 and I_{low}, `Rank-Core-Fed` finds a model that each and every agent prefers to all but at most 5% of possible models, whereas for `FedAvg` and `CoreFed` this value is 12% and 11%, respectively.

Table 2. Comparison of fairness (β-rankwise proportionality) and utilitiarian social welfare ($\sum_{i\in[n]} u_i(\theta)$) on `Rank-Core-Fed` and baselines.

FL Instance	Method	MNIST		CIFAR-10	
		β	$\sum_{i\in[n]} u_i(\theta)$	β	$\sum_{i\in[n]} u_i(\theta)$
I_{high}	`FedAvg`	0.86	9.21	0.79	36.73
	`CoreFed`	0.87	9.23	0.81	36.69
	`q-FFL`	0.87	9.26	0.80	36.75
	`Rank-Core-Fed`	**0.89**	9.23	**0.85**	36.81
I_{mid}	`FedAvg`	0.88	9.52	0.83	37.21
	`CoreFed`	0.89	9.56	0.85	37.23
	`q-FFL`	0.88	9.52	0.89	37.21
	`Rank-Core-Fed`	**0.92**	9.63	**0.91**	37.19
I_{low}	`FedAvg`	0.91	9.62	0.88	37.29
	`CoreFed`	0.93	9.59	0.89	37.34
	`q-FFL`	0.93	9.65	0.91	37.32
	`Rank-Core-Fed`	**0.97**	9.70	**0.95**	37.41

5. Discussion

We believe our work is a step forward in understanding and realizing fairness in federated learning. Naturally, many questions remain open.

One interesting theoretical challenge is to investigate the design of provably efficient methods to compute PVC-stable models. Previous work (Ianovski & Kondratev, 2021) has addressed this question when the feasible set of models is discrete. However, existing algorithms do not extend to the setting with a continuous set of models as far as we know. To develop an initial intuition for this question, we present a case study in linear utility functions in Appendix B. Specifically, we design a *fully polynomial time randomized scheme (FPRAS)* to find a welfare-optimal, PVC-stable model when the utility functions are linear and the feasible set of models is defined by a polyhedron. We hope this will instigate further study of richer utility functions.

Another question is whether there exists an objective function that naturally captures PVC guarantees. In the context of concave utility functions, Chaudhury et al. (2022) show that the Nash product of the utilities organically implies core-stability. One could seek analogous functions in terms of the ranks of the agents that can directly imply the guarantees that we prove in the paper. We believe that such objective functions can also lead to better distributed learning algorithms that guarantee PVC-stability.

Impact Statement

The goal of this paper is to advance fairness in the field of machine learning. Potential societal consequences of our work include the design of fairer federated learning protocols. As with any notion of fairness, it is conceivable that PVC-stability is at odds with other criteria and therefore can sometimes lead to outcomes that are undesirable.

References

Chaudhury, B. R., Li, L., Kang, M., Li, B., and Mehta, R. Fairness in federated learning via core-stability. In *Thirty-sixth Conference on Neural Information Processing Systems*, 2022.

Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, A., Costa, A. B., Wood, B. J., Tsai, C.-S., et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature medicine*, 27(10):1735–1743, 2021.

Donahue, K. and Kleinberg, J. M. Models of fairness in federated learning. *CoRR*, abs/2112.00818, 2021.

Dyer, M. E., Frieze, A. M., and Kannan, R. A random polynomial time algorithm for approximating the volume of convex bodies. In *STOC*, pp. 375–381. ACM, 1989.

Elbir, A. M., Soner, B., and Coleri, S. Federated learning in vehicular networks. *arXiv preprint arXiv:2006.01412*, 2020.

Ghosh, A., Chung, J., Yin, D., and Ramchandran, K. An efficient framework for clustered federated learning. *Advances in Neural Information Processing Systems*, 33: 19586–19597, 2020.

Huang, T., Lin, W., Wu, W., He, L., Li, K., and Zomaya, A. Y. An efficiency-boosting client selection scheme for federated learning with fairness guarantee. *IEEE Trans. Parallel Distributed Syst.*, 32(7):1552–1564, 2021.

Ianovski, E. and Kondratev, A. Y. Computing the proportional veto core. In *AAAI*, pp. 5489–5496, 2021.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

LeCun, Y., Cortes, C., and Burges, C. MNIST handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2, 2010.

Li, T., Beirami, A., Sanjabi, M., and Smith, V. Tilted empirical risk minimization. *CoRR*, abs/2007.01162, 2020a.

Li, T., Sanjabi, M., Beirami, A., and Smith, V. Fair resource allocation in federated learning. In *ICLR*, 2020b.

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

Mo, J. and Walrand, J. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, 2000.

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

Moulin, H. The proportional veto principle. *The Review of Economic Studies*, 48(3):407–416, 1981.

