Towards Socially and Economically Beneficial Machine Learning

Wenshuo Guo

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2022-272
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2022/EECS-2022-272.html

December 19, 2022

Towards Socially and Economically Beneficial Machine Learning

by

Wenshuo Guo

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael I. Jordan, Chair
Professor Peng Ding
Professor Nika Haghtalab

Fall 2022

Towards Socially and Economically Beneficial Machine Learning

Abstract

Towards Socially and Economically Beneficial Machine Learning

by

Wenshuo Guo

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Michael I. Jordan, Chair

From digital platforms, automated transportation to healthcare, the rapid deployment of machine learning has in many ways changed our everyday life. However, when learning systems are deployed in the real world, they immediately face a complex social and economic context which poses feasibility constraints, drives the underlying dynamics, and influences the kinds of data that the systems can actually obtain. Optimizing a single offline objective in isolation to these contexts can lead to severe unintended consequences at deployment and hinder the improvement of social welfare that the system has the potential to bring.

In this thesis, I summarize my research works on developing learning algorithms that incorporate such social economic contexts into the design from three aspects: (i) Learning with noisy input data; (ii) Learning with bandit-type user feedback; (iii) Learning under causal dynamics. I will situate each of these with the particular applications of machine learning on fair classification, resource allocation, auction and platform design.

To my parents

Contents

List of Figures

List of Tables

Acknowledgments

This thesis is impossible without the love and support from a group of people that I am forever lucky to have met before or throughout my PhD journey.

First and foremost, I would like to thank my advisor Michael I. Jordan. Mike has been a role model of mine who is not only able to build a successful research career, but also always keep being extremely respectful, passionate, encouraging, and full-of-ideas when interacting with students and other people around. I remember our first face-to-face conversation during the EECS Visit Days in a meeting room in Soda: I described my interest in working on "learning in complex systems" and expressed my excitement to learn more in optimization and statistics. That was a super vague description and I was actually very nervous; but Mike was very warm, very encouraging, and gave me a full assurance that those were great thoughts and I would certainly be able to work out something great and have peers with shared interests in his group. That kind of strong encouragement, doubtlessness, and support has been always there during my entire PhD, whenever I talked to him. Mike's support made it possible for me to explore many research topics that I wanted to learn about and work on at the interconnections of machine learning, economics, and society. From all the interactions with Mike, I also learnt that good research needs to be motivated by true passion and curiosity, and the importance of creativity, respectfulness, open-mindedness and collaborations in research. Thank you Mike for all the guidance and support to me.

I would like to express my gratitude to Christian Borgs, Peng Ding, Nika Haghtalab, Ben Recht and Jacob Steinhardt for kindly serving on my preliminary exam committee, qualification exam committee, and thesis committee. Thank you Christian for the many joyful chats with me about my research and providing valuable feedback to me. Thank you Peng for teaching me my first formal course on causal inference and working together on my first research project related to that. Thank you Nika for introducing me to many interesting works at the intersection of learning theory and economics, and always being able to provide critical ideas and timely suggestions on our project. Thank you Ben for introducing me to Modest Yachts where I started one of my first research projects in grad school and had so much fun interacting with this group of lovely people over the years. Thank you Jacob for having me in group meetings and gatherings where I had the most energetic discussions on half-baked research ideas, and for having many research discussions with me when I was at a time trying to figure out new directions.

I am very grateful to work together with a group of phenomenal collaborators throughout my PhD, many of whom also served as my mentors. I would like to thank: Rediet Abebe, Kumar Krishna Agarwal, David Alvarez-Melis, Stephen Bates, Romil Bhardwaj, Kush Bhatia, Asim Biswal, Derek Cheng, Andrew Cotter, Mihaela Curmei, Sarah Dean, Peng Ding, Meryem Essaidi, Alex Fang, Nikhil Garg, Matthew Fahrbach, Sara Fridovich-Keil, Joseph E. Gonzalez, Aditya Grover, Maya Gupta, Nika Haghtalab, Benjamin Hindman, Nhat Ho, Michael I. Jordan, Kirthevasan Kandasamy, Wang-Cheng Kang, Sai Praneeth Karimireddy, Karl Krauth, Aditi Krishnapriyan, Tianyi Lin, Celestine Mendler-Dünner, Vidya Muthukumar, Harikrishna Narasimhan, Ashwin Pananjady, Jonathan Ragan-Kelly, Ben Recht, An-

drew Rosen, Daniel Rothchild, Ludwig Schmidt, Vaishaal Shankar, Jacob Steinhardt, Ion Stoica, Sam Taggart, Eric Taw, Ellen Vitercik, Ruoxi Wang, Serena Wang, Yixin Wang, Mingzhang Yin, Manolis Zampetakis, Alex Zhao, Angela Zhou, Yichen Zhou.

I would not have enjoyed my PhD journey with the accompany of my friends at Berkeley. Thank you to my friends and research group peers in SAIL, the Foundations of Machine Learning cohort, the EECS Theory Group and the BAIR community. Thank you to my office mates in Soda 523 for all the jigsaw puzzles we solved. Thank you Patrick Hernan, Jean Nguyen and Shirley Salanio for making many of my applications' processes so much easier. Thank you Kattt Atchley, Kosta Ilov, Jon Kuroda and Naomi Yamasaki for always be there providing research support to me across SAIL, ADEPT, Sky Computing Lab and BAIR. Thank you to Berkeley Graduate Fellowship and Google PhD Fellowship for providing funding for my PhD study.

Before Berkeley, I would like to thank Professors Szeto Kwok Yip, Dit-Yan Yeung, and Mordecai J. Golin for mentoring me through my undergraduate research, which led me to pursue further in grad school. I owe my special thanks to Professor Szeto Kwok Yip. Thank you for welcoming me to your group where I started my very first formal research project in statistical physics as a freshman and developed more and more passion in research after that. Thank you for all the advice, thought-provoking discussions, encouragement and care. I will miss you forever.

Finally, I would like to thank my mom and dad for their unconditional love and support. Thank you for being my role models of many good qualities that a person can have. Thank you for always believing in me. You are the best parents in the world. This thesis is dedicated to you.

Chapter 1

Introduction

From digital platforms, automated transportation to healthcare, the rapid deployment of machine learning has in many ways changed our everyday life. However, when learning systems are deployed in the real world, they immediately face a complex social and economic context which poses feasibility constraints, drives the underlying dynamics, and influences the kinds of data that the systems can actually obtain. Optimizing a single offline objective in isolation to these contexts can lead to severe unintended consequences at deployment and hinder the improvement of social welfare that the system has the potential to bring.

In this thesis, I summarize my research works on developing learning algorithms that incorporate such social economic contexts into the design from three aspects: (i) Learning with noisy input data; (ii) Learning with bandit-type user feedback; (iii) Learning under causal dynamics.

1.1 Robust Learning with Noisy Input Data

Real-world data almost always suffers from noise, and potential corruptions from adversaries. The first part of this thesis is focused on developing robust learning algorithms towards input noise and corruptions.

In Chapter 2, we studied the learning of revenue-optimal auctions for multiple bidders, in a setting in which the samples can be corrupted adversarially [101]. We first consider the information-theoretic limit in a population model, assuming exact knowledge of the adversarially perturbed valuation distribution. We develop a theoretical algorithm which obtains a tight upper bound on the revenue for the MHR and regular distributions, obtaining the information-theoretic limit of the robustness guarantee. We then relax the population model and derive sample complexity bounds for learning optimal auctions from samples. We propose a practical algorithm which takes the corrupted samples as input, and provide the sample complexity upper bounds for the MHR distribution case and the single-bidder regular distribution case. We also provide accompanying sample complexity lower bounds, and demonstrate a small gap relative to the corresponding upper bounds.

In Chapter 3, we studied robust optimization algorithms in another context: fair classifications [217]. In particular, many existing fairness criteria for machine learning involve equalizing some metric across *protected groups* such as race or gender. However, practitioners trying to audit or enforce such group-based criteria can easily face the problem of noisy or biased protected group information. In this work, we study the consequences of naïvely relying on noisy protected group labels: we provide an upper bound on the fairness violations on the true groups G when the fairness criteria are satisfied on noisy groups $\hat{G}$. We introduce two new approaches using robust optimization that, unlike the naïve approach of only relying on $\hat{G}$, are guaranteed to satisfy fairness criteria on the true protected groups G while minimizing a training objective. We provide theoretical guarantees that one such approach converges to an optimal feasible solution. Using two case studies, we show empirically that the robust approaches achieve better true group fairness guarantees than the naïve approach.

1.2 Online Learning with Bandit-Type User Feedback

Machine learning systems in the real world are constantly interacting with users. In the second part of this thesis, we focus on developing online learning algorithms which learn through the repeated interactions with users and feedback in an online environment.

In Chapter 4, we study online algorithms in the context of resource allocation in distributed systems [104]. In particular, the sharing of scarce resources among multiple rational agents is one of the classical problems in economics. In exchange economies, which are used to model such situations, agents begin with an initial endowment of resources and exchange them in a way that is mutually beneficial until they reach a competitive equilibrium (CE). The allocations at a CE are Pareto efficient and fair. Consequently, they are used widely in designing mechanisms for fair division. However, computing CEs requires the knowledge of agent preferences which are unknown in several applications of interest. In this work, we explore a new online learning mechanism, which, on each round, allocates resources to the agents and collects stochastic feedback on their experience in using that allocation. Its goal is to learn the agent utilities via this feedback and imitate the allocations at a CE in the long run. We quantify CE behavior via two losses and propose a randomized algorithm which achieves sublinear loss under a parametric class of utilities. Empirically, we demonstrate the effectiveness of this mechanism through numerical simulations.

In Chapter 5, we develop no-regret algorithms in the context of auction platforms where the seller and buyer interact over multiple rounds [102]: Auctions with partially-revealed information about items are broadly employed in real-world applications, but the underlying mechanisms have limited theoretical support. In this work, we study a machine learning formulation of these types of mechanisms, presenting algorithms that are no-regret from the buyer's perspective. Specifically, a buyer who wishes to maximize his utility interacts repeatedly with a platform over a series of T rounds. In each round, a new item is drawn from an unknown distribution and the platform publishes a price together with incomplete,

"masked" information about the item. The buyer then decides whether to purchase the item. We formalize this problem as an online learning task where the goal is to have low regret with respect to a myopic oracle that has perfect knowledge of the distribution over items and the seller's masking function. When the distribution over items is known to the buyer and the mask is a SimHash function mapping $\mathbb{R}^d$ to $\{0,1\}^\ell$, our algorithm has regret $\tilde{\mathcal{O}}((Td\ell)^{1/2})$. In a fully agnostic setting when the mask is an arbitrary function mapping to a set of size n and the prices are stochastic, our algorithm has regret $\tilde{\mathcal{O}}((Tn)^{1/2})$.

1.3 Causal Learning with Optimization-Dependent Responses

In Chapter 6, we investigate in the intersection of causal inference and machine learning. In particular, the intersection of causal inference and machine learning for decision-making is rapidly expanding, but the default decision criterion remains an *average* of individual causal outcomes across a population. In practice, various operational restrictions ensure that a decision-maker's utility is not realized as an *average* but rather as an *output* of a downstream decision-making problem (such as matching, assignment, network flow, minimizing predictive risk). In Guo et al. [103], we develop a new framework for off-policy evaluation with *policy-dependent* linear optimization responses: causal outcomes introduce stochasticity in objective function coefficients. Under this framework, a decision-maker's utility depends on the policy-dependent optimization, which introduces a fundamental challenge of *optimization* bias even for the case of policy evaluation. We construct unbiased estimators for the policy-dependent estimand by a perturbation method, and discuss asymptotic variance properties for a set of adjusted plug-in estimators. Lastly, attaining unbiased policy evaluation allows for policy optimization: we provide a general algorithm for optimizing causal interventions. We corroborate our theoretical results with numerical simulations.

Chapter 2

Robust Learning of Optimal Auctions

2.1 Introduction

Arguably the fundamental difficulty in the design of optimal auctions is that real valuations are private and unknown to the auction designer. Consider specifically the problem of selling one item to multiple buyers. Suppose that we model the buyers' valuations as arising as independent draws from buyer-specific prior distributions. In this scenario, what is the optimal mechanism in terms of the expected revenue? This problem was solved by Myerson [171] through a characterization of *virtual value functions*. In particular, we can define a virtual value function of each buyer based on their prior distributions. An optimal auction then lets the buyer with the largest non-negative virtual value win the item, and charges the winner a price that equals the threshold value above which she wins.[1]

Unfortunately, there is a further fundamental challenge in deploying these theoretical results in practice, which is that in real-world settings the auction designer may not even know the prior distributions on valuations. Instead, what the designer might hope for is that there is a stream of previous transactions, or some other relevant auxiliary data, that is helpful in inferring the buyers' private distributions. This perspective has motivated an active recent literature learning optimal auctions from samples [22, 52, 65, 75, 95, 96, 100, 116, 166, 167, 186, 187, 201]. In this line of work, the central question is: suppose we are only able to access the prior distributions in the form of independent samples, how many samples are sufficient and necessary for finding an approximately optimal auction?

While this merging of mechanism design and learning theory is appealing, a further concern arises. Given the potentially adversarial setting of auction design, do we really believe that the data that we observe are drawn in accord with our assumptions? More concretely, is the learning of optimal auctions robust to adversarial corruptions of the samples? This problem is arguably at the core of what it means to learn an optimal auction. It is a challenging problem; indeed, as we show in Counterexample 1 in Section 2.4, auction designs

[1] More generally, the optimal auction picks the winner based on the virtual value after an "ironing" procedure.

that are optimal in the absence of corruptions can become arbitrarily bad even if a small portion of the samples are corrupted. Building on earlier work by Cai and Daskalakis [44] and Brustle et al. [39], we tackle a key open problem—what is the best approximation to the optimal revenue for arbitrary levels of corruption for distributions with unbounded support? And what is the mechanism that achieves it?

In summary, in this work we explore the problem of the robust learning of optimal auctions, where the samples of bidders' valuations are subject to corruption and their support is unbounded. In particular, we consider having access to samples that are drawn from some distribution $\tilde{\mathcal{D}}$ which is within a Kolmogorov-Smirnov (KS) distance α of the true distribution $\mathcal{D}^*$. Denote OPT as the maximum revenue we can achieve under the true valuation distributions. Our goal is to design mechanisms that are guaranteed to achieve a revenue of at least $(1 - \rho(\alpha)) \cdot \text{OPT}$ for the smallest possible error $\rho(\alpha)$ and with the use of a minimal number of samples.

Our results

We study the problem of learning revenue-optimal multi-bidder auctions from samples when the samples of bidders' valuations can be adversarially corrupted or drawn from distributions that are adversarially perturbed. We summarize our main results as follows:

1. We derive tight upper bounds on the revenue we can obtain with a corrupted distribution under a population model. For distributions with monotone hazard rate (MHR), and with total corruption α, we obtain an approximation ratio of $1 - O(\alpha)$ compared to the optimal revenue under the true distribution (see Theorem 2.3.6). For regular valuation distributions, where for total corruption α, we get an approximation ratio of $1 - O(\sqrt{\alpha})$ (see Theorem 2.3.8).

2. To achieve these upper bounds, we propose a new *theoretical* algorithm for the population model (see Algorithm 1) that, given only an "approximate distribution" for the bidder's valuation, can learn a mechanism whose revenue is nearly optimal simultaneously for all "true distributions" that are α-close to the given distribution in Kolmogorov-Smirnov distance. The proposed algorithm operates beyond the setting of bounded distributions that have been studied in prior works; indeed, they apply to general unbounded MHR and regular distributions.

3. We further show that these upper bounds under the population model cannot be further improved (up to constant log factors), by providing matching lower bounds for both the MHR and regular distributions (see Theorem 2.3.7 and Theorem 2.3.9).

4. Lastly, we derive sample complexity upper bounds for learning a near-optimal auction for both MHR and regular distributions with multiple bidders (Theorem 2.4.3 and Theorem 2.4.4), and propose a *practical* algorithm (see Algorithm 2) which takes samples as input. We also provide accompanying sample complexity lower bounds

(Theorem 2.4.5), and demonstrate a small gap relative to the corresponding upper bounds which is of interest for future work.

Related work

Designing revenue optimal auctions is a classic problem in economic theory that has attracted much research attention. We survey the most closely related work in two main areas.

Learning optimal auctions from samples. Recent work has explored settings of learning approximately optimal auction from samples, both for single-item auctions [52], and multi-item auctions [21, 22, 166, 201]. Most recently, Guo et al. [100] provide a complete set of sample complexity bounds for single-item auctions, by deriving matching upper and lower bounds up to a poly-logarithmic factor. While these approaches have obtained fruitful results on the sample complexity of learning optimal auctions, a key assumption that is commonly made in this work is that the samples are independently and identically drawn from the bidders' valuation distributions, with the goal of learning an auction which maximizes the expected revenue on the underlying, unknown distribution over bidder valuations. A major difference in our work is that we consider that the samples can suffer from potential corruptions, which is a significantly more challenging setting.

Robustness of learning optimal auctions. Our paradigm on the robust learning of optimal auctions is closely related to recent work that considers the learning of auctions from mismatched distributions or corrupted samples. Cai and Daskalakis [44] consider a multi-item auction setting, where there is a given "approximate distribution," and the goal is to compute an auction whose revenue is approximately optimal simultaneously for all "true distributions" that are close to the given one. They provide an algorithm that achieves a poly-α additive loss compared to the true optimal revenue. More recently, Brustle et al. [39] consider learning multi-item auctions where bidders' valuations are drawn from correlated distributions that can be captured by Markov random fields. However, they make a key simplifying assumption—that the bidders' valuation for the items lie in some bounded interval. Our results, by contrast, apply to the general setting of unbounded valuation distributions, a setting that requires new theoretical machinery. To the best of our knowledge, our work constitutes the first analysis of the learnability of single-item optimal auctions from corrupted samples for unbounded distributions.

Organization. In Section 2.2, we provide background on auction models and formally state our problem. Section 2.3 contains our main theoretical statements for the population model. We propose an algorithm that achieves optimal theoretical upper bounds, by providing matching lower bounds. Section 2.4 contains our main results on learning with finite samples. We provide a practical algorithm that takes samples from the corrupted distribu-

tion, and provides sample complexity upper and lower bounds for both the regular and MHR distributions cases. We conclude in Section 2.5.

2.2 Preliminaries

We begin by formally defining the setting we study for robust learning of optimal auctions, which includes the revenue objective and the general classes of valuation distributions that we consider.

Auction models

Single-bidder setting. Consider one item for sale to one bidder. The bidder has a private valuation $v \in \mathbb{R}_+$ for this item. We assume that v is a random variable distributed according to the distribution $\mathcal{D}^*$, with support $\mathbb{R}_+$, cumulative distribution function F, and probability density function f.

It is well known that the optimal auction in this setting is a reserve price auction, such that the task for the seller is to compute a reserve price p that optimizes revenue [171]. We assume that the bidder has a quasi-linear utility that is equal to $u(p) = v - p$ if she decides to buy the item and $u(p) = 0$ otherwise. The seller aims to set p such that her expected revenue—i.e., the received payment—is maximized. We consider the setting where both v and $\mathcal{D}^*$ are unknown to the seller. However, the seller can access i.i.d. samples that are drawn from a distribution $\tilde{\mathcal{D}}$, which is α-close to $\mathcal{D}$ with regard to the Kolmogorov distance:

Definition 2.2.1. (Kolmogorov-Smirnov distance) For probability measures μ and ν on $\mathbb{R}$, define

$$d_k(\mu, \nu) = \sup_{x \in \mathbb{R}} |\mu((-\infty, x)) - \nu((-\infty, x))|.$$

It is well known that $d_k(\mu, \nu) \leqslant d_{TV}(\mu, \nu)$, where d_{TV} denotes the total variation (TV) distance between μ and ν. The closeness of $\tilde{\mathcal{D}}$ to $\mathcal{D}^*$ is thus formalized as follows:

$$d_k(\mathcal{D}^*, \tilde{\mathcal{D}}) \leqslant \alpha,$$

for some $\alpha > 0$.

Multi-bidder setting. Consider one item for sale to n bidders. Each bidder has a private valuation, $v_i \in \mathbb{R}_+$, where v_i is independently drawn from the corresponding prior distribution $\mathcal{D}_i^*$. Thus, the valuations $\mathbf{v} = (v_1, v_2, \cdots, v_n)$ follow a product distribution $\mathbf{D}^* = \mathcal{D}_1^* \times \cdots \times \mathcal{D}_n^*$. Each bidder submits a bid $b_i \geqslant 0$. Denote all the bids as $\mathbf{b} = (b_1, \cdots, b_n)$. A mechanism in this setting consists of two rules: the allocation rule $\mathbf{x}(\mathbf{b})$ that takes the bids $\mathbf{b}$ and outputs the probability $x_i(\mathbf{b})$ that each bidder i will receive the item, and the payment rule $\mathbf{p}(\mathbf{b})$ that takes the bids $\mathbf{b}$ and outputs the payment of bidder i. Bidder i's utility is then $u_i(\mathbf{b}) = v_i \cdot x_i(\mathbf{b}) - p_i(\mathbf{b})$. The goal of the seller is to find a mechanism that maximizes the expected revenue $\mathbb{E}[\sum_{i \in [n]} p_i(\mathbf{b})]$, where the expectation is over $\mathbf{v} \sim \mathbf{D}^*$,

under the following *Dominant Strategy Incentive Compatibility (DSIC)* and the *Individual Rationality (IR)* constraints:

$$u_i(v_i, \mathbf{b}_{-i}) \geqslant u_i(b_i, \mathbf{b}_{-i}) \qquad \text{for all } v_i, b_i \in \mathbb{R}_+ \text{ and all } \mathbf{b}_{-i} \in \mathbb{R}_+^{n-1} \qquad \text{(DSIC)}$$
$$u_i(v_i, \mathbf{b}_{-i}) \geqslant 0 \qquad \text{for all } v_i \in \mathbb{R}_+ \text{ and all } \mathbf{b}_{-i} \in \mathbb{R}_+^{n-1}. \qquad \text{(IR)}$$

We consider the setting in which the valuations and the prior distributions are unknown to the seller. Instead, the seller has access to a finite number of i.i.d. samples drawn from the product distribution $\tilde{\mathcal{D}} = \tilde{\mathcal{D}}_1 \times \cdots \times \tilde{\mathcal{D}}_n$, where each $\tilde{\mathcal{D}}_i$ satisfies

$$d_k(\mathcal{D}_i^*, \tilde{\mathcal{D}}_i) \leqslant \alpha_i,$$

for some $\alpha_i > 0, \forall i \in [n]$.

Revenue objective. Letting $\mathbf{D}$, $\mathbf{D}'$ be product or single bidder distributions as described above, we define $M_{\mathbf{D}}$ to be the mechanism that achieves the optimal revenue for the value distributions $\mathbf{D}$ and $\mathrm{OPT}(\mathbf{D})$ its expected revenue. Let also $\mathrm{Rev}(M_{\mathbf{D}}, \mathbf{D}')$ be the expected revenue of the mechanism $M_{\mathbf{D}}$ when applied to a setting where the values are drawn with respect to $\mathbf{D}'$.

Monotone hazard rate (MHR) and regular distributions

For any bidder i with a valuation $v_i \sim \mathcal{D}_i$, define the *virtual value function* for this bidder as $\phi_i(v) \stackrel{\text{def}}{=} v - \frac{1-F_i(v)}{f_i(v)}$, where F_i and f_i are the CDF and PDF of $\mathcal{D}_i$. The *hazard rate* of the distribution $\mathcal{D}_i$ is defined as the function $\frac{f_i(v)}{1-F_i(v)}$. Then, the distribution $\mathcal{D}_i$ is said to be *regular* if the virtual value $\phi_i(v)$ is monotonically non-decreasing in v. Further, distribution $\mathcal{D}_i$ has *monotone hazard rate* (MHR) if $\frac{f_i(v)}{1-F_i(v)}$ is monotone non-decreasing.

2.3 The population model

In this section, we study the problem of learning optimal auction assuming that we have the exact knowledge of the adversarially perturbed distributions $\tilde{\mathbf{D}}$. We relax this assumption in Section 2.4 where we show how to learn optimal auctions when we only have sample access to $\tilde{\mathbf{D}}$.

We begin in Section 2.3 with the description of our mechanism in the population model. Then, in Section 2.3, we present our analysis for the population mechanism for Monotone Hazard Rate distributions and we also present the sketch of our proof for the single-bidder case. Similarly, in Section 2.3 we state our analysis for the population mechanism for regular distributions and we present a proof sketch for the single-bidder case. Finally, we show that our proposed mechanism achieves optimal (up to constants) guarantees among any mechanism in the population model.

Robust Myerson auction in the population model

Our algorithm assumes as an input the exact knowledge of a product distribution, $\tilde{\mathbf{D}} = \tilde{\mathcal{D}}_1 \times \cdots \times \tilde{\mathcal{D}}_n$, such that the $d_k(\mathcal{D}_i^*, \tilde{\mathcal{D}}_i) \leqslant \alpha_i$ and its goal is to find a mechanism that achieves approximately optimal revenue for $\mathbf{D}^*$, where $\mathbf{D}^* = \Pi_i \mathcal{D}_i^*$. Without further assumptions, this is an impossible task, as we explain in Section 2.4 via an example. Thus we assume that the algorithm possesses some additional knowledge regarding $\mathcal{D}_i^*$, either that it is MHR or regular, and the mechanism needs to exploit this additional property.

To utilize the additional property of the distributions $\mathcal{D}_i^*$, our mechanism uses the important concept of the *link function* for MHR and regular distributions.

Definition 2.3.1 (Link Function)**.** The link function $h_M(x; F)$ for MHR distributions is defined as $h_M(x; F) = -\ln(1-F(x))$ and the link function $h_r(x; F)$ for regular distributions is defined as $h_r(x; F) = 1/(1-F(x))$. We also define the corresponding inverse link functions $h_M^{-1}(x; h) = 1 - \exp(-h(x))$ and $h_r^{-1}(x; h) = 1 - 1/h(x)$. Observe that $h_M^{-1}(x; h_M(x; F)) = F(x)$ and $h_r^{-1}(x; h_r(x; F)) = F(x)$. We may write $h_M(x)$ or $h_r(x)$ when F is clear from the context.

We provide some intuition on the link function. First, by construction, the link function of either an MHR distribution or a regular distribution is convex and non-decreasing. Second, the link function is monotone with regard to F. These two properties are important when we define the notion of a minimal MHR/regular distribution in a Kolmogorov ball, momentarily, which will be used as a necessary step in our algorithm.

Importantly, the link function provides a convenient characterization of the optimal reserve price and optimal revenue for a distribution F that is MHR or regular. To see this, first consider a single bidder with a valuation distribution F. Denote the optimal reserve price for selling one item to her as x^*, and the optimal expected revenue as $\text{OPT}(F)$. Then, when F is MHR, we show that x^* is also the unique minimizer of $(h_M(x) - \log(x))$. On the other hand, when F is regular, v^* is the point where $h_r(x)$ intersects with its tangent line kx, with $k = 1/\text{OPT}(F)$ (proof details in Appendix). Figure 2.1 illustrates such a useful property for h_M and h_r explicitly, for a single-item, single-bidder auction.

Next, we formally define stochastic dominance between two distributions, and state the property of strong revenue monotonicity.

Definition 2.3.2 (Stochastic dominance)**.** Given two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ with CDFs as F_1 and F_2. Then, we say $\mathcal{D}_1$ (first-order) stochastically dominates $\mathcal{D}_2$ if for every $x \in \mathcal{X}$,

$$F_1(x) \leqslant F_2(x),$$

denoted as $\mathcal{D}_1 \succeq \mathcal{D}_2$. We say a product distribution $\mathbf{D} = \Pi_i \mathcal{D}_i$ (component-wise) stochastically dominates another product distribution $\mathbf{D}' = \Pi_i \mathcal{D}_i'$ if for every i, we have $\mathcal{D}_i \succeq \mathcal{D}_i'$.

Lemma 2.3.3 (Strong revenue monotonicity [100])**.** *Let $\mathbf{D}$, $\mathbf{D}'$ be two product distributions such that $\mathbf{D}' \succeq \mathbf{D}$, then, for M that is the optimal mechanism for $\mathbf{D}$, we have:*

$$\text{Rev}(M, \mathbf{D}) \leqslant \text{Rev}(M, \mathbf{D}').$$

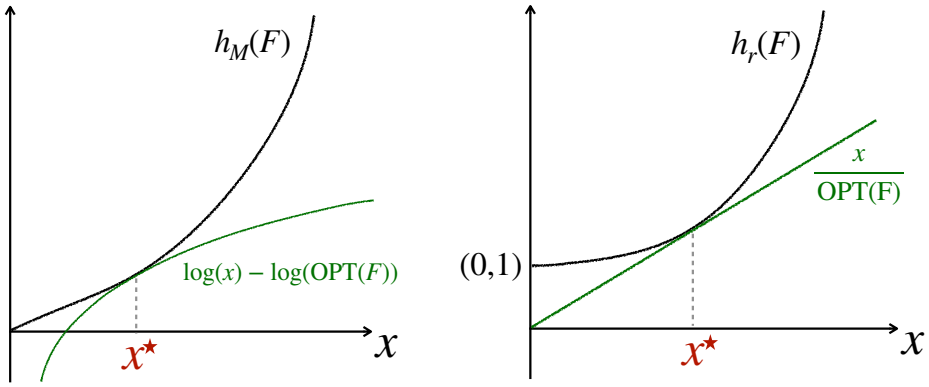


Figure 2.1: Optimal reserve price x^* with regard to the link function, for a single-item single-bidder auction with a valuation distribution F. *(left)* F is MHR; *(right)* F is regular.

Algorithm 1 Robust Myerson Auction in the Population Model

1: **Input:** $\alpha_1 \dots \alpha_n > 0$, link function $h(\cdot)$, possibly corrupted valuation distribution $\tilde{F} = \Pi_{i=1}^n \tilde{F}_i$.
2: **for** i = 1...n **do**
3: Compute a minimal regular / MHR distribution in $B_{d_k,\alpha_i}(\tilde{F}_i)$ according to Eq (2.1), denote as $\hat{F}_i$.
4: **end for**
5: Set $\hat{F} = \Pi_{i=1}^n \hat{F}_i$.
6: Output Myerson's optimal auction $M_{\hat{F}}$ w.r.t. the distribution $\hat{F}$.

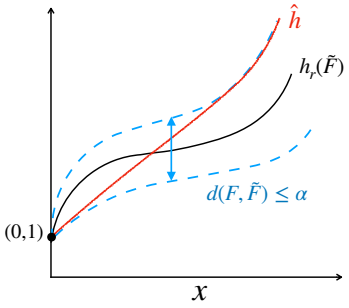


Figure 2.2: A minimal regular distribution in $B_{d_k,\alpha}$, in the space transformed by applying the link function.

The following lemma illustrates the importance of the link functions as well as their connection with first-order stochastic dominance. The proof of this lemma is given in Appendix 2.6.

Lemma 2.3.4. *A distribution with CDF F is MHR if and only if $h_M(x;F)$ is a convex function of x. Similarly, F is regular if and only if $h_r(x;F)$ is a convex function of x. Moreover, for two MHR (resp. regular) distributions F_1 and F_2, such that $F_1 \succeq F_2$, we have that $h_M(x;F_1) \leqslant h_M(x;F_2)$ (resp. $h_r(x;F_1) \leqslant h_r(x;F_2)$) for all x.*

A key idea used in our algorithm is the minimal MHR/regular distribution within a Kolmogorov distance divergence ball. Formally,

Definition 2.3.5. For a given distribution with its cumulative distribution function as F, denote the set of all the distributions that are α-close to F in Kolmogorov distance as $B_{d_k,\alpha}(F)$:

$$B_{d_k,\alpha}(F) \stackrel{\text{def}}{=} \{F' : d_k(F', F) \leqslant \alpha\}.$$

Further, define a minimal MHR/regular distribution within $B_{d_k,\alpha}(F)$ as:

$$\hat{F}(x) = h^{-1}(x; \hat{h}), \quad \text{where} \quad \hat{h}(x) \stackrel{\text{def}}{=} \max_{\substack{\tilde{F} \in B_{d_k,\alpha}(F) \\ \tilde{F} \text{ is MHR / regular}}} h\left(\tilde{F}(x)\right) \quad \forall x \in \mathbb{R}_+. \tag{2.1}$$

Figure 2.2 gives an illustration of a minimal regular distribution within $B_{d_k,\alpha}(F)$, in the space transformed by the link function of regular distributions.

Analysis for MHR distributions

In this section we state the results for the performance of Algorithm 1 for MHR distributions and we provide a proof sketch for the single-bidder case. The full proof of the following theorem can be found in Appendix 2.7.

Theorem 2.3.6. *Let $\mathbf{D}^* = \mathcal{D}_1^* \times \cdots \times \mathcal{D}_n^*$ be a product distribution where every $\mathcal{D}_i^*$ is MHR. Let also $\tilde{\mathbf{D}} = \tilde{\mathcal{D}}_1 \times \cdots \tilde{\mathcal{D}}_n$ be any product distribution such that for all $i \in [n]$ it holds that $d_k(\mathcal{D}_i^*, \tilde{\mathcal{D}}_i) \leqslant \alpha_i$. If $\tilde{M}$ is the mechanism that Algorithm 1 outputs with input $\tilde{\mathbf{D}}$ then it holds that*

$$\mathrm{Rev}(\tilde{M}, \mathbf{D}^*) \geqslant \left(1 - \tilde{O}\left(\sum_{i=1}^{n} \alpha_i\right)\right) \cdot \mathrm{OPT}(\mathbf{D}^*).$$

In particular for $n = 1$, if $\alpha = \alpha_1$, then we have that $\mathrm{Rev}(\tilde{M}, \mathcal{D}^*) \geqslant (1 - O(\alpha)) \cdot \mathrm{OPT}(\mathcal{D}^*)$.

Proof sketch for $n = 1$. The first key step in our proof is the observation that, by construction, Algorithm 1 runs the Myerson optimal auction on an MHR distribution $\hat{F}$, such that $\hat{F}$ is stochastically dominated by any other MHR distribution that is within $B_{d_k,\alpha}(\tilde{F}')$. On the other hand we have $d_k(F^*(x), \tilde{F}(x)) \leqslant \alpha$. Applying the triangle inequality, we have $d_k(F^*(x), \hat{F}(x)) \leqslant 2\alpha$. It is then sufficient for us to bound the ratio of the optimal revenue for any two MHR distributions F_1 and F_2, with $d_k(F_1, F_2) \leqslant 2\alpha$, and where F_1 is stochastically dominated by F_2.

The key part of our proof then considers such F_1, F_2, and due to the fact that the ratio of the revenues, $\mathrm{OPT}_{F_1}/\mathrm{OPT}_{F_2}$, is scale invariant, we assume without loss of generality that $\mathrm{OPT}_{F_1} = 1$. We then prove that this leads to $h(P^*_{F_1}) \leqslant 1$. The result then follows from two further key lemmas. First, for any reserve price $x < P^*_{F_1}$, $|h_1(x) - h_2(x)| = \left|\log\left(\frac{1-F_2(x)}{1-F_1(x)}\right)\right|$. Further applying the fact that by assumption $|F_1(x) - F_2(x)| \leqslant \alpha$ we show that $|h_1(x) - h_2(x)| = O(\alpha)$ for any reserve price $x < P^*_{F_1}$. Second, using the fact that F_1 is stochastically dominated by F_2, we derive that $P^*_{F_2} \leqslant P^*_{F_1}$. The conclusion then follows

from bounding the ratio of $s_1(x) = h_1(x) - \log(x)$, and $s_2(x) = h_2(x) - \log(x)$, based on the definition of $P^*_{F_1}$ and $P^*_{F_2}$. ■

Next we show that the information-theoretic Algorithm 1 is optimal up to constants for MHR distributions. We provide the proof of the following theorem in Appendix 2.8.

Theorem 2.3.7. *Let M be any DSIC and IR mechanism that takes as input a product distribution $\tilde{\mathbf{D}} = \tilde{\mathcal{D}}_1 \times \cdots \times \tilde{\mathcal{D}}_n$. Then there exists a product distribution $\mathbf{D}^* = \mathcal{D}^*_1 \times \cdots \times \mathcal{D}^*_n$ such that $d_k(\mathcal{D}^*_i, \tilde{\mathcal{D}}_i) \leqslant \alpha$, $\mathcal{D}^*_i$ is MHR for every i, and*

$$\mathrm{Rev}(M, \mathbf{D}^*) \leqslant (1 - \tilde{\Omega}(n \cdot \alpha)) \cdot \mathrm{OPT}(\mathbf{D}^*).$$

Analysis for regular distributions

In this section we state the results for the performance of Algorithm 1 for regular distributions and we provide a proof sketch for the single-bidder case. The full proof of the following theorem can be found in Appendix 2.7.

Theorem 2.3.8. *Let $\mathbf{D}^* = \mathcal{D}^*_1 \times \cdots \times \mathcal{D}^*_n$ be a product distribution where every $\mathcal{D}^*_i$ is regular. Let also $\tilde{\mathbf{D}} = \tilde{\mathcal{D}}_1 \times \cdots \tilde{\mathcal{D}}_n$ be any product distribution such that for all $i \in [n]$ it holds that $d_k(\mathcal{D}^*_i, \tilde{\mathcal{D}}_i) \leqslant \alpha_i$. If $\tilde{M}$ is the mechanism that Algorithm 1 outputs with input $\tilde{\mathbf{D}}$ then it holds that*

$$\mathrm{Rev}(\tilde{M}, \mathbf{D}^*) \geqslant \left(1 - 5 \cdot \sqrt{\sum_{i=1}^{n} \alpha_i}\right) \cdot \mathrm{OPT}(\mathbf{D}^*).$$

Proof sketch for $n = 1$. We first prove a general result that for two regular distributions F and $\bar{F}$, such that $d_k(F, \bar{F}) \leqslant \alpha$, where $F(x)$ is stochastically dominated by $\bar{F}(x)$ for $x \in \mathbb{R}_+$. The optimal revenue of these two distributions is close, formally $\frac{\mathrm{OPT}(F)}{\mathrm{OPT}(\bar{F})} \geqslant 1 - O(\sqrt{\alpha})$. The first key step replies on using the link function $h_r(x) = \frac{1}{1-F(x)}$ for regular distributions. Since $h_r(x)$ preserves the same monotonicity property as $F(x)$, we first derive a lower bound on $\bar{h}_r(x, \bar{F})$ that is $\bar{h}_r(x, \bar{F}) \geqslant h_r(x, F) - \alpha h^2_r(x, F)$, using the fact that $d_k(F, \bar{F}) \leqslant \alpha$. This bound gives us useful constraints to discuss in different cases in the following part of the proof. Denote the corresponding optimal reserve prices for F and $\bar{F}$ as P and $\bar{P}$. We discuss separately two cases for $h(\bar{P})$, where, for case 1 we have $h(\bar{P}) \leqslant \frac{1}{\sqrt{\alpha}}$, and for case 2, we have $h(\bar{P}) > \frac{1}{\sqrt{\alpha}}$. Using the connection from the link function to the revenue (see Figure 2.1), case 1 directly leads to the conclusion that $\frac{\mathrm{OPT}(F)}{\mathrm{OPT}(\bar{F})} \geqslant 1 - \sqrt{\alpha}$. Case 2 is more subtle and requires a more careful argument. Lastly, by construction, Algorithm 1 runs the Myerson optimal auction on a regular distribution $\hat{F}$, such that $\hat{F} \geqslant \hat{F}'(x)$ for all $x \in \mathbb{R}_+$, for any other regular distribution $F'(x)$ such that $d_k(F'(x), \tilde{F}(x)) \leqslant \alpha$. Applying the triangle inequality and combining with the conclusions obtained from the two cases concludes the proof. ■

Finally, we show that the information-theoretic Algorithm 1 is optimal up to constants for regular distributions. We provide the proof of the following theorem in Appendix 2.8.

Theorem 2.3.9. *Let M be any DSIC and IR mechanism that takes as input a product distribution $\tilde{\mathbf{D}} = \tilde{\mathcal{D}}_1 \times \cdots \times \tilde{\mathcal{D}}_n$. Then there exists a product distribution $\mathbf{D}^* = \mathcal{D}_1^* \times \cdots \times \mathcal{D}_n^*$ such that $d_k(\mathcal{D}_i^*, \tilde{\mathcal{D}}_i) \leqslant \alpha$, $\mathcal{D}_i^*$ is regular for every i, and*

$$\mathrm{Rev}(M, \mathbf{D}^*) \leqslant (1 - \Omega(\sqrt{n \cdot \alpha})) \cdot \mathrm{OPT}(\mathbf{D}^*).$$

2.4 Finite samples

We provide a practical algorithm that takes samples from the corrupted distribution $\tilde{\mathbf{D}}$ as an input. We show that this algorithm achieves almost optimal sample complexity for the MHR distribution case and the single-bidder regular distribution case, whereas for the multi-bidder regular distributions there is a small gap between our upper and lower bounds.

An important notion to explain our algorithm for the finite-sample case is the following notion of the convex envelope.

Definition 2.4.1 (Convex Envelope)**.** The convex envelope $Conv(f)$ of a function f is a function with the following property

$$Conv(f)(x) = \sup\{g(x) \mid g \text{ is convex and } g \leqslant f \text{ over } \mathbb{R}_+\}.$$

In words, $Conv(f)$ is the maximum convex function that is below f.

For our algorithm one important property of the convex envelope is expressed in the following lemma whose proof is presented in Appendix 2.6.

Lemma 2.4.2. *Let f be a non-decreasing piecewise constant function with k pieces, then $Conv(f)$ can be computed in time* $\mathrm{poly}(k)$ *and is a piecewise linear function with $O(k)$ pieces.*

Algorithm 2 Robust Empirical Myerson Auction

1: **Input:** m i.i.d. samples from (possibly corrupted) value distribution $\mathbf{D} = \Pi_{i=1}^{n} \mathcal{D}_i$, link function $h(\cdot)$.

2: Let $\mathbf{E} = \Pi_{i=1}^{n} E_i$ be the empirical distribution, i.e., the uniform distribution over the samples.

3: **for** $i = 1 \dots n$ **do**

4: Construct $\hat{E}_i$ as following: let $q^{E_i}(v)$ be the quantile of E_i; the quantile of $\hat{E}_i$ is as follows:

$$q^{\hat{E}_i}(v) = \begin{cases} \max\left\{0, q^{E_i}(v) - \sqrt{\frac{2q^{E_i}(v)\left(1-q^{E_i}(v)\right)\ln(2mn\delta^{-1})}{m}} - \frac{4\ln(2mn\delta^{-1})}{m} - \alpha_i\right\} & \text{if } v > 0 \\ 1 & \text{if } v = 0 \end{cases}$$

5: Construct $\tilde{E}_i$ such that $h\left(\tilde{E}_i(\cdot)\right)$ is the convex envelope of $h\left(\hat{E}(\cdot)\right)$, i.e.

$$\tilde{E}_i(\cdot) = h^{-1}\left(Conv\left(h(\hat{E}_i(\cdot))\right)\right)$$

6: **end for**

7: Set $\tilde{\mathbf{E}} = \Pi_{i=1}^{n} \tilde{E}_i$

8: Output Myerson's optimal auction $M_{\tilde{\mathbf{E}}}$ w.r.t. $\tilde{\mathbf{E}}$.

The above algorithm resembles the main algorithm of [100] with the addition of step 5. We first show that step 5 is necessary if we wish to obtain any non-trivial result in the robust auction learning setting that we explore in this paper.

Counterexample 1. Imagine we have just one agent, i.e., $n = 1$, with true distribution $\mathcal{D}^*$ equal to an exponential distribution with parameter $\lambda = 1$. Also, to strengthen our counterexample imagine that we have available an infinite number of samples, i.e., $m \to \infty$. Now consider $\tilde{\mathcal{D}}$ to be the corrupted distribution where probability mass α is removed from the mass closer to 0 and it is placed as a point mass at the point c/α for some number c. In this case, running Algorithm 2 without step 5 will result is implementing an auction with reserve price that is very close to c/α. The probability though that the true agent with distribution $\mathcal{D}^*$ will buy this item goes to zero with a rate $\exp(-c/\alpha)$ as $c \to \infty$. Hence, the total revenue will be at most $(c/\alpha) \cdot \exp(-c/\alpha)$ and therefore we can make the total revenue to go to zero as we increase $c \to \infty$. Observe that this counterexample works even though we assumed that the initial distribution $\mathcal{D}^*$ is MHR.

We next provide the analysis of the performance of Algorithm 2 for MHR and regular distributions. The proof of the following result can be found in Appendix 2.9.

Theorem 2.4.3 (Finite samples, Regular distribution)**.** *Let $\mathbf{D}^* = \mathcal{D}_1^* \times \cdots \times \mathcal{D}_n^*$ be a product distribution where every $\mathcal{D}_i^*$ is regular. Let also $\tilde{\mathbf{D}} = \tilde{\mathcal{D}}_1 \times \cdots \tilde{\mathcal{D}}_n$ be any product distribution such that for all $i \in [n]$ it holds that $d_k(\mathcal{D}_i^*, \tilde{\mathcal{D}}_i) \leqslant \alpha_i$. If $\tilde{M}$ is the mechanism that Algorithm 2 outputs with input m samples from $\tilde{\mathbf{D}}$ and assume that $m = \tilde{\Omega}\left(\max_{i\in[n]}\left\{\log(\frac{1}{\delta})/\alpha_i^2\right\}\right)$ then it holds that*

$$\Pr\left(\mathrm{Rev}(\tilde{M}, \mathbf{D}^*) \geqslant \left(1 - O\left(\sqrt{\sum_{i=1}^{n}\alpha_i}\right)\right) \cdot \mathrm{OPT}(\mathbf{D}^*)\right) \geqslant 1 - \delta.$$

Additionally, in the single-bidder case with $n = 1$ and $\alpha = \alpha_1$ the sample requirement becomes $m = \tilde{\Omega}\left(\log(\frac{1}{\delta})/\alpha^{3/2}\right)$.

The corresponding theorem for MHR distributions is the following, whose proof can be found in Appendix 2.9.