Procaccia, A. D. Axioms should explain solutions. *The Future of Economic Design: The Continuing Development of a Field as Envisioned by Its Researchers*, pp. 195–199, 2019.

Shi, Y., Yu, H., and Leung, C. A survey of fairness-aware federated learning. *CoRR*, abs/2111.01872, 2021.

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Sze, V., Chen, Y.-H., Yang, T.-J., and Emer, J. S. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329, 2017.

Tao, T. *An Introduction to Measure Theory*. American Mathematical Society, 2021.

Wang, Z., Fan, X., Qi, J., Wen, C., Wang, C., and Yu, R. Federated learning with fair averaging. In *IJCAI*, pp. 1615–1623, 2021.

Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., and Wang, F. Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, 5(1):1–19, 2021.

Yang, M., Wang, X., Zhu, H., Wang, H., and Qian, H. Federated learning with class imbalance reduction. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pp. 2174–2178. IEEE, 2021.

Zhang, G., Malekmohammadi, S., Chen, X., and Yu, Y. Proportional fairness in federated learning. *Trans. Mach. Learn. Res.*, 2023, 2023.

Zhang, Z. and Sabuncu, M. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.

A. Additional Background on Measurability

We refer the reader to standard texts on measure theory (Tao, 2021) for the definition of a Lebesgue-measurable set. We show below that all sets and functions defined in the current work are Lebesgue-measurable (measurable for short).

Let us first state the definition of a measurable function.

Definition 4. Let $f : D \to \mathbb{R}$ be a function defined on a measurable set D. Then f is measurable if and only if any of the following statements hold.

(i) For every $r \in \mathbb{R}$ the set $\{x \in D : f(x) \geq r\}$ is measurable.

(ii) For every $r \in \mathbb{R}$ the set $\{x \in D : f(x) > r\}$ is measurable.

(iii) For every $r \in \mathbb{R}$ the set $\{x \in D : f(x) \leq r\}$ is measurable.

(iv) For every $r \in \mathbb{R}$ the set $\{x \in D : f(x) < r\}$ is measurable.

In our work, the only assumption we make on the utility functions $\{u_i(\cdot)\}_{i\in[n]}$ is that they are Lebesgue-measurable. With Definition 4, it is easy to see that the lower-level sets $L_i(\ell) = \{\theta : u_i(\theta) \geq \ell\}$ are measurable when the utility function u_i is measurable. In particular, the sets $P_i(\theta) = \{\theta' : u_i(\theta') \leq u_i(\theta)\} = L_i(u_i(\theta))$ are also measurable.

Measurability of the utility functions is a very mild assumption. In particular, any piece-wise continuous function is measurable, as we show below.

Proposition 2. *Let $f : D \to \mathbb{R}$ be a piece-wise continuous function defined on a measurable set D. Then f is Lebesgue-measurable.*

Proof. Define $f^{-1}(U) = \{x \in D : f(x) \in U\}$ to be the inverse image of U under f. One can show using Definition 4 that f is measurable iff $f^{-1}(U)$ is measurable for every open set U of $\mathbb{R}$.

We first prove the theorem for a continuous function f. Consider any open set U of $\mathbb{R}$, and take any $x \in f^{-1}(U)$. Since U is open, there is a small-enough neighborhood of $f(x)$ contained in U, i.e., $\mathcal{B}(f(x), \varepsilon) \subseteq U$ for some $\varepsilon > 0$[8]. By continuity of f, there is a neighborhood $\mathcal{B}(x, \varepsilon')$ s.t. $f(x') \in \mathcal{B}(f(x), \varepsilon)$ for all $x' \in B(x, \varepsilon')$. implying that $\mathcal{B}(x, \varepsilon') \subseteq f^{-1}(U)$. Thus for every $x \in f^{-1}(U)$, there is a small neighborhood of x contained in $f^{-1}(U)$, showing that $f^{-1}(U)$ is an open set, and hence is measurable. Thus, f is a measurable function.

Now suppose that f is piece-wise continuous. Then there is a partition of $\mathbb{R}$ into countably many intervals $X_1, X_2, \ldots$ s.t. f is continuous on each A_i. Let $Y_i = f(X_i)$. Then for any open set U, we have

$$f^{-1}(U) = f^{-1}(\cup_i (U \cap Y_i)) = \cup_i f^{-1}(U \cap Y_i).$$

Since f is continuous, $f^{-1}(U \cap Y_i)$ is measurable. Since countable union of measurable sets is measurable, $f^{-1}(U)$ is also measurable for any open set U, thus showing f is measurable. □

We also require the following lemma for the definition of Pareto optimality in Section 2.3.