Theorem 2.4.4 (Finite samples, MHR distribution)**.** *Let $\mathbf{D}^* = \mathcal{D}_1^* \times \cdots \times \mathcal{D}_n^*$ be a product distribution where every $\mathcal{D}_i^*$ is MHR. Let also $\tilde{\mathbf{D}} = \tilde{\mathcal{D}}_1 \times \cdots \tilde{\mathcal{D}}_n$ be any product distribution such that for all $i \in [n]$ it holds that $d_k(\mathcal{D}_i^*, \tilde{\mathcal{D}}_i) \leqslant \alpha_i$. If $\tilde{M}$ is the mechanism that Algorithm 2 outputs with input m samples from $\tilde{\mathbf{D}}$ and assume that $m = \tilde{\Omega}\left(\max_{i\in[n]}\left\{\log\left(\frac{1}{\delta}\right)/\alpha_i^2\right\}\right)$ then it holds that*

$$\Pr\left(\mathrm{Rev}(\tilde{M}, \mathbf{D}^*) \geqslant \left(1 - \tilde{O}\left(\sum_{i=1}^{n}\alpha_i\right)\right) \cdot \mathrm{OPT}(\mathbf{D}^*)\right) \geqslant 1 - \delta.$$

We make a few remarks about the sample complexity upper bounds in the sequel.

First, in both Theorem 2.4.3 and Theorem 2.4.4, the sample complexity upper bounds depend in a simple way on the sum of all the fractions of corruptions for each bidder; i.e., $\sum_{i=1}^{n}\alpha_i$, indicating the important effect of the *total* amount of corruption. Second, for regular distributions, in Theorem 2.4.3 we obtain a tight sample complexity bound for the single-bidder case, with $m = \tilde{\Omega}\left(\log(\frac{1}{\delta})/\alpha^{3/2}\right)$. For multi-bidder settings, our upper bound contains a small gap, with $m = \tilde{\Omega}\left(\max_{i\in[n]}\left\{\log(\frac{1}{\delta})/\alpha_i^2\right\}\right)$. Whether such a gap can be matched is an interesting open question for future work. Lastly, comparing Theorem 2.4.3 and Theorem 2.4.4, it appears that for the multi-bidder settings the sample complexity bounds are of the same order, but we emphasize the key difference that for regular distributions this sample size is needed to provide a much *weaker* guarantee on the revenue objective, which is a $\left(1 - O\left(\sqrt{\sum_{i=1}^{n}\alpha_i}\right)\right)$ fraction of the optimal revenue, while the guarantee for MHR distributions is a $\left(1 - O\left(\sum_{i=1}^{n}\alpha_i\right)\right)$ fraction of the optimal revenue.

We next provide an information-theoretic lower bound that establishes the tightness of our upper bounds for the single-bidder single-item case with regular and MHR distributions.

Theorem 2.4.5 (Sample complexity lower bounds)**.** *Let M be any DSIC and IR mechanism for a single-item single-buyer setting that takes as input m samples from a distribution $\tilde{\mathcal{D}}$. If*

$$\mathrm{Rev}(M, \mathcal{D}^*) \geqslant (1 - O(\sqrt{\alpha})) \cdot \mathrm{OPT}(\mathcal{D}^*),$$

for all distributions $\mathcal{D}^*$ *such that* $d_k(\mathcal{D}^*, \tilde{\mathcal{D}}) \leqslant \alpha$, *where* $\mathcal{D}^*$ *is regular, then* $m \geqslant \tilde{\Omega}\left(\log(\frac{2}{\delta})/\alpha^{3/2}\right)$. *Additionally, if*

$$\mathrm{Rev}(M, \mathcal{D}^*) \geqslant (1 - O(\alpha)) \cdot \mathrm{OPT}(\mathcal{D}^*),$$

for all distributions $\mathcal{D}^*$ *such that* $d_k(\mathcal{D}^*, \tilde{\mathcal{D}}) \leqslant \alpha$, *where* $\mathcal{D}^*$ *is MHR, we have* $m \geqslant \tilde{\Omega}\left(\log(\frac{2}{\delta})/\alpha^{3/2}\right)$.

Theorem 2.4.5 provides a general sample complexity lower bound on learning a near-optimal auction with at least a $(1 - O(\sqrt{n \cdot \alpha}))$ fraction of the optimal revenue under the true valuation distribution. In comparison to our upper bounds (see Theorem 2.4.3 and Theorem 2.4.4), there is a small gap and we leave the nature of this gap as an open question for future work.

2.5 Conclusion and future directions

We have studied the learning of revenue-optimal auctions for multiple bidders, in a setting in which the samples can be corrupted adversarially. We first consider the information-theoretic limit in a population model, assuming exact knowledge of the adversarially perturbed valuation distribution. We develop a theoretical algorithm which obtains a tight upper bound on the revenue for the MHR and regular distributions, obtaining the information-theoretic limit of the robustness guarantee. We then relax the population model and derive sample complexity bounds for learning optimal auctions from samples. We propose a practical algorithm which takes the corrupted samples as input, and provide the sample complexity upper bounds for the MHR distribution case and the single-bidder regular distribution case. We also provide accompanying sample complexity lower bounds, and demonstrate a small gap relative to the corresponding upper bounds.

2.6 Appendix: Proofs of technical lemmas

Lemma 2.3.4. *A distribution with CDF* F *is MHR if and only if* $h_M(x; F)$ *is a convex function of* x. *Similarly,* F *is regular if and only if* $h_r(x; F)$ *is a convex function of* x. *Moreover, for two MHR (resp. regular) distributions* F_1 *and* F_2, *such that* $F_1 \succeq F_2$, *we have that* $h_M(x; F_1) \leqslant h_M(x; F_2)$ *(resp.* $h_r(x; F_1) \leqslant h_r(x; F_2)$*) for all* x.

Proof. We first show that given the CDF of any MHR distribution $F(x) : \mathbb{R}_+ \to [0, 1]$, $h_M(x) \stackrel{\text{def}}{=} -\log(1 - F(x))$ is a convex, non-decreasing function with $h(0) = 0$. (Without loss of generality, we consider $x \in [0, \infty]$, i.e. $\arg\min_x h(x) = 0$.) We first present the analysis for the case when the distribution is continuous and smooth, and then generalize the same statement to discrete distributions.

MHR continuous distributions:

Denote the corresponding PDF of $F(x)$ as $f(x)$, and $g(x) \stackrel{\text{def}}{=} \frac{f(v)}{1-F(v)}$. By definition, $F(0) = 0$

implies $h_M(0) = 0$. Then, given that $F(x)$ is MHR, we have that $g(x)$ is monotone non-decreasing. By construction,

$$(h_M(x))'' = \left(\frac{f(v)}{1 - F(v)}\right)' = g'(x) \geqslant 0.$$

Therefore, $h_M(x)$ is convex. Moreover, since $F(x)$ is a CDF thus non-decreasing, $h_M(x) = -\log(1 - F(x))$ is also non-decreasing. We show that given any $h_M(x) : \mathbb{R}_+ \to \mathbb{R}_+$, such that $h_M(x)$ is convex, non-decreasing, $h_M(0) = 0$, and $\max_x h_M(x) = \infty$. Then, $F(x) \stackrel{\text{def}}{=} 1 - \exp(-h_M(x))$ is CDF of an MHR distribution.

By construction, $h_M(0) = 0$ implies $F(0) = 0$, and $\max_x h_M(x)$ implies $\max_x F(x) = 1$. Also given that $h_M(x)$ is convex, $g'(x) = \left(\frac{f(v)}{1-F(v)}\right)' = (h_M(x))'' \geqslant 0$, which by definition implies $F(x)$ is MHR.

MHR discrete distributions:

The lemma statement generalizes to the case when the valuation is discrete. We assume that the valuation can take a discrete set of values $\{x_i\}, i = 1, \cdots, n$. Without loss of generality, we will restrict these values to the set $\mathbb{N}_0$ with probability mass function $P(x = i) = p_i; i = 0 \cdots n$. We define the *discrete* hazard rate as:

$$g(x_i) = \frac{P(x = i)}{P(x \geqslant i)}.$$

Then, the valuation distribution is MHR iff the discrete hazard rate is non-decreasing:

$$g(x_{i+1}) \geqslant g(x_i), \tag{2.2}$$

for all $i = 0 \cdots n$.

In this case, our link function will also be discrete. Further, denote $s_i \stackrel{\text{def}}{=} P(x \geqslant i)$, then

$$h(x_i) = -\log(P(x \geqslant x_i)) = -\log(s_i).$$

Then $h(x)$ is convex if and only if for any $i \geqslant 0$,

$$h(x_{i+2}) - h(x_{i+1}) \geqslant h(x_{i+1} - h(x_i). \tag{2.3}$$

We show that Eq (2.2) and Eq (2.3) are equivalent. Notice that

$$\begin{aligned}
& h(x_{i+2}) - h(x_{i+1}) \geqslant h(x_{i+1} - h(x_i) \\
\iff & \frac{s_{i+1}}{s_{i+1} - p_{i+1}} \geqslant \frac{s_i}{s_i - p_i} \\
\iff & p_{i+1}s_i \geqslant p_i s_{i+1} \\
\iff & \frac{p_{i+1}}{s_{i+1}} \geqslant \frac{p_i}{s_i} \\
\iff & g(x_{i+1}) \geqslant g(x_i),
\end{aligned}$$

which completes the proof.

Regular continuous distributions:

We further prove a similar statement for regular continuous distributions. First, given a CDF of a regular distribution $F(x)$,

$$\left(\frac{1}{1-F(x)}\right)'' = \frac{(1-F(x))f(x)' + 2f(x)^2}{(1-F(x))^3}.$$

By definition, the virtual value function is $\phi(x) \stackrel{\text{def}}{=} v - \frac{1-F(x)}{f(x)}$, and

$$\phi'(x) = \frac{(1-F(x))f(x)' + 2f(x)^2}{f(x)^2}.$$

Therefore, $\left(\frac{1}{1-F(x)}\right)''$ and $\phi'(x)$ share the same sign. Moreover, the distribution with CDF as $F(x)$ is regular if and only if the virtual value $\phi(x)$ is monotonically non-decreasing, which is $\phi'(x) \geqslant 0$. Hence the regularity of $F(x)$ implies that $h_r(x) \stackrel{\text{def}}{=} \frac{1}{1-F(x)}$ is convex. Since $F(x)$ is a CDF thus non-decreasing, $h_r(x) = \frac{1}{1-F(x)}$ is also non-decreasing.

Regular discrete distributions:

Similar to the MHR distributions, the lemma statement generalizes to the case when the valuation is discrete for regular distributions. Assume that the valuation can take a discrete set of values $\{x_i\}, i = 1, \cdots, n$. Without loss of generality, we will restrict these values to the set $\mathbb{N}_0$ with probability mass function $P(x = i) = p_i; i = 0 \cdots n$. Further, consistent with the proof for MHR distributions, we denote $s_i \stackrel{\text{def}}{=} P(x \geqslant i)$.

The *discrete* virtual value function is defined as:

$$\phi(x_i) = x_i - \frac{s_i}{p_i},$$

and the valuation distribution is regular iff $\phi(x)$ is non-decreasing:

$$\phi(x_{i+1}) \geqslant \phi(x_i), \tag{2.4}$$

for all $i = 0 \cdots n$.

In this case, our link function will again be discrete:

$$h(x_i) = \frac{1}{P(x \geqslant x_i)} = \frac{1}{s_i}.$$

and $h(x)$ is convex if and only if for any $i \geqslant 0$,

$$h(x_{i+2}) - h(x_{i+1}) \geqslant h(x_{i+1}) - h(x_i). \tag{2.5}$$

We show that Eq (2.4) and Eq (2.5) are equivalent.

$$\begin{aligned}
&h(x_{i+2}) - h(x_{i+1}) \geqslant h(x_{i+1}) - h(x_i) \\
&\iff \frac{1}{s_{i+2}} + \frac{1}{s_i} \geqslant \frac{2}{s_{i+1}} \\
&\iff \frac{1}{s_{i+1} - p_{i+1}} + \frac{1}{s_i} \geqslant \frac{2}{s_{i+1}} \\
&\iff s_{i+1}^2 + p_i p_{i+1} \geqslant s_i s_{i+1} - s_i p_{i+1}. \\
&\iff p_i p_{i+1} + p_{i+1} s_i + s_{i+1}(s_{i+1} - s_i) \geqslant 0 \\
&\iff p_i p_{i+1} + p_{i+1} s_i - s_{i+1} p_i \geqslant 0
\end{aligned} \tag{2.6}$$

Moreover, from the regularity condition Eq (2.4), we have

$$\begin{aligned}
&\phi(x_{i+1}) \geqslant \phi(x_i) \\
&\iff i + 1 - \frac{s_{i+1}}{p_{i+1}} \geqslant i - \frac{s_i}{p_i} \\
&\iff 1 - \frac{s_{i+1}}{p_{i+1}} + \frac{s_i}{p_i} \geqslant 0 \\
&\iff p_i p_{i+1} + p_{i+1} s_i - s_{i+1} p_i \geqslant 0.
\end{aligned} \tag{2.7}$$

Combining (2.6) and (2.7) together completes the proof.
Stochastic dominance:
Lastly, we show that for two MHR (resp. regular) distributions F_1 and F_2, such that $F_1 \succeq F_2$, then we have that $h_M(x; F_1) \leqslant h_M(x; F_2)$ (resp. $h_r(x; F_1) \leqslant h_r(x; F_2)$) for all x. This follows directly from the monotonicity of the link functions and the definition of stochastic dominance (see Definition 2.3.2).

Recall that the link function $h_M(x; F)$ for MHR distributions is defined as $h_M(x; F) = -\ln(1-F(x))$, and the link function $h_r(x; F)$ for regular distributions is defined as $h_r(x; F) = 1/(1-F(x))$. Therefore, for two MHR (resp. regular) distributions F_1 and F_2, $F_1(x) < F_2(x)$ implies $h_M(x, F_1) < h_M(x, F_2)$ (resp. $h_r(x, F_1) < h_r(x, F_2)$), which completes the proof.

■

Lemma 2.4.2. *Let f be a non-decreasing piecewise constant function with k pieces, then $Conv(f)$ can be computed in time* poly(k) *and is a piecewise linear function with $O(k)$ pieces.*

Proof. Given that $f(x)$ is a non-decreasing piecewise constant function with k pieces, we show that the following iterative procedure outputs its lower convex envelope $Conv(f)$, which can be computed in time poly(k) and is a piecewise linear function with $O(k)$ pieces. Figure 2.3 provides an illustration of the construction according to this procedure.

Procedure 1 Computing lower convex envelope for non-decreasing piecewise constant functions

1: **Input: a piecewise constant function** $f(x) : \mathbb{R} \to \mathbb{R}$ **with** k **pieces.** Denote the left starting point of each piece and the end point as $x_0, \ldots, x_k$.
2: **Initialize:** $i \leftarrow 0, i' \leftarrow 0$.
3: **while** $i \leqslant k-1$ **do**
4: $\quad \bar{x}_{i'} \leftarrow x_i, g(\bar{x}_{i'}) \leftarrow f(x_i)$.
5: $\quad i' \leftarrow i' + 1$.
6: $\quad$ Compute $i \leftarrow \arg\min_{i<j\leqslant k} \frac{f(x_j)-f(x_i)}{x_j-x_i}$.
7: **end while**
8: $\bar{x}_{i'} \leftarrow x_i, g(\bar{x}_{i'}) \leftarrow f(x_i)$; $k' \leftarrow i'$.
9: **Return: a piecewise linear function** $g(x) : \mathbb{R} \to \mathbb{R}$ **with** $k' < k$ **pieces.** The left starting points of each piece and the end points are $\bar{x}_0, \ldots, \bar{x}_{i'}$, with the corresponding function values as specified in the procedure.

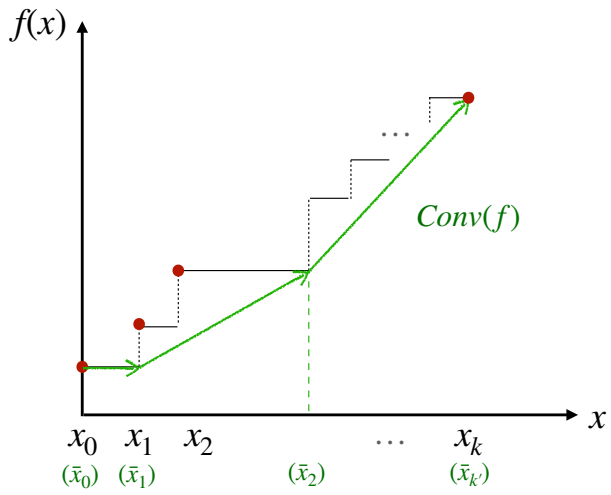


Figure 2.3: Lower convex envelope of a non-decreasing piecewise constant function $f(x)$.

First, the above procedure requires at most k^2 rounds. We show that its output, $g(x)$, is the lower convex envelope for $f(x)$. It is clear from construction that $g(x)$ is piecewise linear, with vertices at $\bar{x}_0, \ldots, \bar{x}_{k'}$. Moreover, $g(x) \leqslant f(x)$ for all x by construction.

Next we show that $g(x)$ is convex. Consider at a round t with $i = i_t, 1 < 1 < k$. Then, step (6) computes $i_{t+1} = \arg\min_{i_t<j\leqslant k} \frac{f(x_j)-f(x_{i_t})}{x_j-x_{i_t}}$. Further denote $\min_{i_t<j\leqslant k} \frac{f(x_j)-f(x_{i_t})}{x_j-x_{i_t}}$ as $s(i_t)$. We show that $s(i_{t+1}) \geqslant s_{i_t}$.

Suppose that $s(i_{t+1}) < s_{i_t}$. Then there exists $j^* > i_{t+1} > i_t$, such that

$$\frac{f(x_{j^*}) - f(x_{i_{t+1}})}{x_{j^*} - x_{i_{t+1}}} < \frac{f(x_{i_{t+1}}) - f(x_{i_t})}{x_{i_{t+1}} - x_{i_t}},$$

which further implies that

$$\frac{f(x_{j^*}) - f(x_{i_t})}{x_{j^*} - x_{i_t}} < \frac{f(x_{i_{t+1}}) - f(x_{i_t})}{x_{i_{t+1}} - x_{i_t}}.$$

Since $j^* > i_{t+1} > i_t$, this contradicts the fact that $i_{t+1} = \arg\min_{i_t < j \leqslant k} \frac{f(x_j) - f(x_{i_t})}{x_j - x_{i_t}}$. Therefore $s(i_{t+1}) \geqslant s_{i_t}$, which means that the slope of each piece for $g(x)$ is non-decreasing. Thus $g(x)$ is convex. Lastly, since $g(x)$ has all vertices with the same function values as $f(x)$, i.e. $g(x) = f(x)$ at all its vertices, and given that $g(x) \leqslant f(x)$ for all x, the values at these vertices are maximized and cannot be further improved. This completes the proof. ■

We further provide two lemmas which present useful properties of the link functions in connection to the revenue.

Lemma 2.6.1. *Given an MHR distribution with the CDF as $F(x) : \mathbb{R}_+ \to [0, 1]$. Define $h(x) \stackrel{\text{def}}{=} -\log(1 - F(x))$. Then, at any reserve price x, the expected revenue $R(x) = \exp(-h(x) + \log(x))$. Moreover, the optimal reserve price P_F^* is the minimizer of $(h(x) - \log(x))$.*

Proof. First by construction, $h(x) - \log(x) = -\log(R(x))$. By definition, the optimal reserve price maximizes the revenue $R(x) = x(1 - F(x))$, thus

$$\begin{aligned} &\max \quad x(1 - F(x)) \\ \iff &\min \quad -\log(x(1 - F(x))) \\ \iff &\min \quad -\log(x) - \log(1 - F(x)) \\ \iff &\min \quad h(x) - \log(x), \end{aligned}$$

which completes the proof. ■

Lemma 2.6.2. *Consider a valuation distribution $\mathcal{D}$ with CDF as $F(x)$. Denote the optimal reserve price as P_F^* and the optimal expected revenue at P_F^* as OPT_F. Then P_F^* should be $P_F^* \leqslant e$, assuming that $\mathrm{OPT}_F \leqslant 1$ and $F(x)$ is MHR.*

Proof. By Lemma 2.6.1, $\mathrm{OPT}_F \leqslant 1$ implies that,

$$h(P_F^*) = \log(P_F^*) + b,$$

for some $b \geqslant 0$. Also by Lemma 2.3.4, h is convex. Combined with the fact that OPT_F is the optimal reserve price and the concavity of $\log(x)$, OPT_F is the only point where $h(P_F^*) = \log(P_F^*) + b$ holds.

Now consider a linear function $y = ax, a > 0$, which is a tangent line of the function $\log(x)+b$. Denote the tangent point as x^*. Solving the equation that $a = (\log(x))' = \frac{1}{x}$, and $ax = \log(x) + b$ give that:

$$x^* = e^{1-b} \leqslant e.$$

Suppose that $P_F^* > x^*$. Consider the linear function $g(x) = \frac{h(P_F^*)}{P_F^*}x$. Since x^* is the tangent point, there exists a point $\bar{x} < P_F^*$, such that $g(\bar{x}) = \log(\bar{x}) + b$. Further, since h is convex, for any point $0 < x < P_F^*$, we have $h(x) < g(x)$. By the continuity of $\log(x)$ and $h(x)$, there exists $\bar{x}' < P_F^*$, such that $h(\bar{x}') = \log(\bar{x}) + b$. This implies that $\bar{x}'$ achieves a larger revenue than P_F^*, and contradicts the fact that P_F^* is the optimal reserve price. Hence, $P_F^* < x^* \leqslant e$, which completes the proof. ■

2.7 Appendix: Proof of upper bounds for the population model

We first prove the following technical lemma that connects the coordinate Kolmogorov distance with the difference in expectation of increasing functions.

Definition 2.7.1 (Increasing Functions and Sets)**.** Let $u : \mathbb{R}^n \to \mathbb{R}$, we say that u is increasing if for every $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{v}' = (v_1', \ldots, v_n')$ such that $v_i' \geqslant v_i$, it holds that $u(\mathbf{v}') \geqslant u(\mathbf{v})$. We say that the subset $A \subseteq \mathbb{R}^n$ is increasing if and only if its characteristic function $\mathbf{1}_A(\mathbf{x})$ is an increasing function of $\mathbf{x}$.

Lemma 2.7.2. *Let $\mathbf{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$, $\mathbf{D}' = \mathcal{D}_1' \times \cdots \times \mathcal{D}_n'$ be product n-dimensional distributions with $d_k(\mathcal{D}_i, \mathcal{D}_i') \leqslant \alpha_i$. Then for every increasing function $u : \mathbb{R}^n \to [0, \bar{u}]$ it holds that*

$$\left| \mathop{\mathbb{E}}_{\mathbf{v}\sim\mathbf{D}}[u(\mathbf{v})] - \mathop{\mathbb{E}}_{\mathbf{v}'\sim\mathbf{D}'}[u(\mathbf{v}')] \right| \leqslant \bar{u} \cdot \left(\sum_{i=1}^{n} \alpha_i \right).$$

Proof. Our first step is to prove that the lemma holds for any function u that is a characteristic function of an increasing set A and then we extend to all increasing functions.