Lemma 2. *Let $\{u_i(\cdot)\}_{i\in[n]}$ be Lebesgue-measurable utility functions let P be a non-empty Lebesgue-measurable set of feasible models. For a model θ and a subset $\emptyset \neq T \subseteq [n]$ of agents, let $B = \{\theta' \in P : \forall i \in T, u_i(\theta') \geq u_i(\theta)$ and $\exists h \in T, u_h(\theta') > u_h(\theta)\}$ be the set of models that Pareto-dominate θ for agents in T. Then B is Lebesgue-measurable.*

Proof. Let $B_i = \{\theta' \in P : u_i(\theta') \geq u_i(\theta)\}$, let $B'_i = \{\theta' \in P : u_i(\theta') > u_i(\theta)\}$, and $C_i = \{\theta' \in P : u_i(\theta') = u_i(\theta)\}$. Note that by Definition 4, B_i and B'_i are measurable. Since $C_i = B_i \setminus B'_i$ is the set difference of measurable sets where $B'_i \subseteq B_i$, C_i is also measurable. Now notice that $B = \bigcap_{i\in T} B_i \setminus \bigcap_{i\in T} C_i$. Since the intersection of finitely many measurable sets is measurable, $\bigcap_{i\in T} B_i$ and $\bigcap_{i\in T} C_i$ are measurable. Thus it follows that B is measurable since $\bigcap_{i\in T} C_i \subseteq \bigcap_{i\in T} B_i$ and both sets are measurable. □

Algorithm 3 Algorithm to compute a PVC-stable model for linear utilities

Input: $P = \{\theta : A\theta \leq b\}$, linear utilities $u_i(\cdot)$
Output: Model θ^* in the proportional veto core
Set $r = \frac{1}{n} - \frac{\varepsilon}{n+1}$
for $i = 1, \ldots, n$ **do**
  $\ell_0 \leftarrow 0, \ell_1 \leftarrow 1, \ell \leftarrow 0$
  while $\ell_0 \leq \ell_1$ **do**
    $\ell \leftarrow (\ell_0 + \ell_1)/2$
    if $\lambda(\{\theta \in Q : u_i(\theta) \leq \ell\}) < (r - \varepsilon)\lambda(P)$ **then**
      $\ell_0 = \ell$
    else if $\lambda(\{\theta \in Q : u_i(\theta) \leq \ell\}) > (r + \varepsilon)\lambda(P)$ **then**
      $\ell_1 = \ell$
    else
      Break
    end if
  end while
  $P_i \leftarrow \{\theta \in Q : u_i(\theta) \leq \ell\}$
  $Q \leftarrow Q \setminus P_i$
end for
Return $\theta^* \in Q$

B. Case Study in Linear Utilities

In this section we design an algorithm to compute a model in the proportional veto core when agents have linear utility functions and the space of models is a bounded polyhedron. Let $P = \{\theta \in \mathbb{R}^d | A\theta \leq b\}$ be the polyhedral feasible space of models. Let $u_i(\theta) = \sum_{j \in [d]} u_{ij} \cdot \theta_j$ be the linear utility function of agent $i \in [n]$.

Description of the Algorithm. Algorithm 3 iteratively computes the sets $P_1, \ldots, P_n$ described in the proof of Theorem 1. The variable Q maintains the set of remaining models after each iteration. That is, $Q = P$ initially, and after iteration i, $Q = P \setminus (\bigcup_{k \leq i} P_k)$. Recall that for each $i \in [n]$, we have $P \setminus (\bigcup_{k \leq i} P_k) \geq_i P_i$. For linear utilities, each level set $\{\theta : u_i(\theta) = \ell\}$ is a hyperplane in $\mathbb{R}^{d-1}$, and hence has measure zero in $\mathbb{R}^d$. Hence the sets P_i can be found using lower level sets of the form $\{\theta : u_i(\theta) \leq \ell\}$. For each $i \in [n]$, Algorithm 3 performs binary search on ℓ to find the appropriate value of ℓ such that the Lebesgue measure of the lower level set of u_i at ℓ is $r \cdot \lambda(P)$. This lower level set is assigned to be P_i. Algorithm 3 then updates the space of remaining models Q by deleting P_i. Note that since P is a polyhedron and all P_i are described by lower-level sets of linear functions, Q is also a polyhedron. Thus Q can simply be described by a list of linear inequalities. We note that the Lebesgue measure of polyhedrons encountered in Algorithm 3 is computable by using the algorithm of Dyer et al. (1989), which is a fully polynomial randomized approximation scheme (FPRAS). As the proof of Theorem 3 shows, the final set $Q = P \setminus (\bigcup_{i \leq n} P_i)$ is non-empty, and any model $\theta^* \in Q$ is in the proportional veto core.

[8] $\mathcal{B}(x, \varepsilon)$ is the open ball of radius ε centered at x.