Let $u = \mathbf{1}_A$ we have that $\mathbb{E}_{\mathbf{v}\sim\mathbf{D}}[u(\mathbf{v})] = \Pr_{\mathbf{v}\sim\mathbf{D}}(\mathbf{v} \in A)$. We define the sequence of distributions $\mathbf{D}_j = \mathcal{D}_1' \times \cdots \times \mathcal{D}_j' \times \mathcal{D}_{j+1} \times \cdots \times \mathcal{D}_n$ for $j = 0, \ldots, n$, where obviously $\mathbf{D}_0 = \mathbf{D}$ and $\mathbf{D}_n = \mathbf{D}'$. Now via triangle inequality we have that

$$\left| \Pr_{\mathbf{v}\sim\mathbf{D}}(\mathbf{v} \in A) - \Pr_{\mathbf{v}\sim\mathbf{D}'}(\mathbf{v} \in A) \right| \leqslant \sum_{j=1}^{n} \left| \Pr_{\mathbf{v}\sim\mathbf{D}_j}(\mathbf{v} \in A) - \Pr_{\mathbf{v}\sim\mathbf{D}_{j-1}}(\mathbf{v} \in A) \right|. \tag{2.8}$$

Let $b_j(\mathbf{v}_{-j})$ be the threshold of the step function $\mathbf{1}_A(v_j, \mathbf{v}_{-j})$ when we fix $\mathbf{v}_{-j}$ and we view it as a function of v_j. Now we have that

$$\begin{aligned}\Pr_{\mathbf{v}\sim\mathbf{D}_j}(\mathbf{v}\in A) &= \int_{\mathbb{R}^n} \mathbf{1}_A(x_j,\mathbf{x}_{-j})\ \ d\mathcal{D}'_1(x_1)\cdots d\mathcal{D}'_j(x_j)\cdot d\mathcal{D}_{j+1}(x_{j+1})\cdots d\mathcal{D}_n(x_n)\\ &= \int_{\mathbb{R}^{n-1}} (1-\mathcal{D}'_j(b_j(\mathbf{x}_{-j})))\ \ d\mathcal{D}'_1(x_1)\cdots d\mathcal{D}'_{j-1}(x_{j-1})\cdot d\mathcal{D}_{j+1}(x_{j+1})\cdots d\mathcal{D}_n(x_n)\end{aligned}$$

similarly we have

$$\Pr_{\mathbf{v}\sim\mathbf{D}_{j-1}}(\mathbf{v}\in A) = \int_{\mathbb{R}^{n-1}} (1-\mathcal{D}_j(b_j(\mathbf{x}_{-j})))\ \ d\mathcal{D}'_1(x_1)\cdots d\mathcal{D}'_{j-1}(x_{j-1})\cdot d\mathcal{D}_{j+1}(x_{j+1})\cdots d\mathcal{D}_n(x_n).$$

Combining these we get that

$$\begin{aligned}&\left|\Pr_{\mathbf{v}\sim\mathbf{D}_j}(\mathbf{v}\in A) - \Pr_{\mathbf{v}\sim\mathbf{D}_{j-1}}(\mathbf{v}\in A)\right|\\ &\quad\leqslant \int_{\mathbb{R}^{n-1}} \left|\mathcal{D}'_j(b_j(\mathbf{x}_{-j})) - \mathcal{D}_j(b_j(\mathbf{x}_{-j}))\right|\ \ d\mathcal{D}'_1(x_1)\cdots d\mathcal{D}'_{j-1}(x_{j-1})\cdot d\mathcal{D}_{j+1}(x_{j+1})\cdots d\mathcal{D}_n(x_n).\end{aligned}$$

from the latter we can use the fact that $d_k(\mathcal{D}_j, \mathcal{D}'_j) \leqslant \alpha_j$ and we get that

$$\left|\Pr_{\mathbf{v}\sim\mathbf{D}_j}(\mathbf{v}\in A) - \Pr_{\mathbf{v}\sim\mathbf{D}_{j-1}}(\mathbf{v}\in A)\right| \leqslant \alpha_j.$$

Applying the above to (2.8) we get that

$$\left|\Pr_{\mathbf{v}\sim\mathbf{D}}(\mathbf{v}\in A) - \Pr_{\mathbf{v}\sim\mathbf{D}'}(\mathbf{v}\in A)\right| \leqslant \sum_{j=1}^{n}\alpha_j. \tag{2.9}$$

The last steps is to extend the above to arbitrary increasing functions. We are going to approximate the increasing function u via a sequence of functions u_k which uniformly converges to u. Then we will show the statement of the lemma for every function u_k which by uniform convergence implies the lemma for u as well. We set $A_{i,k} \triangleq \{\mathbf{x}\in\mathbb{R}^n \mid u(\mathbf{x}) \geqslant \frac{i}{k}\bar{u}\}$ and we define

$$u_k(\mathbf{x}) = \frac{\bar{u}}{k}\sum_{i=1}^{k}\mathbf{1}_{A_{i,k}}(\mathbf{x}).$$

Observe from the above definition that $u_k \to u$ uniformly and since u is increasing we also have that all the sets A_i are increasing. Also observe that

$$\mathop{\mathbb{E}}_{\mathbf{v}\sim\mathbf{D}}[u_k(\mathbf{v})] = \frac{\bar{u}}{k}\sum_{i=1}^{k}\Pr_{\mathbf{v}\sim\mathbf{D}}(\mathbf{v}\in A_{i,k})$$

therefore we get that

$$\left| \mathop{\mathbb{E}}_{\mathbf{v}\sim\mathbf{D}}[u_k(\mathbf{v})] - \mathop{\mathbb{E}}_{\mathbf{v}\sim\mathbf{D}'}[u_k(\mathbf{v})] \right| \leqslant \frac{\bar{u}}{k} \sum_{i=1}^{k} \left| \Pr_{\mathbf{v}\sim\mathbf{D}}(\mathbf{v} \in A_{i,k}) - \Pr_{\mathbf{v}\sim\mathbf{D}'}(\mathbf{v} \in A_{i,k}) \right|.$$

Now we can apply (2.9) and we get

$$\left| \mathop{\mathbb{E}}_{\mathbf{v}\sim\mathbf{D}}[u_k(\mathbf{v})] - \mathop{\mathbb{E}}_{\mathbf{v}\sim\mathbf{D}'}[u_k(\mathbf{v})] \right| \leqslant \bar{u} \cdot \left(\sum_{j=1}^{n} \alpha_j \right).$$

Finally, since this is true for every u_k and u converges uniformly to u the above should be true for u as well and hence the lemma follows. ■

We are going to use Lemma 2.7.2 both for the regular distributions case and for the MHR distributions case.

Monotone Hazard Rate Distributions—Proof of Theorem 2.3.6

In this section we show the part of the Theorem 2.3.6 related to $n > 1$. For the stronger result for the case $n = 1$ we refer to Section 2.7.

Let $\tilde{\mathbf{D}}$ be the corrupted product distribution that we observe, $\hat{\mathbf{D}}$ be the output distribution of Algorithm 1, $\mathbf{D}^*$ be the original distribution that we are interested in. We know from the description of Algorithm 1 for $\hat{\mathbf{D}} = \hat{\mathcal{D}}_1 \times \cdots \times \hat{\mathcal{D}}_n$ that $\hat{\mathcal{D}}_i$ is MHR, that $d_k(\hat{\mathcal{D}}_i, \mathcal{D}_i^*) \leqslant \alpha_i$ and that $\hat{\mathcal{D}}_i \preceq \mathcal{D}_i^*$. We also know that $\mathcal{D}_i^*$ is MHR. Finally, we know that the output M of Algorithm 1 is the Myerson optimal mechanism for the distribution $\hat{\mathbf{D}}$ and hence $\mathrm{Rev}(M, \hat{\mathbf{D}}) = \mathrm{OPT}(\hat{\mathbf{D}})$. So applying the strong revenue monotonicity lemma 2.3.3 we have that

$$\mathrm{OPT}(\hat{\mathbf{D}}) = \mathrm{Rev}(M, \hat{\mathbf{D}}) \leqslant \mathrm{Rev}(M, \mathbf{D}^*). \tag{2.10}$$

Therefore to show Theorem 2.3.6, it suffices to show that

$$\mathrm{OPT}(\hat{\mathbf{D}}) \geqslant \left(1 - \tilde{O}\left(\sum_{i=1}^{n} \alpha_i \right) \right) \cdot \mathrm{OPT}(\mathbf{D}^*). \tag{2.11}$$

We are going to use the following result from [43] but with the formulation obtained in Lemma 17 of [100], combined with the weak revenue monotonicity (Lemma 3 of [100]).

Theorem 2.7.3 ([43])**.** *For any product MHR distribution* $\mathbf{D}$*, and any* $\frac{1}{4} \geqslant \varepsilon \geqslant 0$ *and* $u \geqslant c \cdot \log\left(\frac{1}{\varepsilon}\right) \mathrm{OPT}(\mathbf{D})$*. Let* $t_u(\mathcal{D}_1), \ldots, t_u(\mathcal{D}_n)$ *be the distributions obtained by truncating* $\mathcal{D}_1, \ldots, \mathcal{D}_n$ *at the value* $\bar{u}$ *and let* $t_u(\mathbf{D})$ *be their product distribution, where* c *is an absolute constant. Then, we have that*

$$\mathrm{OPT}(\mathbf{D}) \geqslant \mathrm{OPT}(t_u(\mathbf{D})) \geqslant (1 - \varepsilon) \cdot \mathrm{OPT}(\mathbf{D}).$$

Now let $\bar{u} = c \cdot \log\left(\frac{1}{\varepsilon}\right) \mathrm{OPT}(\mathbf{D}^*)$, then we also have that $\bar{u} \geqslant c \cdot \log\left(\frac{1}{\varepsilon}\right) \mathrm{OPT}(\hat{\mathbf{D}})$ due to weak revenue monotonicity (Lemma 3 of [100]). Hence, applying Theorem 2.7.3 we have that

$$\mathrm{OPT}(\hat{\mathbf{D}}) \geqslant \mathrm{OPT}(t_{\bar{u}}(\hat{\mathbf{D}})) \qquad \text{and} \qquad \mathrm{OPT}(t_{\bar{u}}(\mathbf{D}^*)) \geqslant (1-\varepsilon) \cdot \mathrm{OPT}(\mathbf{D}^*). \tag{2.12}$$

Since we know that $d_k(\hat{\mathcal{D}}_i, \mathcal{D}_i^*) \leqslant \alpha_i$ we also have that $d_k(t_{\bar{u}}(\hat{\mathcal{D}}_i), t_{\bar{u}}(\mathcal{D}_i^*)) \leqslant \alpha_i$. Let now $M_{\bar{u}}^*$ be the optimal mechanism for the distribution $t_{\bar{u}}(\mathbf{D}^*)$. It is easy to see that the ex-post revenue obtained from the mechanism $M_{\bar{u}}^*$ is an increasing function of the observed bids. Hence, we can apply Lemma 2.7.2 to the $[0, \bar{u}]$ bounded distributions $t_{\bar{u}}(\hat{\mathbf{D}})$ and $t_{\bar{u}}(\mathbf{D}^*)$ and we get that

$$\begin{aligned}\mathrm{OPT}(t_{\bar{u}}(\hat{\mathbf{D}})) &\geqslant \mathrm{Rev}(M_{\bar{u}}^*, t_{\bar{u}}(\hat{\mathbf{D}})) \geqslant \mathrm{Rev}(M_{\bar{u}}^*, t_{\bar{u}}(\mathbf{D}^*)) - \bar{u} \cdot \left(\sum_{i=1}^{n} \alpha_i\right) \\ &= \mathrm{OPT}(t_{\bar{u}}(\mathbf{D}^*)) - \bar{u} \cdot \left(\sum_{i=1}^{n} \alpha_i\right). \end{aligned} \tag{2.13}$$

If we combine (2.12) and (2.13) then we have that

$$\mathrm{OPT}(\hat{\mathbf{D}}) \geqslant (1-\varepsilon) \cdot \mathrm{OPT}(\mathbf{D}^*) - \bar{u} \cdot \left(\sum_{i=1}^{n} \alpha_i\right). \tag{2.14}$$

Now we can substitute the value of $\bar{u}$ to the above inequality and we get that

$$\mathrm{OPT}(\tilde{\mathbf{D}}) \geqslant \left(1 - c \cdot \log\left(\frac{1}{\epsilon}\right) \cdot \left(\sum_{i=1}^{n} \alpha_i\right) - \varepsilon\right) \cdot \mathrm{OPT}(\mathbf{D}).$$

Finally, setting $\varepsilon = \sum_{i=1}^{n} \alpha_i$ we get

$$\mathrm{OPT}(\tilde{\mathbf{D}}) \geqslant \left(1 - (c+1) \cdot \left(\sum_{i=1}^{n} \alpha_i\right) \cdot \log\left(\frac{1}{\sum_{i=1}^{n} \alpha_i}\right)\right) \cdot \mathrm{OPT}(\mathbf{D}).$$

Hence, (2.11) follows and as we explained this proves Theorem 2.3.6.

Regular Distributions—Proof of Theorem 2.3.8

Let $\tilde{\mathbf{D}}$ be the corrupted product distribution that we observe, $\hat{\mathbf{D}}$ be the output distribution of Algorithm 1, $\mathbf{D}^*$ be the original distribution that we are interested in. We know from the description of Algorithm 1 for $\hat{\mathbf{D}} = \hat{\mathcal{D}}_1 \times \cdots \times \hat{\mathcal{D}}_n$ that $\hat{\mathcal{D}}_i$ is a regular distribution, that $d_k(\hat{\mathcal{D}}_i, \mathcal{D}_i^*) \leqslant \alpha_i$ and that $\hat{\mathcal{D}}_i \preceq \mathcal{D}_i^*$. We also know that $\mathcal{D}_i^*$ is regular. Finally, we know that the output M of Algorithm 1 is the Myerson optimal mechanism for the distribution $\hat{\mathbf{D}}$ and

hence $\mathrm{Rev}(M, \hat{\mathbf{D}}) = \mathrm{OPT}(\hat{\mathbf{D}})$. So applying the strong revenue monotonicity lemma 2.3.3 we have that

$$\mathrm{OPT}(\hat{\mathbf{D}}) = \mathrm{Rev}(M, \hat{\mathbf{D}}) \leqslant \mathrm{Rev}(M, \mathbf{D}^*). \tag{2.15}$$

Therefore to show Theorem 2.3.8, it suffices to show that

$$\mathrm{OPT}(\hat{\mathbf{D}}) \geqslant \left(1 - \tilde{O}\left(\sum_{i=1}^{n} \alpha_i\right)\right) \cdot \mathrm{OPT}(\mathbf{D}^*). \tag{2.16}$$

We are going to use the following theorem from [65], combined with the weak revenue monotonicity (Lemma 3 of [100]).

Theorem 2.7.4 (Lemma 2 of [65])**.** *Let* $\mathbf{D}$ *be a product of* n *regular distributions and* $\mathrm{OPT}(\mathbf{D})$ *be the optimal revenue of* $\mathbf{D}$*. Suppose* $\frac{1}{4} \geqslant \varepsilon \geqslant 0$ *and* $u \geqslant \frac{1}{\varepsilon}\mathrm{OPT}(\mathbf{D})$*. Let* $t_u(\mathcal{D}_1)$*,* $\ldots$*,* $t_u(\mathcal{D}_n)$ *be the distributions obtained by truncating* $\mathcal{D}_1$*,* $\ldots$*,* $\mathcal{D}_n$ *at the value* u *and let* $t_u(\mathbf{D})$ *be their product distribution. Then, we have that*

$$\mathrm{OPT}(\mathbf{D}) \geqslant \mathrm{OPT}(t_u(\mathbf{D})) \geqslant (1 - 4\varepsilon) \cdot \mathrm{OPT}(\mathbf{D}).$$

Now let $\bar{u} = \frac{1}{\varepsilon}\mathrm{OPT}(\mathbf{D}^*)$, then we also have that $\bar{u} \geqslant \frac{1}{\varepsilon}\mathrm{OPT}(\hat{\mathbf{D}})$ due to weak revenue monotonicity (Lemma 3 of [100]). Hence, applying Theorem 2.7.4 we have that

$$\mathrm{OPT}(\hat{\mathbf{D}}) \geqslant \mathrm{OPT}(t_{\bar{u}}(\hat{\mathbf{D}})) \qquad \text{and} \qquad \mathrm{OPT}(t_{\bar{u}}(\mathbf{D}^*)) \geqslant (1 - \varepsilon) \cdot \mathrm{OPT}(\mathbf{D}^*). \tag{2.17}$$

Since we know that $d_k(\hat{\mathcal{D}}_i, \mathcal{D}_i^*) \leqslant \alpha_i$ we also have that $d_k(t_{\bar{u}}(\hat{\mathcal{D}}_i), t_{\bar{u}}(\mathcal{D}_i^*)) \leqslant \alpha_i$. Let now $M_{\bar{u}}^*$ be the optimal mechanism for the distribution $t_{\bar{u}}(\mathbf{D}^*)$. It is easy to see that the ex-post revenue obtained from the mechanism $M_{\bar{u}}^*$ is an increasing function of the observed bids. Hence, we can apply Lemma 2.7.2 to the $[0, \bar{u}]$ bounded distributions $t_{\bar{u}}(\hat{\mathbf{D}})$ and $t_{\bar{u}}(\mathbf{D}^*)$ and we get that

$$\begin{aligned}\mathrm{OPT}(t_{\bar{u}}(\hat{\mathbf{D}})) \geqslant \mathrm{Rev}(M_{\bar{u}}^*, t_{\bar{u}}(\hat{\mathbf{D}})) &\geqslant \mathrm{Rev}(M_{\bar{u}}^*, t_{\bar{u}}(\mathbf{D}^*)) - \bar{u} \cdot \left(\sum_{i=1}^{n} \alpha_i\right) \\ &= \mathrm{OPT}(t_{\bar{u}}(\mathbf{D}^*)) - \bar{u} \cdot \left(\sum_{i=1}^{n} \alpha_i\right). \end{aligned} \tag{2.18}$$

If we combine (2.17) and (2.18) then we have that

$$\mathrm{OPT}(\hat{\mathbf{D}}) \geqslant (1 - \varepsilon) \cdot \mathrm{OPT}(\mathbf{D}^*) - \bar{u} \cdot \left(\sum_{i=1}^{n} \alpha_i\right). \tag{2.19}$$

Now we can substitute the value of $\bar{u}$ to the above inequality and we get that

$$\mathrm{OPT}(\tilde{\mathbf{D}}) \geqslant \left(1 - \frac{1}{\varepsilon} \cdot \left(\sum_{i=1}^{n} \alpha_i\right) - 4\varepsilon\right) \cdot \mathrm{OPT}(\mathbf{D}).$$

Finally, setting $\varepsilon = \sqrt{\sum_{i=1}^{n} \alpha_i}$ we get

$$\mathrm{OPT}(\tilde{\mathbf{D}}) \geqslant \left(1 - 5 \cdot \sqrt{\sum_{i=1}^{n} \alpha_i}\right) \cdot \mathrm{OPT}(\mathbf{D}).$$

Hence, (2.16) follows and as we explained this proves Theorem 2.3.8.

MHR Distributions – Proof of Theorem 2.3.6, $n = 1$ Case

In this subsection we show the part of the Theorem 2.3.6 related to $n = 1$, for which we obtain a stronger result compared to the case $n > 1$. We first show a useful proposition:

Proposition 2.7.5. *Consider two MHR distributions $\mathcal{D}_1$, $\mathcal{D}_2$ with CDFs as F_1 and F_2, such that $d_k(\mathcal{D}_1, \mathcal{D}_1) \leqslant \alpha$, and $F_1(x) \geqslant F_2(x)$ for all $x \in \mathbb{R}_+$. Denote the optimal expected revenue under $\mathcal{D}_1$ and $\mathcal{D}_2$ as OPT_{F_1} and OPT_{F_2}, and the corresponding optimal reserve prices as $P^*_{F_1}$ and $P^*_{F_2}$. Then,*

$$(1 + \alpha e)^{-1} \leqslant \frac{\mathrm{OPT}_{F_1}}{\mathrm{OPT}_{F_2}} \leqslant 1 + \alpha e.$$

Proof. Consider two MHR distributions $\mathcal{D}_1, \mathcal{D}_2$ with CDFs as F_1 and F_2, such that $d_k(\mathcal{D}_1, \mathcal{D}_1) \leqslant \alpha$, and $F_1(x) \geqslant F_2(x)$ for all $x \in \mathbb{R}_+$. Denote the optimal expected revenue under $\mathcal{D}_1$ and $\mathcal{D}_2$ as OPT_{F_1} and OPT_{F_2}, and the corresponding optimal reserve prices as $P^*_{F_1}$ and $P^*_{F_2}$. Without loss of generality, we consider $\mathrm{OPT}_{F_1} \geqslant \mathrm{OPT}_{F_2}$. Further, since the ratio of the revenues, e.g. $\frac{\mathrm{OPT}_{F_1}}{\mathrm{OPT}_{F_2}}$ is scale invariant, we assume without loss of generality that $\mathrm{OPT}_{F_1} = 1$.

By Lemma 2.6.2, we have $P^*_{F_1} \leqslant e$. By Lemma 2.6.1, $\mathrm{OPT}_{F_1} = 1$ implies that $h_1(P^*_{F_1}) = \log(P^*_{F_1})$. Since $P^*_{F_1} \leqslant e$, we have

$$\begin{aligned}
&h_1(P^*_{F_1}) \leqslant 1 \\
&\iff -\log(1 - F_1(P^*_{F_1})) \leqslant 1 \\
&\iff F_1(P^*_{F_1}) \leqslant 1 - \frac{1}{e} \\
&\iff 1 - F_1(P^*_{F_1}) \geqslant \frac{1}{e}.
\end{aligned}$$

Therefore, since F_1 is non-decreasing, for any $x < P^*_{F_1}$, $1 - F_1(x) \geqslant \frac{1}{e}$. So for any $x < P^*_{F_1}$, we have

$$\begin{aligned}
|h_1(x) - h_2(x)| &= \left|\log\left(\frac{1 - F_2(x)}{1 - F_1(x)}\right)\right| \\
&= \left|\log\left(1 + \frac{F_1(x) - F_2(x)}{1 - F_1(x)}\right)\right| \\
&\leqslant \log(1 + \alpha e) \\
&= O(\alpha),
\end{aligned}$$

where the at the second last step, the inequality follows from the fact that $d_k(\mathcal{D}_1, \mathcal{D}_1) \leqslant \alpha$, and $x < P^*_{F_1}$.

Further, $F_1(x) \geqslant F_2(x)$ for all $x \in \mathbb{R}_+$ implies that $h_1(x) \geqslant h_2(x)$ for all $x \in \mathbb{R}_+$. Therefore, $h_1(P^*_{F_1}) = \log(P^*_{F_1}) \geqslant h_2(P^*_{F_1})$. Therefore, we have $P^*_{F_2} \leqslant P^*_{F_1}$, and

$$|h_1(P^*_{F_2}) - h_2(P^*_{F_2})| \leqslant \log(1 + \alpha e).$$

Now define functions $s_1(x) = h_1(x) - \log(x)$, and $s_2(x) = h_2(x) - \log(x)$. Then by the definition of $P^*_{F_1}$, $P^*_{F_2}$ and Lemma 2.6.1,

$$\begin{aligned}\min_{x \leqslant P^*_{F_1}} s_1(x) = s_1(P^*_{F_1}) &\leqslant s_1(P^*_{F_2}) \\ &\leqslant s_2(P^*_{F_2}) + \log(1 + \alpha e) \\ &= \min_{x \leqslant P^*_{F_2}} s_2(x) + \log(1 + \alpha e).\end{aligned}$$

Therefore, by the definitions of s_1 and s_2,

$$\begin{aligned}&\left|\min_{x \leqslant P^*_{F_1}} s_1(x) - \min_{x \leqslant P^*_{F_2}} s_2(x)\right| \leqslant \log(1 + \alpha e) \\ \iff &|\log(\mathrm{OPT}_{F_2}) - \log(\mathrm{OPT}_{F_1})| \leqslant \log(1 + \alpha e) \\ \iff &-\log(1 + \alpha e) \leqslant \log(\mathrm{OPT}_{F_2}) \leqslant \log(1 + \alpha e) \\ \iff &(1 + \alpha e)^{-1} \leqslant \mathrm{OPT}_{F_2} \leqslant 1 + \alpha e.\end{aligned}$$

The above directly implies:

$$(1 + \alpha e)^{-1} \leqslant \frac{\mathrm{OPT}_{F_1}}{\mathrm{OPT}_{F_2}} \leqslant 1 + \alpha e.$$

which completes the proof. ■

Now we are ready to prove Theorem 2.3.6 for the $n = 1$ case.

Proof. First, by construction, Algorithm 1 runs the Myerson optimal auction on an MHR distribution $\hat{F}$, such that $\hat{F} \geqslant \hat{F}'(x)$ for all $x \in \mathbb{R}_+$, for any MHR distribution $F'(x)$ such that $d_k(F'(x), \tilde{F}(x)) \leqslant \alpha$. Also by assumption, $d_k(F^*(x), \tilde{F}(x)) \leqslant \alpha$. Therefore by triangle inequality, $d_k(F^*(x), \hat{F}(x)) \leqslant d_k(F^*(x), \tilde{F}(x)) + d_k(\tilde{F}(x), \hat{F}(x)) \leqslant 2\alpha$.

Denote $\alpha' = 2\alpha$. By Proposition 2.7.5,

$$(1 + \alpha' e)^{-1} \leqslant \frac{\mathrm{OPT}_{F_1}}{\mathrm{OPT}_{F_2}} \leqslant 1 + \alpha' e.$$

Note that $(1 + \alpha' e)^{-1} = (1 + 2\alpha e)^{-1} = 1 - O(\alpha)$, which completes the proof. ■

2.8 Appendix: Proof of optimality for the upper bounds

For these lower bounds we follow the idea of the lower bounds from [100] adapted to the corrupted case that we consider in this paper. The lower bound constructions of [100] are based on a family of distributions

$$\mathcal{H} = \{\mathbf{D} \mid \mathcal{D}_1 = \mathcal{D}^b, \mathcal{D}_i = \mathcal{D}^h \quad \text{or} \quad \mathcal{D}_i = \mathcal{D}^\ell \text{ for all } 2 \leqslant i \leqslant n\}.$$

Observe that this family is characterized by the triplet of distributions $\mathcal{D}^b$, $\mathcal{D}^h$, and $\mathcal{D}^\ell$ for which we ask for the following conditions.

a) $\mathcal{D}^b$ is a point mass at v_0.

b) The propability of $v \geqslant v_2$ is at most $1/n$ both when $v \sim \mathcal{D}^h$ and when $v \sim \mathcal{D}^\ell$.

c) The probability of $v_1 > v \geqslant v_2$ is at least p both when $v \sim \mathcal{D}^h$ and when $v \sim \mathcal{D}^\ell$.

d) For any value v such that $v_1 > v \geqslant v_2$, we have $\phi^\ell(v) + \Delta \leqslant v_0 \leqslant \phi^h(v) - \Delta$, where ϕ^ℓ is the virtual value function of $\mathcal{D}^\ell$ and correspondingly for ϕ^h.

e) For any value v such that $v < v_2$, we have that $\phi^h(v), \phi^\ell(v) \leqslant v_0$.

f) For any value $v_1 > v \geqslant v_2$ we have that the ratio $\frac{d\mathcal{D}^h}{d\mathcal{D}^\ell}(v)$ is upper and lower bounded by a constant, where $\frac{d\mathcal{D}^h}{d\mathcal{D}^\ell}$ is the Radon–Nikodym derivative between $\mathcal{D}^h$ and $\mathcal{D}^\ell$.

g) $\mathcal{D}^h$ is regular.

h) The point v_1 is either $+\infty$ or is a point mass and an upper bound on the support in both $\mathcal{D}^\ell$ and $\mathcal{D}^h$.

Under these conditions and using the exact same proof as the Lemma 18 from [100] we can show the following.

Lemma 2.8.1. *Let $\mathcal{H}$ be a class of distributions that satisfies the conditions a) - h) and additionally satisfies the following.*

i) We have that $d_k(\mathcal{D}^\ell, \mathcal{D}^h) \leqslant \alpha/n$.

Then any algorithm that is robust to a total corruption α in Kolmogorov distance across all bidders achieves revenue of at most

$$\mathrm{OPT}(\mathbf{D}) - \Omega(n \cdot p \cdot \Delta)$$

for any distribution $\mathbf{D} \in \mathcal{H}$.

MHR Distributions – Proof of Theorem 2.3.7

Let $a = \ln(n) - \ln(1-\beta)$, $b = \ln(n)$, $v_0 = a - 1$, $v_1 = \ln(n) - 2 \cdot \ln(1-\beta)$, $v_2 = a$, $p = \beta \cdot (1-\beta)/n$, $\Delta = 1/2$. Then we define $\mathcal{D}^\ell$ and $\mathcal{D}^h$ according to their CDFs F^ℓ and F^h which are the following:

$$F^\ell(v) = \begin{cases} 1 - \exp(-v) & v < v_1 \\ 0 & v \geqslant v_1 \end{cases},$$

$$F^h(v) = \begin{cases} 1 - \exp\left(-\frac{b}{a} \cdot v\right) & v < v_2 \\ 1 - \exp\left(-\frac{v_1 - b}{v_1 - a} \cdot (v - a) + b\right) & v_2 \leqslant v < v_1 \\ 0 & v \geqslant v_1 \end{cases}.$$

Observe also that for this choice of distributions it holds that

$$\phi^\ell(v) = \begin{cases} v - 1 & v < v_1 \\ v_1 & v \geqslant v_1 \end{cases},$$

$$\phi^h(v) = \begin{cases} v - \frac{a}{b} & v < v_2 \\ v - \frac{v_1 - a}{v_1 - b} & v_2 \leqslant v < v_1 \\ v_1 & v \geqslant v_1 \end{cases}.$$

Now the conditions a) - h) are easy to verify. For the condition i) we observe that the maximum difference between the two CDFs is at $v = v_2$ for which we have that $\left|F^\ell(v_2) - F^h(v_2)\right| \leqslant \beta/n$. Hence, Lemma 2.8.1 implies that the maximum revenue achievable by any robust mechanism is

$$\mathrm{OPT}(\mathbf{D}) - \Omega(n \cdot p \cdot \Delta) = \mathrm{OPT}(\mathbf{D}) - \Omega(\beta).$$

Observe that since the maximum value of any bidder is at most $\ln(n)$ we have that the maximum revenue is

$$\left(1 - \frac{\beta}{\ln(n)}\right) \cdot \mathrm{OPT}(\mathbf{D}).$$

If we write this expression with respect to the amount of corruption per bidder, then we have that the maximum possible revenue is

$$\left(1 - \frac{n \cdot \alpha}{\ln(n)}\right) \cdot \mathrm{OPT}(\mathbf{D}).$$

Finally, we observe that all of $\mathcal{D}^b$, $\mathcal{D}^\ell$, and $\mathcal{D}^h$ are MHR and hence Theorem 2.3.7 follows.

Regular Distributions – Proof of Theorem 2.3.9

For the case of regular distributions we will use the same distributions used by [100] in their proof of their Theorem 2. In particular, let $v_0 = 3/2$, $v_1 = +\infty$, $v_2 = 1 + \frac{1}{\beta}$, $p = \frac{\beta}{n}$, and

$\Delta = 1/2$. We define $\mathcal{D}^\ell$ and $\mathcal{D}^h$ through their CDFs as follows

$$F^\ell(v) = 1 - \frac{1}{n \cdot (v-1)},$$

$$F^h(v) = \begin{cases} 0 & v < 1 + \frac{1}{n} \\ 1 - \frac{1}{n \cdot (v-1)} & 1 + \frac{1}{n} \geqslant v < v_2 \\ 1 - \frac{1-\beta}{n \cdot (v-2)} & v \geqslant v_2 \end{cases}.$$

The fact that these distributions satisfy a) - h) can be found in [100]. We will focus on proving i). It is not hard to see that the two CDFs appears when $v = \bar{v} = 1 + \frac{1}{\sqrt{1-\beta}}$. For this value we have

$$\left|F^\ell(\bar{v}) - F^h(\bar{v})\right| = \frac{1}{n}\left(2 - \beta - 2\sqrt{1-\beta}\right) \leqslant \frac{\beta^2}{n},$$

where the last inequality can be easily verifies for $\beta \leqslant 1$. Now setting $\alpha = \frac{\beta^2}{n}$, observing that $n \cdot p \cdot \Delta = \Omega(\beta)$, and observing that $\mathrm{OPT}(\mathbf{D}) \leqslant O(1)$ we can apply Lemma 2.8.1 and we get that the maximum possible revenue is

$$\left(1 - \Omega\left(\sqrt{n \cdot \alpha}\right)\right) \cdot \mathrm{OPT}(\mathbf{D}).$$

Finally by observing that all of $\mathcal{D}^b$, $\mathcal{D}^\ell$, and $\mathcal{D}^h$ are regular Theorem 2.3.9 follows.

2.9 Appendix: Proofs of sample complexity bounds

Proof of Theorem 2.4.3, $n > 1$ case

This follows easily from Theorem 2.3.8 and the DKW inequality Dvoretzky et al. [76], Massart [161] that states that the empirical CDF with m samples is close to the population CDF with an error of at most

$$O\left(\sqrt{\frac{\log(1/\delta)}{m}}\right)$$

with probability at least $1 - \delta$. ■

Proof of Theorem 2.4.3, $n = 1$ Case

We present in this section a proof of Theorem 2.4.3 for the case with $n = 1$ and regular distributions. In this case, we show that Algorithm 2 achieves the optimal sample complexity, up to a poly-logarithmic factor.

First, by [Lemma 5, Guo et al. [100]], we have that with probability at least $1-\delta$, for any value $v \geqslant 0$, the quantiles of $\tilde{\mathcal{D}}$ and its empirical counterpart E satisfy that:

$$|q^E(v) - q^{\tilde{\mathcal{D}}}(v)| \leqslant \sqrt{\frac{2q^{\tilde{\mathcal{D}}}(v)(1-q^{\tilde{\mathcal{D}}(v)})\ln(2m\delta^{-1})}{m}} + \frac{\ln(2m\delta^{-1})}{m}. \tag{2.20}$$

Further note that by construction, we have

$$q^E - q^{\hat{E}} \leqslant \sqrt{\frac{2q^E(v)\left(1-q^E(v)\right)\ln(2m\delta^{-1})}{m}} + \frac{4\ln(2m\delta^{-1})}{m} + \alpha.$$

Given that Algorithm 2 runs the Myerson optimal auction on $\tilde{E}$, which is a minimal regular distribution that dominates $\tilde{E}$. Further, $\hat{E} \succeq D^*$ by construction, assuming Eq (2.20) holds. Therefore, we have $D^* \succeq \tilde{E}$ assuming Eq (2.20) holds. Applying Lemma 2.3.3 yields:

$$\mathrm{Rev}(M_{\tilde{E}}, \mathcal{D}^*) \geqslant \mathrm{Rev}(M_{\tilde{E}}, \tilde{E}) = \mathrm{OPT}(\tilde{E}).$$

Therefore, the remaining task is to ensure that m is sufficiently large such that

$$\mathrm{OPT}(\tilde{E}) \geqslant (1-\sqrt{\alpha})\mathrm{OPT}(\mathcal{D}^*).$$

We will use a useful lemma below which connects the ratio of revenues that we are interested in with the value of link function at an optimal reserve price.

Lemma 2.9.1. *Given two regular distributions $\mathcal{D}, \bar{\mathcal{D}}$ with CDFs $F, \bar{F}$, such that $\bar{F} \succeq F$ and $d_k(\mathcal{D}, \bar{\mathcal{D}}) \leqslant \beta$. Denote the optimal reserve price for $\bar{F}$ as $\bar{P}$, and the optimal expected revenue for $F, \bar{F}$ as $\mathrm{OPT}_F, \mathrm{OPT}_{\bar{F}}$. Then we have*

$$\frac{\mathrm{OPT}_F}{\mathrm{OPT}_{\bar{F}}} \geqslant 1 - \beta h_r(\bar{P})$$

Proof. Recall that $h_r(x) = \frac{1}{1-F(x)}$, and $\bar{h}_r(x) = \frac{1}{1-\bar{F}(x)}$. Then, $F(x) \geqslant \bar{F}(x)$ implies $h_r(x) \geqslant \bar{h}_r(x)$.

By definition, $d_k(\mathcal{D}, \bar{\mathcal{D}}) \leqslant \beta$ implies that $\max_x F(x) - \bar{F}(x) \leqslant \beta$. So we have:

$$h_r(x) - \bar{h}_r(x) = \frac{F(x)-\bar{F}(x)}{(1-F(x))(1-\bar{F}(x)} = (F(x)-\bar{F}(x))h_r(x)\bar{h}_r(x) \leqslant \beta h_r^2(x),$$

where the last inequality follows from the fact that $\max_x F(x) - \bar{F}(x) \leqslant \beta$, and $h_r(x) \geqslant \bar{h}_r(x)$. Thus, for all x,

$$\bar{h}_R(x) \geqslant h_r(x) - \beta h_r^2(x). \tag{2.21}$$

Note that the expected revenue, $R(x) = x(1-F(x))$, at any x, equals to $\frac{x}{h_r(x)}$, which is the reciprocal of the slope for the linear function $g(a) = h_r(x) \cdot a$. Hence, the revenue is maximized when the slope for the linear function $g(a) = h_r(x) \cdot a$ is minimized.

Denote the corresponding optimal reserve prices for F and $\bar{F}$ as P and $\bar{P}$. Then at $\bar{P}$,

$$\bar{h}_r(\bar{P}) = \frac{1}{1-\bar{F}(\bar{P})} = \frac{1}{\mathrm{OPT}_{\bar{F}}} \cdot \bar{P}.$$

Denote $\mathrm{Rev}(F, x)$ as the expected revenue with a reserve price at x for a valuation distribution with CDF as F. Then,

$$\frac{\mathrm{OPT}_F}{\mathrm{OPT}_{\bar{F}}} \geqslant \frac{\mathrm{Rev}(F,\bar{P})}{\mathrm{OPT}_{\bar{F}}} = \frac{\bar{h}_r(\bar{P})}{h(\bar{P})} \geqslant \frac{h_r(\bar{P}) - \beta h_r^2(\bar{P})}{h_r(\bar{P})} = 1 - \beta h_r(\bar{P}),$$

where the first inequality follows directly from the definition of the optimal revenue, and the second inequality is from Eq (2.21). ■

Now we will use Lemma 2.9.1 to proceed. Denote the optimal reserve price for $\mathcal{D}^*$ as P^*. Denote the link function applied to $\tilde{E}$ and $\mathcal{D}^*$ as $\tilde{h}$, h^*, respectively. Then, we will discuss two cases for $\tilde{h}(P^*)$.

Case 1: $\tilde{h}(P^*) > \frac{1}{\sqrt{\alpha}}$. For this case, $\tilde{h}(P^*) > \frac{1}{\sqrt{\alpha}}$ implies that $q^{\tilde{E}}(P^*) < \sqrt{\alpha}$. Applying [Lemma 5, Guo et al. [100]] and triangle inequalities, we have

$$|q^{\tilde{E}} - q^{\mathcal{D}^*}| \leqslant \sqrt{\frac{2q^{\tilde{E}}(v)\left(1-q^{\tilde{E}}(v)\right)\ln(2m\delta^{-1})}{m}} + \frac{4\ln(2m\delta^{-1})}{m} + \alpha.$$

Given that $q^{\tilde{E}}(P^*) < \sqrt{\alpha}$, we have $q^{\tilde{E}}(1-q^{\tilde{E}}) \leqslant q^{\tilde{E}} \leqslant \sqrt{\alpha}$. Therefore, it suffices to have

$$\sqrt{\frac{\sqrt{\alpha}}{m}} \leqslant C_1\alpha,$$

for some universal constant C_1 to ensure that $|q^{\tilde{E}} - q^{\mathcal{D}^*}| = O(\alpha)$, which implies $m \geqslant 1/\{C_1^2\alpha^{3/2}\}$ for some universal constant C_1.

Case 2: $\tilde{h}(P^*) \leqslant \frac{1}{\sqrt{\alpha}}$. For this case, $\tilde{h}(P^*) \leqslant \frac{1}{\sqrt{\alpha}}$ implies that $q^{\tilde{E}}(P^*) \geqslant \sqrt{\alpha}$.

By lemma 2.9.1, we have that

$$\frac{\mathrm{OPT}_{\tilde{E}}}{\mathrm{OPT}_{\mathcal{D}^*}} \geqslant 1 - \beta\tilde{h}_r(P^*),$$

therefore it suffice to ensure that $1 - \beta\tilde{h}_r(P^*) \geqslant 1 - C_2\sqrt{\alpha}$ for some universal constant C_2, which implies that $\beta \leqslant q^{\tilde{E}}(P^*) \cdot C_2\sqrt{\alpha}$. Applying [Lemma 5, Guo et al. [100]], it suffices to have that $\sqrt{\frac{q^{\tilde{E}}(P^*)}{m}} \leqslant \beta \leqslant q^{\tilde{E}}(P^*) \cdot C_2\sqrt{\alpha}$, which yields that $m > \frac{1}{C_2^2\alpha q^{\tilde{E}}}$. Lastly, applying the fact that we are in the case where $q^{\tilde{E}}(P^*) \geqslant \sqrt{\alpha}$ we get that is suffices to have $m > \frac{1}{C_2^2\alpha^{3/2}}$ for some universal constant C_2. This completes the proof.

■

Proof of Theorem 2.4.4

This follows easily from Theorem 2.3.6 and the DKW inequality [76, 161] that states that the empirical CDF with m samples is close to the population CDF with an error of at most

$$O\left(\sqrt{\frac{\log(1/\delta)}{m}}\right)$$

with probability at least $1 - \delta$. ■

Proof of Theorem 2.4.5

We omit the details of this proof since it follows from Theorem 2 and Appendix E of [100] applied for the case $n = 1$. The reason is that if we could get a better bound in our corrupted case then this algorithm could be used to improve our sample complexity result in the non-corrupted case.

Chapter 3

Robust Optimization for Fairness with Noisy Protected Groups

3.1 Introduction

As machine learning becomes increasingly pervasive in real-world decision making, the question of ensuring *fairness* of ML models becomes increasingly important. The definition of what it means to be "fair" is highly context dependent. Much work has been done on developing mathematical fairness criteria according to various societal and ethical notions of fairness, as well as methods for building machine-learning models that satisfy those fairness criteria [see, e.g., 54, 77, 87, 110, 141, 189, 216, 222].

Many of these mathematical fairness criteria are *group-based*, where a target metric is equalized or enforced over subpopulations in the data, also known as *protected groups*. For example, the *equality of opportunity* criterion introduced by Hardt et al. [110] specifies that the true positive rates for a binary classifier are equalized across protected groups. The *demographic parity* [77] criterion requires that a classifier's positive prediction rates are equal for all protected groups.

One important practical question is whether or not these fairness notions can be reliably measured or enforced if the protected group information is noisy, missing, or unreliable. For example, survey participants may be incentivized to obfuscate their responses for fear of disclosure or discrimination, or may be subject to other forms of response bias. Social desirability response bias may affect participants' answers regarding religion, political affiliation, or sexual orientation [138]. The collected data may also be outdated: census data collected ten years ago may not an accurate representation for measuring fairness today.

Another source of noise arises from estimating the labels of the protected groups. For various image recognition tasks (e.g., face detection), one may want to measure fairness across protected groups such as gender or race. However, many large image corpora do not include protected group labels, and one might instead use a separately trained classifier to estimate group labels, which is likely to be noisy [41]. Similarly, zip codes can act as a noisy

indicator for socioeconomic groups.

In this work, we focus on the problem of training binary classifiers with fairness constraints when only noisy labels, $\hat{G} \in \{1, ..., \hat{m}\}$, are available for m true protected groups, $G \in \{1, ..., m\}$, of interest. We study two aspects: First, if one satisfies fairness constraints for noisy protected groups $\hat{G}$, what can one say with respect to those fairness constraints for the true groups G? Second, how can side information about the noise model between $\hat{G}$ and G be leveraged to better enforce fairness with respect to the true groups G?

Contributions: Our contributions are three-fold:

1. We provide a bound on the fairness violations with respect to the true groups G when the fairness criteria are satisfied for the noisy groups $\hat{G}$.

2. We introduce two new robust-optimization methodologies that satisfy fairness criteria on the true protected groups G while minimizing a training objective. These methodologies differ in convergence properties, conservatism, and noise model specification.

3. We show empirically that unlike the naïve approach, our two proposed approaches are able to satisfy fairness criteria with respect to the true groups G on average.

The first approach we propose (Section 3.5) is based on distributionally robust optimization (DRO) [29, 71]. Let p denote the full distribution of the data, $X, Y \sim p$. Let p_j be the distribution of the data conditioned on the true groups being j, so $X, Y|G = j \sim p_j$; and $\hat{p}_j$ be the distribution of X, Y conditioned on the noisy groups, so $X, Y|\hat{G} = j \sim \hat{p}_j$. Given an upper bound on the total variation (TV) distance $\gamma_j \geqslant TV(p_j, \hat{p}_j)$ for each $j \in \{1, ..., m\}$, we define $\tilde{p}_j$ such that the conditional distributions $(X, Y|\tilde{G} = j \sim \tilde{p}_j)$ fall within the bound γ_j with respect to $\hat{p}_j$: $\gamma_j \geqslant TV(\tilde{p}_j, \hat{p}_j)$. Thus, the set of all such $\tilde{p}_j$ is guaranteed to include the unknown true group distribution p_j, for all j. Because it is based on the well-studied DRO setting, this approach has the advantage of being easy to analyze. However, the results may be overly conservative unless tight bounds $\{\gamma_j\}_{j=1}^m$ can be given.

Our second robust optimization strategy (Section 3.6) uses a robust re-weighting of the data from soft protected group assignments, inspired by criteria proposed by Kallus et al. [128] for auditing the fairness of ML models given imperfect group information. Extending their work, we *optimize* a constrained problem to achieve their robust fairness criteria, and provide a theoretically ideal algorithm that is guaranteed to converge to an optimal feasible point, as well as an alternative practical version that is more computationally tractable. Compared to DRO, this second approach uses a more precise noise model, $P(\hat{G} = k|G = j)$, between $\hat{G}$ and G for all pairs of group labels j, k, that can be estimated from a small auxiliary dataset containing ground-truth labels for both G and $\hat{G}$. An advantage of this more detailed noise model is that a practitioner can incorporate knowledge of any bias in the relationship between G and $\hat{G}$ (for instance, survey respondents favoring one socially preferable response over others), which causes it to be less likely than DRO to result in an overly-conservative model. Notably, this approach does *not* require that $\hat{G}$ be a direct

approximation of G—in fact, G and $\hat{G}$ can represent distinct (but related) groupings, or even groupings of different sizes, with the noise model tying them together. For example, if G represents "language spoken at home," then $\hat{G}$ could be a noisy estimate of "country of residence."

3.2 Related work

Constrained optimization for group-based fairness metrics: The simplest techniques for enforcing group-based constraints apply a post-hoc correction of an existing classifier [110, 220]. For example, one can enforce *equality of opportunity* by choosing different decision thresholds for an existing binary classifier for each protected group [110]. However, the classifiers resulting from these post-processing techniques may not necessarily be optimal in terms of accuracy. Thus, constrained optimization techniques have emerged to train machine-learning models that can more optimally satisfy the fairness constraints while minimizing a training objective [4, 53, 54, 69, 94, 173, 222].

Fairness with noisy protected groups: Group-based fairness notions rely on the knowledge of *protected group* labels. However, practitioners may only have access to noisy or unreliable protected group information. One may naïvely try to enforce fairness constraints with respect to these noisy protected groups using the above constrained optimization techniques, but there is no guarantee that the resulting classifier will satisfy the fairness criteria with respect to the true protected groups [105].

Under the conservative assumption that a practitioner has no information about the protected groups, Hashimoto et al. [112] applied DRO to enforce what Lahoti et al. [142] refer to as *Rawlsian Max-Min fairness*. In contrast, here we assume some knowledge of a noise model for the noisy protected groups, and are thus able to provide tighter results with DRO: we provide a practically meaningful maximum total variation distance bound to enforce in the DRO procedure. We further extend Hashimoto et al. [112]'s work by applying DRO to problems equalizing fairness metrics over groups, which may be desired in some practical applications [137].

Concurrently, Lahoti et al. [142] proposed an adversarial reweighting approach to improve group fairness by assuming that non-protected features and task labels are correlated with unobserved groups. Like Hashimoto et al. [112], Lahoti et al. [142] also enforce *Rawlsian Max-Min fairness* with unknown protected groups, whereas our setup includes constraints for parity based fairness notions.

Kallus et al. [128] considered the problem of *auditing* fairness criteria given noisy groups. They propose a "robust" fairness criteria using soft group assignments and show that if a given model satisfies those fairness criteria with respect to the noisy groups, then the model will satisfy the fairness criteria with respect to the true groups. Here, we build on that work by providing an algorithm for training a model that satisfies their robust fairness criteria while minimizing a training objective.

Lamy et al. [145] showed that when there are only two protected groups, one need only tighten the "unfairness tolerance" when enforcing fairness with respect to the noisy groups. Mozannar et al. [168] showed that if the predictor is independent of the protected attribute, then fairness with respect to the noisy groups is the same as fairness with respect to the true groups. When there are more than two groups, and when the noisy groups are included as an input to the classifier, other robust optimization approaches may be necessary. When using post-processing instead of constrained optimization, Awasthi et al. [15] showed that under certain conditional independence assumptions, post-processing using the noisy groups will not be worse in terms of fairness violations than not post-processing at all. In our work, we consider the problem of training the model subject to fairness constraints, rather than taking a trained model as given and only allowing post-processing, and we do not rely on conditional independence assumptions. Indeed, the model may include the noisy protected attribute as a feature.

Robust optimization: We use a minimax set-up of a two-player game where the uncertainty is adversarial, and one minimizes a worst-case objective over a feasible set [28, 35]; e.g., the noise is contained in a unit-norm ball around the input data. As one such approach, we apply a recent line of work on DRO which assumes that the uncertain distributions of the data are constrained to belong to a certain set [71, 153, 172].

3.3 Optimization problem setup

We begin with the training problem for incorporating group-based fairness criteria in a learning setting [4, 54, 69, 94, 110]. Let $X \in \mathcal{X} \subseteq \mathbb{R}^D$ be a random variable representing a feature vector, with a random binary label $Y \in \mathcal{Y} = \{0, 1\}$ and random protected group membership $G \in \mathcal{G} = \{1, ..., m\}$. In addition, let $\hat{G} \in \hat{\mathcal{G}} = \{1, ..., \hat{m}\}$ be a random variable representing the noisy protected group label for each (X, Y), which we assume we have access to during training. For simplicity, assume that $\hat{\mathcal{G}} = \mathcal{G}$ (and $\hat{m} = m$). Let $\phi(X; \theta)$ represent a binary classifier with parameters $\theta \in \Theta$ where $\phi(X; \theta) > 0$ indicates a positive classification.

Then, training with fairness constraints [4, 54, 69, 94, 110] is:

$$\min_{\theta} \quad f(\theta) \quad \text{s.t.} \quad g_j(\theta) \leqslant 0, \forall j \in \mathcal{G}, \tag{3.1}$$

The objective function $f(\theta) = \mathbb{E}[l(\theta, X, Y)]$, where $l(\theta, X, Y)$ is any standard binary classifier training loss. The constraint functions $g_j(\theta) = \mathbb{E}[h(\theta, X, Y)|G = j]$ for $j \in \mathcal{G}$, where $h(\theta, X, Y)$ is the target fairness metric, e.g. $h(\theta, X, Y) = \mathbb{1}\left(\phi(X; \theta) > 0\right) - \mathbb{E}[\mathbb{1}\left(\phi(X; \theta) > 0\right)]$ when equalizing positive rates for the *demographic parity* [77] criterion (see [54] for more examples). Algorithms have been studied for problem (3.1) when the true protected group labels G are given [see, e.g., 4, 54, 78].

3.4 Bounds for the naïve approach

When only given the noisy groups $\hat{G}$, one naïve approach to solving problem (3.1) is to simply re-define the constraints using the noisy groups [105]:

$$\min_{\theta} \quad f(\theta) \quad \text{s.t.} \quad \hat{g}_j(\theta) \leqslant 0, \ \ \forall j \in \mathcal{G}, \tag{3.2}$$

where $\hat{g}_j(\theta) = \mathbb{E}[h(\theta, X, Y)|\hat{G} = j], \ \ j \in \mathcal{G}$.

This introduces a practical question: if a model was constrained to satisfy fairness criteria on the noisy groups, how far would that model be from satisfying the constraints on the true groups? We show that the fairness violations on the true groups G can at least be bounded when the fairness criteria are satisfied on the noisy groups $\hat{G}$, provided that $\hat{G}$ does not deviate too much from G.

Bounding fairness constraints using TV distance

Recall that $X, Y|G = j \sim p_j$ and $X, Y|\hat{G} = j \sim \hat{p}_j$. We use the TV distance $TV(p_j, \hat{p}_j)$ to measure the distance between the probability distributions p_j and $\hat{p}_j$ (see Appendix 3.10 and Villani [211]). Given a bound on $TV(p_j, \hat{p}_j)$, we obtain a bound on fairness violations for the true groups when naïvely solving the optimization problem (3.2) using only the noisy groups:

Theorem 3.4.1. *(proof in Appendix 3.10.) Suppose a model with parameters θ satisfies fairness criteria with respect to the noisy groups $\hat{G}$: $\hat{g}_j(\theta) \leqslant 0, \ \ \forall j \in \mathcal{G}$. Suppose $|h(\theta, x_1, y_1) - h(\theta, x_2, y_2)| \leqslant 1$ for any $(x_1, y_1) \neq (x_2, y_2)$. If $TV(p_j, \hat{p}_j) \leqslant \gamma_j$ for all $j \in \mathcal{G}$, then the fairness criteria with respect to the true groups G will be satisfied within slacks γ_j for each group: $g_j(\theta) \leqslant \gamma_j, \ \ \forall j \in \mathcal{G}$.*

Theorem 3.4.1 is tight for the family of functions h that satisfy $|h(\theta, x_1, y_1) - h(\theta, x_2, y_2)| \leqslant 1$ for any $(x_1, y_1) \neq (x_2, y_2)$. This condition holds for any fairness metrics based on rates such as demographic parity, where h is simply some scaled combination of indicator functions. Cotter et al. [54] list many such rate-based fairness metrics. Theorem 3.4.1 can be generalized to functions h whose differences are not bounded by 1 by looking beyond the TV distance to more general Wasserstein distances between p_j and $\hat{p}_j$. We show this in Appendix 3.10, but for all fairness metrics referenced in this work, formulating Theorem 3.4.1 with the TV distance is sufficient.

Estimating the TV distance bound in practice

Theorem 3.4.1 bounds the fairness violations of the naïve approach in terms of the TV distance between the conditional distributions p_j and $\hat{p}_j$, which assumes knowledge of p_j and is not always possible to estimate. Instead, we can estimate an upper bound on $TV(p_j, \hat{p}_j)$ from metrics that are easier to obtain in practice. Specifically, the following lemma shows

that shows that if the prior on class j is unaffected by the noise, $P(G \neq \hat{G}|G = j)$ directly translates into an upper bound on $TV(p_j, \hat{p}_j)$.

Lemma 3.4.2. *(proof in Appendix 3.10.) Suppose $P(G = j) = P(\hat{G} = j)$ for a given $j \in \mathcal{G}$. Then $TV(p_j, \hat{p}_j) \leqslant P(G \neq \hat{G}|G = j)$.*

In practice, an estimate of $P(G \neq \hat{G}|G = j)$ may come from a variety of sources. As assumed by Kallus et al. [128], a practitioner may have access to an *auxiliary* dataset containing G and $\hat{G}$, but not X or Y. Or, practitioners may have some prior estimate of $P(G \neq \hat{G}|G = j)$: if $\hat{G}$ is estimated by mapping zip codes to the most common socioeconomic group for that zip code, then census data provides a prior for how often $\hat{G}$ produces an incorrect socioeconomic group.

By relating Theorem 3.4.1 to realistic noise models, Lemma 3.4.2 allows us to bound the fairness violations of the naïve approach using quantities that can be estimated empirically. In the next section we show that Lemma 3.4.2 can also be used to produce a *robust* approach that will actually guarantee full satisfaction of the fairness violations on the true groups G.

3.5 Robust Approach 1: Distributionally robust optimization (DRO)

While Theorem 3.4.1 provides an upper bound on the performance of the naïve approach, it fails to provide a guarantee that the constraints on the true groups are satisfied, i.e. $g_j(\theta) \leqslant 0$. Thus, it is important to find other ways to do better than the naïve optimization problem (3.2) in terms of satisfying the constraints on the true groups. In particular, suppose in practice we are able to assert that $P(G \neq \hat{G}|G = j) \leqslant \gamma_j$ for all groups $j \in \mathcal{G}$. Then Lemma 3.4.2 implies a bound on TV distance between the conditional distributions on the true groups and the noisy groups: $TV(p_j, \hat{p}_j) \leqslant \gamma_j$. Therefore, any feasible solution to the following constrained optimization problem is guaranteed to satisfy the fairness constraints on the true groups:

$$\min_{\theta \in \Theta} \quad f(\theta) \quad \text{s.t.} \quad \max_{\substack{\tilde{p}_j : TV(\tilde{p}_j, \hat{p}_j) \leqslant \gamma_j \\ \tilde{p}_j \ll p}} \tilde{g}_j(\theta) \leqslant 0, \quad \forall j \in \mathcal{G}, \tag{3.3}$$

where $\tilde{g}_j(\theta) = \mathbb{E}_{X,Y \sim \tilde{p}_j}[h(\theta, X, Y)]$, and $\tilde{p}_j \ll p$ denotes absolute continuity.

General DRO formulation

A DRO problem is a minimax optimization [71]:

$$\min_{\theta \in \Theta} \max_{q : D(q,p) \leqslant \gamma} \mathop{\mathbb{E}}_{X,Y \sim q} [l(\theta, X, Y)], \tag{3.4}$$

where D is some divergence metric between the distributions p and q, and $l : \Theta \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Much existing work on DRO focuses on how to solve the DRO problem for different divergence metrics D. Namkoong and Duchi [172] provide methods for efficiently and optimally solving the DRO problem for f-divergences, and other work has provided methods for solving the DRO problem for Wasserstein distances [81, 153]. Duchi and Namkoong [71] further provide finite-sample convergence rates for the empirical version of the DRO problem.

Solving the DRO problem

An important and often difficult aspect of using DRO is specifying a divergence D and bound γ that are meaningful. In this case, Lemma 3.4.2 gives us the key to formulating a DRO problem that is guaranteed to satisfy the fairness criteria with respect to the true groups G.

The optimization problem (3.3) can be written in the form of a DRO problem (3.4) with TV distance by using the Lagrangian formulation. Adapting a simplified version of a gradient-based algorithm provided by Namkoong and Duchi [172], we are able to solve the empirical formulation of problem (3.4) efficiently. Details of our empirical Lagrangian formulation and pseudocode are in Appendix 3.11.

3.6 Robust Approach 2: Soft group assignments

While any feasible solution to the distributionally robust constrained optimization problem (3.3) is guaranteed to satisfy the constraints on the true groups G, choosing each $\gamma_j = P(G \neq \hat{G}|G = j)$ as an upper bound on $TV(p_j, \hat{p}_j)$ may be rather conservative. Therefore, as an alternative to the DRO constraints in (3.3), in this section we show how to optimize using the robust fairness criteria proposed by Kallus et al. [128].

Constraints with soft group assignments

Given a trained binary predictor, $\hat{Y}(\theta) = \mathbb{1}(\phi(\theta; X) > 0)$, Kallus et al. [128] proposed a set of robust fairness criteria that can be used to audit the fairness of the given trained model with respect to the true groups $G \in \mathcal{G}$ using the noisy groups $\hat{G} \in \hat{\mathcal{G}}$, where $\mathcal{G} = \hat{\mathcal{G}}$ is not required in general. They assume access to a *main dataset* with the noisy groups $\hat{G}$, true labels Y, and features X, as well an *auxiliary dataset* containing both the noisy groups $\hat{G}$ and the true groups G. From the main dataset, one obtains estimates of the joint distributions $(\hat{Y}(\theta), Y, \hat{G})$; from the auxiliary dataset, one obtains estimates of the joint distributions $(\hat{G}, G)$ and a noise model $P(G = j|\hat{G} = k)\ \forall j \in \mathcal{G}, k \in \hat{\mathcal{G}}$.

These estimates are used to associate each example with a vector of weights, where each weight is an estimated probability that the example belongs to the true group j. Specifically, suppose that we have a function $w : \mathcal{G} \times \{0, 1\} \times \{0, 1\} \times \hat{\mathcal{G}} \to [0, 1]$, where $w(j \mid \hat{y}, y, k)$ estimates $P(G = j|\hat{Y}(\theta) = \hat{y}, Y = y, \hat{G} = k)$. We rewrite the fairness constraint $E[h(\theta, X, Y)|G = j] = \frac{E[h(\theta,X,Y)P(G=j|\hat{Y}(\theta),Y,\hat{G})]}{P(G=j)}$ (derivation in Appendix 3.12), and estimate

this using w. We also show how h can be adapted to the *equality of opportunity* setting in Appendix 3.12.

Given the main dataset and auxiliary dataset, we limit the possible values of the function $w(j \mid \hat{y}, y, k)$ using the law of total probability (as in [128]). The set of possible functions w is given by:

$$\mathcal{W}(\theta) = \left\{ w : \begin{smallmatrix} \sum_{\hat{y},y\in\{0,1\}} w(j|\hat{y},y,k) P(\hat{Y}(\theta)=\hat{y}, Y=y|\hat{G}=k) = P(G=j|\hat{G}=k), \\ \sum_{j=1}^{m} w(j|\hat{y},y,k)=1, w(j|\hat{y},y,k)\geqslant 0 \quad \forall \hat{y},y\in\{0,1\}, j\in\mathcal{G}, k\in\hat{\mathcal{G}} \end{smallmatrix} \right\}. \tag{3.5}$$

The robust fairness criteria can now be written in terms of $\mathcal{W}(\theta)$ as:

$$\max_{w\in\mathcal{W}(\theta)} g_j(\theta, w) \leqslant 0, \;\; \forall j \in \mathcal{G} \quad \text{where} \quad g_j(\theta, w) = \frac{\mathbb{E}[h(\theta, X, Y) w(j|\hat{Y}(\theta), Y, \hat{G})]}{P(G=j)}. \tag{3.6}$$

Robust optimization with soft group assignments

We extend Kallus et al. [128]'s work by formulating a robust optimization problem using soft group assignments. Combining the robust fairness criteria above with the training objective, we propose:

$$\min_{\theta\in\Theta} \quad f(\theta) \quad \text{s.t.} \max_{w\in\mathcal{W}(\theta)} g_j(\theta, w) \leqslant 0, \;\; \forall j \in \mathcal{G}, \tag{3.7}$$

where Θ denotes the space of model parameters. Any feasible solution is guaranteed to satisfy the original fairness criteria with respect to the true groups. Using a Lagrangian, problem (3.7) can be rewritten as:

$$\min_{\theta\in\Theta} \max_{\lambda\in\Lambda} \mathcal{L}(\theta, \lambda) \tag{3.8}$$

where the Lagrangian $\mathcal{L}(\theta, \lambda) = f(\theta) + \sum_{j=1}^{m} \lambda_j \max_{w\in\mathcal{W}(\theta)} g_j(\theta, w)$, and $\Lambda \subseteq \mathbb{R}_+^m$.

When solving this optimization problem, we use the empirical finite-sample versions of each expectation. As described in Proposition 9 of Kallus et al. [128], the inner maximization (3.6) over $w \in \mathcal{W}(\theta)$ can be solved as a linear program for a given fixed θ. However, the Lagrangian problem (3.8) is not as straightforward to optimize, since the feasible set $\mathcal{W}(\theta)$ depends on θ through $\hat{Y}$. While in general the pointwise maximum of convex functions is convex, the dependence of $\mathcal{W}(\theta)$ on θ means that even if $g_j(\theta, w)$ were convex, $\max_{w\in\mathcal{W}(\theta)} g_j(\theta, w)$ is not necessarily convex. We first introduce a theoretically ideal algorithm that we prove converges to an optimal, feasible solution. This ideal algorithm relies on a minimization oracle, which is not always computationally tractable. Therefore, we further provide a practical algorithm using gradient methods that mimics the ideal algorithm in structure and computationally tractable, but does not share the same convergence guarantees.

Ideal algorithm

The minimax problem in (3.8) can be interpreted as a zero-sum game between the θ-player and λ-player. In Algorithm 3, we provide an iterative procedure for solving (3.8), where at each step, the θ-player performs a full optimization, i.e., a *best response* over θ, and the λ-player responds with a gradient ascent update on λ.

For a fixed θ, the gradient of the Lagrangian $\mathcal{L}$ with respect to λ is given by $\partial\mathcal{L}(\theta,\lambda)/\partial\lambda_j = \max_{w\in\mathcal{W}(\theta)} g_j(\theta,w)$, which is a linear program in w. The challenging part, however, is the best response over θ; that is, finding a solution $\min_\theta \mathcal{L}(\theta,\lambda)$ for a given λ, as this involves a max over constraints $\mathcal{W}(\theta)$ which depend on θ. To implement this best response, we formulate a nested minimax problem that decouples this intricate dependence on θ, by introducing Lagrange multipliers for the constraints in $\mathcal{W}(\theta)$. We then solve this problem with an oracle that jointly minimizes over both θ and the newly introduced Lagrange multipliers (details in Algorithm 5 in Appendix 3.13).

The output of the best-response step is a stochastic classifier with a distribution $\hat{\theta}^{(t)}$ over a finite set of θs. Algorithm 3 then returns the average of these distributions, $\bar{\theta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\theta}^t$, over T iterations. By extending recent results on constrained optimization [53], we show in Appendix 3.13 that the output $\bar{\theta}$ is near-optimal and near-feasible for the robust optimization problem in (3.7). That is, for a given $\epsilon > 0$, by picking T to be large enough, we have that the objective $\mathbb{E}_{\theta\sim\bar{\theta}}[f(\theta)] \leqslant f(\theta^*) + \epsilon$, for any θ^* that is feasible, and the expected violations in the robust constraints are also no more than ϵ.

Algorithm 3 *Ideal* Algorithm

Require: learning rate $\eta_\lambda > 0$, estimates of $P(G = j|\hat{G} = k)$ to specify $\mathcal{W}(\theta)$, ρ, ρ'

1: **for** $t = 1,\ldots,T$ **do**
2: *Best response on* θ: run the oracle-based Algorithm 5 to find a distribution $\hat{\theta}^{(t)}$ over Θ s.t. $\mathbb{E}_{\theta\sim\hat{\theta}^{(t)}}\left[\mathcal{L}(\theta,\lambda^{(t)})\right] \leqslant \min_{\theta\in\Theta}\mathcal{L}(\theta,\lambda^{(t)}) + \rho$.
3: *Estimate gradient* $\nabla_\lambda\mathcal{L}(\hat{\theta}^{(t)},\lambda^{(t)})$: for each $j\in\mathcal{G}$, choose $\delta_j^{(t)}$ s.t. $\delta_j^{(t)} \leqslant \mathbb{E}_{\theta\sim\hat{\theta}^{(t)}}\left[\max_{w\in\mathcal{W}(\theta)} g_j(\theta,w)\right] \leqslant \delta_j^{(t)} + \rho'$
4: *Ascent step on* λ: $\tilde{\lambda}_j^{(t+1)} \leftarrow \lambda_j^{(t)} + \eta_\lambda\,\delta_j^{(t)},\ \ \forall j\in\mathcal{G};\quad \lambda^{(t+1)} \leftarrow \Pi_\Lambda(\tilde{\lambda}^{(t+1)})$
5: **end for**
6: **return** $\bar{\theta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\theta}^{(t)}$

Practical algorithm

Algorithm 3 is guaranteed to converge to a near-optimal, near-feasible solution, but may be computationally intractable and impractical for the following reasons. First, the algorithm needs a nonconvex minimization oracle to compute a best response over θ. Second, there

are multiple levels of nesting, making it difficult to scale the algorithm with mini-batch or stochastic updates. Third, the output is a distribution over multiple models, which can be be difficult to use in practice [174].

Therefore, we supplement Algorithm 3 with a practical algorithm, Algorithm 6 (see Appendix 3.14) that is similar in structure, but approximates the inner best response routine with two simple steps: a maximization over $w \in \mathcal{W}(\theta^{(t)})$ using a linear program for the current iterate $\theta^{(t)}$, and a gradient step on θ at the maximizer $w^{(t)}$. Algorithm 6 leaves room for other practical modifications such as using stochastic gradients. We provide further discussion in Appendix 3.14.

3.7 Experiments

We compare the performance of the naïve approach and the two robust optimization approaches (DRO and soft group assignments) empirically using two datasets from UCI [70] with different constraints. For both datasets, we stress-test the performance of the different algorithms under different amounts of noise between the true groups G and the noisy groups $\hat{G}$. We take l to be the hinge loss. The specific constraint violations measured and additional training details can be found in Appendix 3.15.

Generating noisy protected groups: Given the true protected groups, we synthetically generate noisy protected groups by selecting a fraction γ of data uniformly at random. For each selected example, we perturb the group membership to a different group also selected uniformly at random from the remaining groups. This way, for a given γ, $P(\hat{G} \neq G) \approx P(\hat{G} \neq G | G = j) \approx \gamma$ for all groups $j, k \in \mathcal{G}$. We evaluate the performance of the different algorithms ranging from small to large amounts of noise: $\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

Case study 1 (Adult): equality of opportunity

We use the Adult dataset from UCI [70] collected from 1994 US Census, which has 48,842 examples and 14 features (details in Appendix 3.15). The classification task is to determine whether an individual makes over \$50K per year. For the true groups, we use $m = 3$ race groups of "white," "black," and "other." As done by [54, 87, 223], we enforce *equality of opportunity* by equalizing true positive rates (TPRs). Specifically, we enforce that the TPR conditioned on each group is greater than or equal to the overall TPR on the full dataset with some slack α, which produces m true group fairness criteria, $\{g_j^{\text{TPR}}(\theta) \leqslant 0\} \quad \forall j \in \mathcal{G}$ (details on the constraint function h in Appendix 3.11 and 3.12).

Case study 2 (Credit): equalized odds

We consider another application of group-based fairness constraints to credit default prediction. Fourcade and Healy [86] provide an in depth study of the effect of credit scoring techniques on the credit market, showing that this scoring system can perpetuate inequity.

Enforcing group-based fairness with credit default predictions has been considered in a variety of prior works [5, 26, 30, 33, 87, 97, 110, 216]. Following Hardt et al. [110] and Grari et al. [97], we enforce *equalized odds* [110] by equalizing both true positive rates (TPRs) and false positive rates (FPRs) across groups.

We use the "default of credit card clients" dataset from UCI [70] collected by a company in Taiwan [221], which contains 30,000 examples and 24 features (details in Appendix 3.15). The classification task is to determine whether an individual defaulted on a loan. We use $m = 3$ groups based on education levels: "graduate school," "university," and "high school/other" (the use of education in credit lending has previously been studied in the algorithmic fairness and economics literature [30, 93, 146]). We constrain the TPR conditioned on each group to be greater than or equal to the overall TPR on the full dataset with a slack α, and the FPR conditioned on each group to be less than or equal to the overall FPR on the full dataset. This produces $2m$ true group-fairness criteria, $\{g_j^{\text{TPR}}(\theta) \leqslant 0, g_j^{\text{FPR}}(\theta) \leqslant 0\} \ \ \forall j \in \mathcal{G}$ (details on constraint functions h in Appendix 3.11 and 3.12).

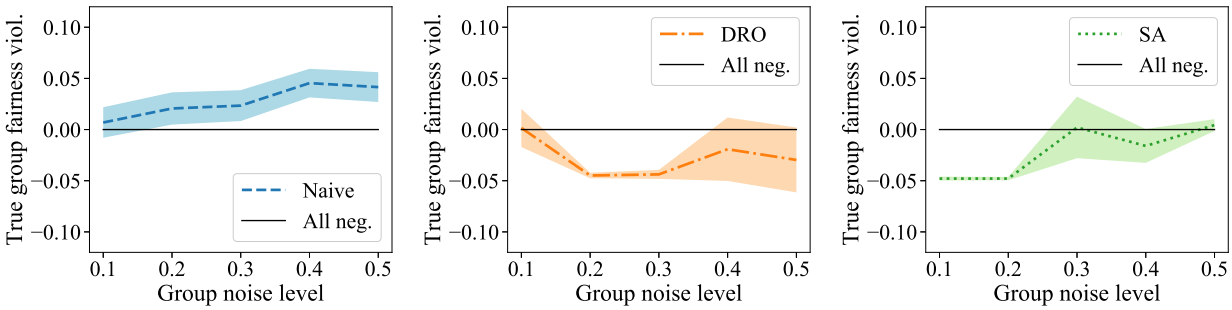


Figure 3.1: Case study 1 (Adult): maximum true group constraint violations on test set for the Naive, DRO, and soft assignments (SA) approaches for different group noise levels γ on the Adult dataset (mean and standard error over 10 train/val/test splits). The black solid line represents the performance of the trivial "all negatives" classifier, which has constraint violations of 0. A negative violation indicates satisfaction of the fairness constraints on the true groups.

Results

In case study 1 (Adult), the unconstrained model achieves an error rate of 0.1447 ± 0.0012 (mean and standard error over 10 splits) and a maximum constraint violation of 0.0234 ± 0.0164 on test set with respect to the true groups. The model that assumes knowledge of the true groups achieves an error rate of 0.1459 ± 0.0012 and a maximum constraint violation of -0.0469 ± 0.0068 on test set with respect to the true groups. As a sanity check, this demonstrates that when given access to the true groups, it is possible to satisfy the constraints on the test set with a reasonably low error rate.

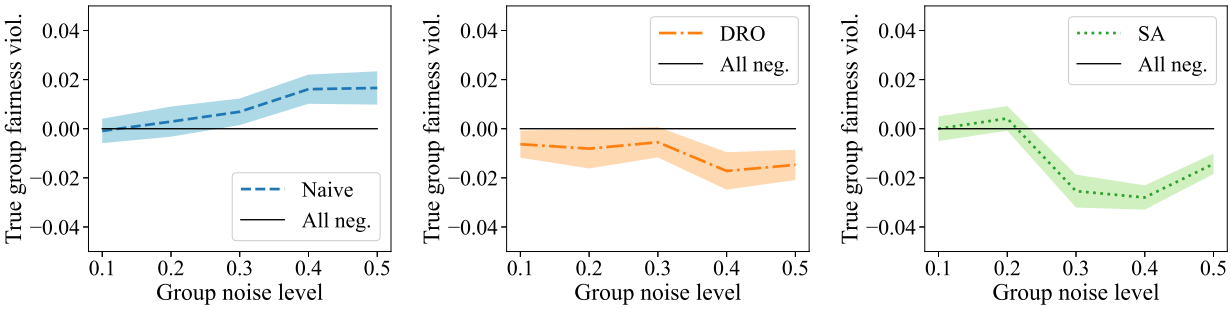


Figure 3.2: Case study 2 (Credit): maximum true group constraint violations on test set for the Naive, DRO, and soft assignments (SA) approaches for different group noise levels γ on the Credit dataset (mean and standard error over 10 train/val/test splits). This figure shows the max constraint violation over all TPR and FPR constraints, and Figure 3.6 in Appendix 3.15 shows the breakdown of these constraint violations into the max TPR and the max FPR constraint violations.

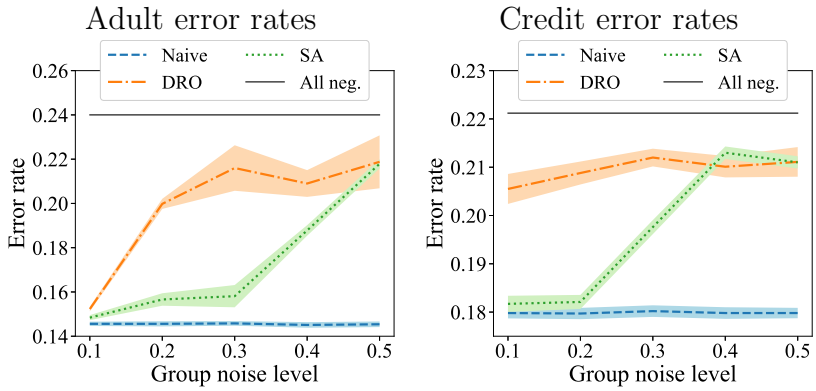


Figure 3.3: Error rates on test set for different group noise levels γ on the Adult dataset (*left*) and the Credit dataset (*right*) (mean and standard error over 10 train/val/test splits). The black solid line represents the performance of the trivial "all negatives" classifier. The soft assignments (SA) approach achieves lower error rates than DRO, and as the noise level increases, the gap in error rate between the naive approach and each robust approach increases.

In case study 2 (Credit), the unconstrained model achieves an error rate of 0.1797 ± 0.0013 (mean and standard error over 10 splits) and a maximum constraint violation of 0.0264 ± 0.0071 on the test set with respect to the true groups. The constrained model that assumes knowledge of the true groups achieves an error rate of 0.1796 ± 0.0011 and a maximum constraint violation of -0.0105 ± 0.0070 on the test set with respect to the true groups. For this dataset, it was possible to satisfy the constraints with approximately the same error rate

on test as the unconstrained model. Note that the unconstrained model achieved a lower error rate on the train set than the constrained model (0.1792 ± 0.0015 unconstrained vs. 0.1798 ± 0.0024 constrained).

For both case studies, Figures 3.1 and 3.2 show that the robust approaches DRO (*center*) and soft group assignments (SA) (*right*) satisfy the constraints on average for all noise levels. As the noise level increases, the naïve approach (*left*) has increasingly higher true group constraint violations. The DRO and SA approaches come at a cost of a higher error rate than the naïve approach (Figure 3.3). The error rate of the naïve approach is close to the model optimized with constraints on the true groups G, regardless of the noise level γ. However, as the noise increases, the naïve approach no longer controls the fairness violations on the true groups G, even though it does satisfy the constraints on the noisy groups $\hat{G}$ (Figures 3.4 and 3.7 in Appendix 3.15). DRO generally suffers from a higher error rate compared to SA (Figure 3.3), illustrating the conservatism of the DRO approach.

3.8 Conclusion and future directions

We explore the practical problem of enforcing group-based fairness for binary classification given noisy protected group information. In addition to providing new theoretical analysis of the naïve approach of only enforcing fairness on the noisy groups, we also propose two new robust approaches that guarantee satisfaction of the fairness criteria on the true groups. For the DRO approach, Lemma 3.4.2 gives a theoretical bound on the TV distance to use in the optimization problem. For the soft group assignments approach, we provide a theoretically ideal algorithm and a practical alternative algorithm for satisfying the robust fairness criteria proposed by Kallus et al. [128] while minimizing a training objective. We empirically show that both of these approaches managed to satisfy the constraints with respect to the true groups, even under difficult noise models.

In follow-up work, Narasimhan et al. [175] provide a general method for enforcing a large number of constraints at once, and enforce constraints concurrently on many possible realizations of noisy protected groups under a given noise model. This can be seen as an extension of the Soft Group Assignments approach that we propose in Section 3.6, which Narasimhan et al. [175] describe in their Appendix.

One additional avenue of future work is to empirically compare the robust approaches when the noisy groups have different dimensionality from the true groups (Appendix 3.11). Second, the looseness of the bound in Lemma 3.4.2 can lead to over-conservatism of the DRO approach, suggesting a need to better calibrate the DRO neighborhood. Finally, it would be valuable to study the impact of distribution mismatch between the main dataset and the auxiliary dataset.

3.9 Discussions on the broader impact

As machine learning is increasingly employed in high stakes environments, any potential application has to be scrutinized to ensure that it will not perpetuate, exacerbate, or create new injustices. Aiming to make machine learning algorithms themselves intrinsically fairer, more inclusive, and more equitable plays an important role in achieving that goal. Group-based fairness [87, 110] is a popular approach that the machine learning community has used to define and evaluate fair machine learning algorithms. Until recently, such work has generally assumed access to clean, correct protected group labels in the data. Our work addresses the technical challenge of enforcing group-based fairness criteria under noisy, unreliable, or outdated group information. However, we emphasize that this technical improvement alone does not necessarily lead to an algorithm having positive societal impact, for reasons that we now delineate.

Choice of fairness criteria First, our work does not address the choice of the group-based fairness criteria. Many different algorithmic fairness criteria have been proposed, with varying connections to prior sociopolitical framing [119, 176]. From an algorithmic standpoint, these different choices of fairness criteria have been shown to lead to very different prediction outcomes and tradeoffs [87]. Furthermore, even if a mathematical criterion may seem reasonable (e.g., equalizing positive prediction rates with *demographic parity*), Liu et al. [155] show that the long-term impacts may not always be desirable, and the choice of criteria should be heavily influenced by domain experts, along with awareness of tradeoffs.

Choice of protected groups In addition to the specification of fairness criteria, our work also assumes that the true protected group labels have been pre-defined by the practitioner. However, in real applications, the selection of appropriate true protected group labels is itself a nontrivial issue.

First, the measurement and delineation of these protected groups should not be overlooked, as "the process of drawing boundaries around distinct social groups for fairness research is fraught; the construction of categories has a long history of political struggle and legal argumentation" [108]. Important considerations include the context in which the group labels were collected, who chose and collected them, and what implicit assumptions are being made by choosing these group labels. One example is the operationalization of race in the context of algorithmic fairness. Hanna et al. [108] critiques "treating race as an attribute, rather than a structural, institutional, and relational phenomenon." The choice of categories surrounding gender identity and sexual orientation have strong implications and consequences as well [98], with entire fields dedicated to critiquing these constructs. Jacobs and Wallach [122] provide a general framework for understanding measurement issues for these sensitive attributes in the machine-learning setting, building on foundational work from the social sciences [24].

Another key consideration when defining protected groups is problems of *intersectionality* [55, 115]. Group-based fairness criteria inherently do not consider within-group inequality

[132]. Even if we are able to enforce fairness criteria robustly for a given set of groups, the intersections of groups may still suffer [41].

Domain specific considerations Finally, we emphasize that group-based fairness criteria simply may not be sufficient to mitigate problems of significant background injustice in certain domains. Abebe et al. [1] argue that computational methods have mixed roles in addressing social problems, where they can serve as *diagnostics*, *formalizers*, and *rebuttals*, and also that "computing acts as synecdoche when it makes long-standing social problems newly salient in the public eye." Moreover, the use of the algorithm itself may perpetuate inequity, and in the case of credit scoring, create stratifying effects of economic classifications that shape life-chances [86]. We emphasize the importance of domain specific considerations ahead of time before applying any algorithmic solutions (even "fair" ones) in sensitive and impactful settings.

3.10 Appendix: Proofs for Section 3.4

This section provides proofs and definitions details for the theorems and lemmas presented in Section 3.4.

Proofs for TV distance

Definition 3.10.1. (TV distance) Let $c(x, y) = \mathbb{1}(x \neq y)$ be a metric, and let π be a coupling between probability distributions p and q. Define the total variation (TV) distance between two distributions p, q as

$$TV(p, q) = \inf_{\pi} \mathop{\mathbb{E}}_{X,Y\sim\pi}[c(X, Y)]$$
$$\text{s.t.} \int \pi(x, y)dy = p(x), \int \pi(x, y)dx = q(y).$$

Theorem 3.4.1. *(proof in Appendix 3.10.) Suppose a model with parameters θ satisfies fairness criteria with respect to the noisy groups $\hat{G}$: $\hat{g}_j(\theta) \leqslant 0, \ \ \forall j \in \mathcal{G}$. Suppose $|h(\theta, x_1, y_1) - h(\theta, x_2, y_2)| \leqslant 1$ for any $(x_1, y_1) \neq (x_2, y_2)$. If $TV(p_j, \hat{p}_j) \leqslant \gamma_j$ for all $j \in \mathcal{G}$, then the fairness criteria with respect to the true groups G will be satisfied within slacks γ_j for each group: $g_j(\theta) \leqslant \gamma_j, \ \ \forall j \in \mathcal{G}$.*

Proof. For any group label j,

$$g_j(\theta) = g_j(\theta) - \hat{g}_j(\theta) + \hat{g}_j(\theta) \leqslant |g_j(\theta) - \hat{g}_j(\theta)| + \hat{g}_j(\theta).$$

By Kantorovich-Rubenstein theorem (provided here as Theorem 3.10.2), we also have

$$|\hat{g}_j(\theta) - g_j(\theta)| = |\mathop{\mathbb{E}}_{X,Y\sim\hat{p}_j}[h(\theta, X, Y)] - \mathop{\mathbb{E}}_{X,Y\sim p_j}[h(\theta, X, Y)]| \leqslant TV(p_j, \hat{p}_j).$$

By assumption that θ satisifes fairness constraints with respect to the noisy groups $\hat{G}$, $\hat{g}_j(\theta) \leqslant 0$. Thus, we have the desired result that $g_j(\theta) \leqslant TV(p_j, \hat{p}_j) \leqslant \gamma_j$.

Note that if p_j and $\hat{p}_j$ are discrete, then the TV distance $TV(p_j, \hat{p}_j)$ could be very large. In that case, the bound would still hold, but would be loose. ■

Theorem 3.10.2. *(Kantorovich-Rubinstein).*[1] *Call a function f Lipschitz in c if $|f(x) - f(y)| \leqslant c(x, y)$ for all x, y, and let $\mathcal{L}(c)$ denote the space of such functions. If c is a metric, then we have*

$$W_c(p, q) = \sup_{f \in \mathcal{L}(c)} \mathop{\mathbb{E}}_{X \sim p}[f(X)] - \mathop{\mathbb{E}}_{X \sim q}[f(X)].$$

As a special case, take $c(x, y) = \mathbb{I}(x \neq y)$ (corresponding to TV distance). Then $f \in \mathcal{L}(c)$ if and only if $|f(x) - f(y)| \leqslant 1$ for all $x \neq y$. By translating f, we can equivalently take the supremum over all f mapping to $[0, 1]$. This says that

$$TV(p, q) = \sup_{f: \mathcal{X} \to [0,1]} \mathop{\mathbb{E}}_{X \sim p}[f(X)] - \mathop{\mathbb{E}}_{X \sim q}[f(X)]$$

Lemma 3.4.2. *(proof in Appendix 3.10.) Suppose $P(G = j) = P(\hat{G} = j)$ for a given $j \in \mathcal{G}$. Then $TV(p_j, \hat{p}_j) \leqslant P(G \neq \hat{G} | G = j)$.*

Proof. For probability measures p_i and $\hat{p}_i$, the TV distance is given by

$$TV(p_i, \hat{p}_i) = \sup\{|p_i(A) - \hat{p}_i(A)| : A \text{ is a measurable event}\}.$$

Fix A to be any measurable event for both p_i and $\hat{p}_i$. This means that A is also a measurable event for p, the distribution of the random variables X, Y. By definition of p_i, $p_i(A) = P(A|G = i)$. Then

$$\begin{aligned}
|p_i(A) - \hat{p}_i(A)| &= |P(A|G = i) - P(A|\hat{G} = i)| \\
&= |P(A|G = i, \hat{G} = i)P(\hat{G} = i|G = i) \\
&\quad + P(A|G = i, \hat{G} \neq i)P(\hat{G} \neq i|G = i) \\
&\quad - P(A|\hat{G} = i, G = i)P(G = i|\hat{G} = i) \\
&\quad - P(A|\hat{G} = i, G \neq i)P(G \neq i|\hat{G} = i)| \\
&= |P(A|G = i, \hat{G} = i) \Big(P(\hat{G} = i|G = i) - P(G = i|\hat{G} = i)\Big) \\
&\quad - P(\hat{G} \neq G|G = i) \Big(P(A|G = i, \hat{G} \neq i) - P(A|\hat{G} = i, G \neq i)\Big)| \\
&= |0 - P(\hat{G} \neq G|G = i) \Big(P(A|G = i, \hat{G} \neq i) - P(A|\hat{G} = i, G \neq i)\Big)| \\
&\leqslant P(\hat{G} \neq G|G = i)
\end{aligned}$$

[1] Edwards, D.A. On the Kantorovich–Rubinstein theorem. *Expositiones Mathematicae*, 20(4):387-398, 2011.

The second equality follows from the law of total probability. The third and the fourth equalities follow from the assumption that $P(G = i) = P(\hat{G} = i)$, which implies that $P(\hat{G} = G|G = i) = P(G = \hat{G}|\hat{G} = i)$ since

$$P(G = \hat{G}|G = i) = \frac{P(G = \hat{G}, G = i)}{P(G = i)} = \frac{P(G = \hat{G}, \hat{G} = i)}{P(\hat{G} = i)} = P(G = \hat{G}|\hat{G} = i).$$

This further implies that $P(\hat{G} \neq i|G = i) = P(G \neq i|\hat{G} = i)$.

Since $|p_i(A) - \hat{p}_i(A)| \leqslant P(\hat{G} \neq G|G = i)$ for any measurable event A, the supremum over all events A is also bounded by $P(\hat{G} \neq G|G = i)$. This gives the desired bound on the TV distance. ■

Generalization to Wasserstein distances

Theorem 3.4.1 can be directly extended to loss functions that are Lipschitz in other metrics. To do so, we first provide a more general definition of Wasserstein distances:

Definition 3.10.3. (Wasserstein distance) Let $c(x, y)$ be a metric, and let π be a coupling between p and q. Define the Wasserstein distance between two distributions p, q as

$$W_c(p, q) = \inf_{\pi} \mathop{\mathbb{E}}_{X,Y\sim\pi}[c(X, Y)] \quad \text{s.t.} \int \pi(x, y)dy = p(x), \int \pi(x, y)dx = q(y).$$

As a familiar example, if $c(x, y) = ||x - y||_2$, then W_c is the earth-mover distance, and $\mathcal{L}(c)$ is the class of 1-Lipschitz functions. Using the Wasserstein distance W_c under different metrics c, we can bound the fairness violations for constraint functions h beyond those specified for the TV distance in Theorem 3.4.1.

Theorem 3.10.4. *Suppose a model with parameters θ satisfies fairness criteria with respect to the noisy groups $\hat{G}$:*

$$\hat{g}_j(\theta) \leqslant 0 \ \ \forall j \in \mathcal{G}.$$

Suppose the function h satisfies $|h(\theta, x_1, y_1) - h(\theta, x_2, y_2)| \leqslant c((x_1, y_1), (x_2, y_2))$ for any $(x_1, y_1) \neq (x_2, y_2)$ w.r.t a metric c. If $W_c(p_j, \hat{p}_j) \leqslant \gamma_j$ for all $j \in \mathcal{G}$, then the fairness criteria with respect to the true groups G will be satisfied within slacks γ_j for each group:

$$g_j(\theta) \leqslant \gamma_j \quad \forall j \in \mathcal{G}.$$

Proof. By the triangle inequality, for any group label j,

$$|g_j(\theta) - g(\theta)| \leqslant |g_j(\theta) - \hat{g}_j(\theta)| + \hat{g}_j(\theta)$$

By Kantorovich-Rubenstein theorem (provided here as Theorem 3.10.2), we also have

$$\begin{aligned}|\hat{g}_j(\theta) - g_j(\theta)| &= |\mathop{\mathbb{E}}_{X,Y\sim\hat{p}_j}[h(\theta, X, Y)] - \mathop{\mathbb{E}}_{X,Y\sim p_j}[h(\theta, X, Y)]| \\ &\leqslant W_c(p_j, \hat{p}_j).\end{aligned}$$

By the assumption that θ satisifes fairness constraints with respect to the noisy groups $\hat{G}$, $\hat{g}_j(\theta) \leqslant 0$. Therefore, combining these with the triangle inequality, we get the desired result. ■

3.11 Appendix: Details on DRO formulation for TV distance

Here we describe the details on solving the DRO problem (3.3) with TV distance using the empirical Lagrangian formulation. We also provide the pseudocode we used for the projected gradient-based algorithm to solve it.

Empirical Lagrangian Formulation

We rewrite the constrained optimization problem (3.3) as a minimax problem using the Lagrangian formulation. We also convert all expectations into expectations over empirical distributions given a dataset of n samples $(X_1, Y_1, G_1), ..., (X_n, Y_n, G_n)$.

Let n_j denote the number of samples that belong to a true group $G = j$. Let the empirical distribution $\hat{p}_j \in \mathbb{R}^n$ be a vector with i-th entry $\hat{p}_j^i = \frac{1}{n_j}$ if the i-th example has a noisy group membership $\hat{G}_i = j$, and 0 otherwise. Replacing all expectations with expectations over the appropriate empirical distributions, the empirical form of (3.3) can be written as:

$$\begin{aligned}\min_{\theta} \quad & \frac{1}{n}\sum_{i=1}^{n} l(\theta, X_i, Y_i) \\ \text{s.t.} \quad & \max_{\tilde{p}_j \in \mathbb{B}_{\gamma_j}(\hat{p}_j)} \sum_{i=1}^{n} \tilde{p}_j^i h(\theta, X_i, Y_i) \leqslant 0 \quad \forall j \in \mathcal{G}\end{aligned} \tag{3.9}$$

where $\mathbb{B}_{\gamma_j}(\hat{p}_j) = \{\tilde{p}_j \in \mathbb{R}^n : \frac{1}{2}\sum_{i=1}^{n} |\tilde{p}_j^i - \hat{p}_j^i| \leqslant \gamma_j, \sum_{i=1}^{n} \tilde{p}_j^i = 1, \tilde{p}_j^i \geqslant 0 \quad \forall i = 1, ..., n\}$.

For ease of notation, for $j \in \{1, 2, ..., m\}$, let

$$f(\theta) = \frac{1}{n}\sum_{i=1}^{n} l(\theta, X_i, Y_i)$$

$$f_j(\theta, \tilde{p}_j) = \sum_{i=1}^{n} \tilde{p}_j^i h(\theta, X_i, Y_i).$$

Then the Lagrangian of the empirical formulation (3.9) is

$$\mathcal{L}(\theta, \lambda) = f(\theta) + \sum_{j=1}^{m} \lambda_j \max_{\tilde{p}_j \in \mathbb{B}_\gamma(\hat{p}_j)} f_j(\theta, \tilde{p}_j)$$

and problem (3.9) can be rewritten as

$$\min_{\theta} \max_{\lambda \geqslant 0} f(\theta) + \sum_{j=1}^{m} \lambda_j \max_{\tilde{p}_j \in \mathbb{B}_\gamma(\hat{p}_j)} f_j(\theta, \tilde{p}_j)$$

Moving the inner max out of the sum and rewriting the constraints as ℓ_1-norm constraints:

$$\begin{aligned} \min_{\theta} \max_{\lambda \geqslant 0} \max_{\substack{\tilde{p}_j \in \mathbb{R}^n, \tilde{p}_j \geqslant 0, \\ j=1,\ldots,m}} & f(\theta) + \sum_{j=1}^{m} \lambda_j f_j(\theta, \tilde{p}_j) \\ & \text{s.t. } ||\tilde{p}_j - \hat{p}_j||_1 \leqslant 2\gamma_j, \quad ||\tilde{p}_j||_1 = 1 \quad \forall j \in \{1, ..., m\} \end{aligned} \tag{3.10}$$

Since projections onto the ℓ_1-ball can be done efficiently [72], we can solve problem (3.10) using a projected gradient descent ascent (GDA) algorithm. This is a simplified version of the algorithm introduced by Namkoong and Duchi [172] for solving general classes of DRO problems. We provide pseudocode in Algorithm 4.

Projected GDA Algorithm for DRO

Algorithm 4 Project GDA Algorithm

Require: learning rates $\eta_\theta > 0$, $\eta_\lambda > 0$, $\eta_p > 0$, estimates of $P(G \neq \hat{G} | \hat{G} = j)$ to specify γ_j.

1: **for** $t = 1, \ldots, T$ **do**
2: *Descent step on* θ*:*
$\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta_\theta \nabla_\theta f(\theta^{(t)}) - \eta_\theta \sum_{j=1}^{m} \lambda_j^{(t)} \nabla_\theta f_j(\theta^{(t)}, \tilde{p}_j^{(t)})$
3: *Ascent step on* λ:
$\lambda_j^{(t+1)} \leftarrow \lambda_j^{(t)} + \eta_\lambda f_j(\theta, \tilde{p}_j^{(t)})$
4: **for** $j = 1, ..., m$ **do**
5: *Ascent step on* $\tilde{p}_j$: $\tilde{p}_j^{(t+1)} \leftarrow \tilde{p}_j^{(t)} + \eta_p \lambda_j^{(t)} \nabla_{\tilde{p}_j} f_j(\theta^{(t)}, \tilde{p}_j^{(t)})$
6: *Project* $\tilde{p}_j^{(t+1)}$ *onto* ℓ_1*-norm constraints*: $||\tilde{p}_j^{(t+1)} - \hat{p}_j||_1 \leqslant 2\gamma_j, ||\tilde{p}_j^{(t+1)}||_1 = 1$
7: **end for**
8: **end for**
9: **return** $\theta^{(t^*)}$ where t^* denotes the *best* iterate that satisfies the constraints in (3.3) with the lowest objective.

Equalizing TPRs and FPRs using DRO

In the two case studies in Section 3.7, we enforce *equality of opportunity* and *equalized odds* [110] by equalizing true positive rates (TPRs) and/or false positive rates (FPRs) within some slack α. In this section, we describe in detail the implementation of the constraints for equalizing TPRs and FPRs under the DRO approach.

To equalize TPRs with slack α under the DRO approach, we set

$$\tilde{g}_j^{\text{TPR}}(\theta) = \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)]} - \frac{\mathbb{E}_{X,Y\sim \tilde{p}_j}[\mathbb{1}(Y=1)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim \tilde{p}_j}[\mathbb{1}(Y=1)]} - \alpha. \tag{3.11}$$

The first term corresponds to the TPR for the full population. The second term estimates the TPR for group j. Setting $\alpha = 0$ exactly equalizes true positive rates.

To equalize FPRs with slack α under the DRO approach, we set

$$\tilde{g}_j^{\text{FPR}}(\theta) = \frac{\mathbb{E}_{X,Y\sim \tilde{p}_j}[\mathbb{1}(Y=0)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim \tilde{p}_j}[\mathbb{1}(Y=0)]} - \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)]} - \alpha. \tag{3.12}$$

The first term estimates the FPR for group j. The second term corresponds to the FPR for the full population. Setting $\alpha = 0$ exactly equalizes false positive rates.

To equalize TPRs for Case Study 1, we apply m constraints, $\left\{\max_{\tilde{p}_j : TV(\tilde{p}_j,\hat{p}_j)\leqslant\gamma_j, \tilde{p}_j \ll p} \tilde{g}_j^{\text{TPR}}(\theta) \leqslant 0\right\}\ \forall j \in \mathcal{G}$.

To equalize both TPRs and FPRs simultaneously for Case Study 2, we apply $2m$ constraints, $\left\{\max_{\tilde{p}_j : TV(\tilde{p}_j,\hat{p}_j)\leqslant\gamma_j, \tilde{p}_j \ll p} \tilde{g}_j^{\text{TPR}}(\theta) \leqslant 0, \max_{\tilde{p}_j : TV(\tilde{p}_j,\hat{p}_j)\leqslant\gamma_j, \tilde{p}_j \ll p} \tilde{g}_j^{\text{FPR}}(\theta) \leqslant 0\right\}\ \forall j \in \mathcal{G}$.

$h(\theta, X, Y)$ for equalizing TPRs and FPRs

Since the notation in Section 3.5 and in the rest of the paper uses generic functions h to express the group-specific constraints, we show in Lemma 3.11.1 that the constraint using $\tilde{g}_j^{\text{TPR}}(\theta)$ in Equation (3.11) can also be written as an equivalent constraint in the form of Equation (3.3), as

$$\tilde{g}_j^{\text{TPR}}(\theta) = \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[h^{\text{TPR}}(\theta, X, Y)]$$

for some function $h^{\text{TPR}} : \Theta \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

Lemma 3.11.1. *Denote $\hat{Y}$ as $\mathbb{1}(\phi(X;\theta) > 0)$. Let $h^{TPR}(\theta, X, Y)$ be given by*

$$h^{TPR}(\theta, X, Y) = \frac{1}{2}\left(-\mathbb{1}(\hat{Y}=1, Y=1) - \mathbb{1}(Y=1)\left(\alpha - \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)]}\right)\right).$$

Then

$$\frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)]} - \frac{\mathbb{E}_{X,Y\sim \tilde{p}_j}[\mathbb{1}(Y=1)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim \tilde{p}_j}[\mathbb{1}(Y=1)]} - \alpha \leqslant 0$$
$$\iff \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[h^{TPR}(\theta, X, Y)] \leqslant 0.$$

Proof. Substituting the given function $h^{\text{TPR}}(\theta, X, Y)$, and using the fact that $\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=1)] \geqslant 0$:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[h^{\text{TPR}}(\theta, X, Y)] \leqslant 0 \\
\iff &\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}\left[\frac{1}{2}\left(-\mathbb{1}(\hat{Y}=1, Y=1) - \mathbb{1}(Y=1)\left(\alpha - \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)]}\right)\right)\right] \leqslant 0 \\
\iff &-\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(\hat{Y}=1, Y=1)] - \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}\left[\mathbb{1}(Y=1)\left(\alpha - \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)]}\right)\right] \leqslant 0 \\
\iff &-\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(\hat{Y}=1, Y=1)] - \alpha \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=1)] \\
&\quad + \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)]} \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=1)] \leqslant 0 \\
\iff &\frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=1)]} - \frac{\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(\hat{Y}=1, Y=1)]}{\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=1)]} - \alpha \leqslant 0
\end{aligned}
$$

■

By similar proof, we also show in Lemma 3.11.2 that the constraint using $\tilde{g}_j^{\text{FPR}}(\theta)$ in Equation (3.12) can also be written as an equivalent constraint in the form of Equation (3.3), as

$$\tilde{g}_j^{\text{FPR}}(\theta) = \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[h^{\text{FPR}}(\theta, X, Y)]$$

for some function $h^{\text{FPR}} : \Theta \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

Lemma 3.11.2. *Denote $\hat{Y}$ as $\mathbb{1}(\phi(X;\theta) > 0)$. Let $h^{FPR}(\theta, X, Y)$ be given by*

$$h^{FPR}(\theta, X, Y) = \frac{1}{2}\left(\mathbb{1}(\hat{Y}=1, Y=0) - \mathbb{1}(Y=0)\left(\alpha + \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)]}\right)\right).$$

Then

$$
\begin{aligned}
\frac{\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=0)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=0)]} - \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)\,\mathbb{1}(\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)]} - \alpha \leqslant 0 \\
\iff \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[h^{FPR}(\theta, X, Y)] \leqslant 0.
\end{aligned}
$$

Proof. Substituting the given function $h^{\text{FPR}}(\theta, X, Y)$, and using the fact that $\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=0)] \geqslant 0$:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[h^{\text{FPR}}(\theta, X, Y)] \leqslant 0 \\
\iff &\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}\left[\frac{1}{2}\left(\mathbb{1}(\hat{Y}=1, Y=0) - \mathbb{1}(Y=0)\left(\alpha + \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)]}\right)\right)\right] \leqslant 0 \\
\iff &\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(\hat{Y}=1, Y=0)] - \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}\left[\mathbb{1}(Y=0)\left(\alpha + \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)]}\right)\right] \leqslant 0 \\
\iff &\mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(\hat{Y}=1, Y=0)] - \alpha \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=0)] \\
&\quad - \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)]} \mathop{\mathbb{E}}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=0)] \leqslant 0 \\
\iff &\frac{\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(\hat{Y}=1, Y=0)]}{\mathbb{E}_{X,Y\sim\tilde{p}_j}[\mathbb{1}(Y=0)]} - \frac{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0,\hat{Y}=1)]}{\mathbb{E}_{X,Y\sim p}[\mathbb{1}(Y=0)]} - \alpha \leqslant 0
\end{aligned}
$$

■

DRO when $\hat{G}$ and G have different dimensionalities

The soft assignments approach is naturally formulated to be able to handle $G \in \mathcal{G} = \{1, ..., m\}$ and $\hat{G} \in \hat{\mathcal{G}} = \{1, ..., \hat{m}\}$ when $\hat{m} \neq m$. The DRO approach can be extended to handle this case by generalizing Lemma 3.4.2 to $TV(p_j, \hat{p}_i) \leqslant P(\hat{G} \neq i | G = j), j \in \mathcal{G}, i \in \hat{\mathcal{G}}$, and generalizing the DRO formulation to have the true group distribution p_j bounded in a TV distance ball centered at $\hat{p}_i$. Empirically comparing this generalized DRO approach to the soft group assignments approach when $\hat{m} \neq m$ is an interesting avenue of future work.

3.12 Appendix: Further details for soft group assignments approach

Here we provide additional technical details regarding the soft group assignments approach introduced in Section 3.7.

Derivation for $\mathbb{E}[h(\theta, X, Y)|G=j]$

Here we show $\mathbb{E}[h(\theta, X, Y)|G=j] = \frac{\mathbb{E}[h(\theta,X,Y)P(G=j|\hat{Y},Y,\hat{G})]}{P(G=j)}$, assuming that $h(\theta, X, Y)$ depends on X through $\hat{Y}$, i.e. $\hat{Y} = \mathbb{1}(\phi(\theta, X) > 0)$. Using the tower property and the definition of

conditional expectation:

$$\begin{aligned}\mathbb{E}[h(\theta, X, Y)|G = j] &= \frac{\mathbb{E}[h(\theta, X, Y)\,\mathbb{1}(G = j)]}{P(G = j)} \\ &= \frac{\mathbb{E}[\mathbb{E}[h(\theta, X, Y)\,\mathbb{1}(G = j)|\hat{Y}, Y, \hat{G}]]}{P(G = j)} \\ &= \frac{\mathbb{E}[h(\theta, X, Y)\,\mathbb{E}[\mathbb{1}(G = j)|\hat{Y}, Y, \hat{G}]]}{P(G = j)} \\ &= \frac{\mathbb{E}[h(\theta, X, Y)P(G = j|\hat{Y}, Y, \hat{G})]}{P(G = j)}\end{aligned} \tag{3.13}$$

Equalizing TPRs and FPRs using soft group assignments

In the two case studies in Section 3.7, we enforce *equality of opportunity* and *equalized odds* [110] by equalizing true positive rates (TPRs) and/or false positive rates (FPRs) within some slack α. In this section, we describe in detail the implementation of the constraints for equalizing TPRs and FPRs under the soft group assignments approach.

To equalize TPRs with slack α under the soft group assignments approach, we set

$$g_j^{\text{TPR}}(\theta, w) = \frac{\mathbb{E}[\mathbb{1}(Y = 1)\,\mathbb{1}(\hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 1)]} - \frac{\mathbb{E}[\mathbb{1}(Y = 1)\,\mathbb{1}(\hat{Y} = 1)w(j|\hat{Y}, Y, \hat{G})]}{\mathbb{E}[\mathbb{1}(Y = 1)w(j|\hat{Y}, Y, \hat{G})]} - \alpha. \tag{3.14}$$

The first term corresponds to the TPR for the full population. The second term estimates the TPR for group j as done by Kallus et al. [128] in Equation (5) and Proposition 8. Setting $\alpha = 0$ exactly equalizes true positive rates.

To equalize FPRs with slack α under the soft group assignments approach, we set

$$g_j^{\text{FPR}}(\theta, w) = \frac{\mathbb{E}[\mathbb{1}(Y = 0)\,\mathbb{1}(\hat{Y} = 1)w(j|\hat{Y}, Y, \hat{G})]}{\mathbb{E}[\mathbb{1}(Y = 0)w(j|\hat{Y}, Y, \hat{G})]} - \frac{\mathbb{E}[\mathbb{1}(Y = 0)\,\mathbb{1}(\hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 0)]} - \alpha. \tag{3.15}$$

The first term estimates the FPR for group j as done previously for the TPR. The second term corresponds to the FPR for the full population. Setting $\alpha = 0$ exactly equalizes false positive rates.

To equalize TPRs for Case Study 1, we apply m constraints, $\left\{\max_{w \in \mathcal{W}(\theta)} g_j^{\text{TPR}}(\theta, w) \leqslant 0\right\}\ \forall j \in \mathcal{G}$. To equalize both TPRs and FPRs simultaneously for Case Study 2, we apply $2m$ constraints, $\left\{\max_{w \in \mathcal{W}(\theta)} g_j^{\text{TPR}}(\theta, w) \leqslant 0, \max_{w \in \mathcal{W}(\theta)} g_j^{\text{FPR}}(\theta, w) \leqslant 0\right\}\ \forall j \in \mathcal{G}$.

$h(\theta, X, Y)$ for equalizing TPRs and FPRs

Since the notation in Section 3.6 and in the rest of the paper uses generic functions h to express the group-specific constraints, we show in Lemma 3.12.1 that the constraint using

$g_j^{\text{TPR}}(\theta, w)$ in Equation (3.14) can also be written as an equivalent constraint in the form of Equation (3.6), as

$$g_j^{\text{TPR}}(\theta, w) = \frac{\mathbb{E}[h^{\text{TPR}}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})]}{P(G = j)}$$

for some function $h^{\text{TPR}} : \Theta \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

Lemma 3.12.1. *Denote $\hat{Y}$ as $\mathbb{1}(\phi(X;\theta) > 0)$. Let $h^{TPR}(\theta, X, Y)$ be given by*

$$h^{TPR}(\theta, X, Y) = \frac{1}{2}\left(-\mathbb{1}(\hat{Y} = 1, Y = 1) - \mathbb{1}(Y = 1)\left(\alpha - \frac{\mathbb{E}[\mathbb{1}(Y = 1, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 1)]}\right)\right).$$

Then

$$\begin{aligned}\frac{\mathbb{E}[\mathbb{1}(Y = 1)\,\mathbb{1}(\hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 1)]} - \frac{\mathbb{E}[\mathbb{1}(Y = 1)\,\mathbb{1}(\hat{Y} = 1) w(j|\hat{Y}, Y, \hat{G})]}{\mathbb{E}[\mathbb{1}(Y = 1) w(j|\hat{Y}, Y, \hat{G})]} - \alpha \leqslant 0 \\ \iff \frac{\mathbb{E}[h^{TPR}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})]}{P(G = j)} \leqslant 0.\end{aligned}$$

for all $j \in \mathcal{G}, P(G = j) > 0$.

Proof. Substituting the given function $h^{\text{TPR}}(\theta, X, Y)$, and using the fact that $P(G = j) > 0$ and $\mathbb{E}[\mathbb{1}(Y = 1) w(j|\hat{Y}, Y, \hat{G})] \geqslant 0$:

$$\begin{aligned}&\frac{\mathbb{E}[h^{\text{TPR}}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})]}{P(G = j)} \leqslant 0 \\ &\iff \mathbb{E}[h^{\text{TPR}}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})] \leqslant 0 \\ &\iff \mathbb{E}\left[\frac{1}{2}\left(-\mathbb{1}(\hat{Y} = 1, Y = 1) - \mathbb{1}(Y = 1)\left(\alpha - \frac{\mathbb{E}[\mathbb{1}(Y = 1, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 1)]}\right)\right) w(j|\hat{Y}, Y, \hat{G})\right] \leqslant 0 \\ &\iff -\mathbb{E}[\mathbb{1}(\hat{Y} = 1, Y = 1) w(j|\hat{Y}, Y, \hat{G})] \\ &\qquad - \mathbb{E}\left[\mathbb{1}(Y = 1)\left(\alpha - \frac{\mathbb{E}[\mathbb{1}(Y = 1, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 1)]}\right) w(j|\hat{Y}, Y, \hat{G})\right] \leqslant 0 \\ &\iff -\mathbb{E}[\mathbb{1}(\hat{Y} = 1, Y = 1) w(j|\hat{Y}, Y, \hat{G})] - \alpha\,\mathbb{E}[\mathbb{1}(Y = 1) w(j|\hat{Y}, Y, \hat{G})] \\ &\qquad + \frac{\mathbb{E}[\mathbb{1}(Y = 1, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 1)]}\,\mathbb{E}[\mathbb{1}(Y = 1) w(j|\hat{Y}, Y, \hat{G})] \leqslant 0 \\ &\iff \frac{\mathbb{E}[\mathbb{1}(Y = 1, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 1)]} - \frac{\mathbb{E}[\mathbb{1}(\hat{Y} = 1, Y = 1) w(j|\hat{Y}, Y, \hat{G})]}{\mathbb{E}[\mathbb{1}(Y = 1) w(j|\hat{Y}, Y, \hat{G})]} - \alpha \leqslant 0\end{aligned}$$

■

By similar proof, we also show in Lemma 3.12.2 that the constraint using $g_j^{\text{FPR}}(\theta, w)$ in Equation (3.15) can also be written as an equivalent constraint in the form of Equation (3.6), as

$$g_j^{\text{FPR}}(\theta, w) = \frac{\mathbb{E}[h^{\text{FPR}}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})]}{P(G = j)}$$

for some function $h^{\text{FPR}} : \Theta \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

Lemma 3.12.2. *Denote $\hat{Y}$ as $\mathbb{1}(\phi(X;\theta) > 0)$. Let $h^{FPR}(\theta, X, Y)$ be given by*

$$h^{FPR}(\theta, X, Y) = \frac{1}{2}\left(\mathbb{1}(\hat{Y} = 1, Y = 0) - \mathbb{1}(Y = 0)\left(\alpha + \frac{\mathbb{E}[\mathbb{1}(Y = 0, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 0)]}\right)\right).$$

Then

$$\begin{aligned}\frac{\mathbb{E}[\mathbb{1}(Y = 0)\,\mathbb{1}(\hat{Y} = 1) w(j|\hat{Y}, Y, \hat{G})]}{\mathbb{E}[\mathbb{1}(Y = 0) w(j|\hat{Y}, Y, \hat{G})]} - \frac{\mathbb{E}[\mathbb{1}(Y = 0)\,\mathbb{1}(\hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 0)]} - \alpha \leqslant 0 \\ \iff \frac{\mathbb{E}[h^{FPR}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})]}{P(G = j)} \leqslant 0.\end{aligned}$$

for all $j \in \mathcal{G}, P(G = j) > 0$.

Proof. Substituting the given function $h^{\text{FPR}}(\theta, X, Y)$, and using the fact that $P(G = j) > 0$ and $\mathbb{E}[\mathbb{1}(Y = 0) w(j|\hat{Y}, Y, \hat{G})] \geqslant 0$:

$$\begin{aligned}&\frac{\mathbb{E}[h^{\text{FPR}}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})]}{P(G = j)} \leqslant 0 \\ &\iff \mathbb{E}[h^{\text{FPR}}(\theta, X, Y) w(j|\hat{Y}, Y, \hat{G})] \leqslant 0 \\ &\iff \mathbb{E}\left[\frac{1}{2}\left(\mathbb{1}(\hat{Y} = 1, Y = 0) - \mathbb{1}(Y = 0)\left(\alpha + \frac{\mathbb{E}[\mathbb{1}(Y = 0, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 0)]}\right)\right) w(j|\hat{Y}, Y, \hat{G})\right] \leqslant 0 \\ &\iff \mathbb{E}[\mathbb{1}(\hat{Y} = 1, Y = 0) w(j|\hat{Y}, Y, \hat{G})] \\ &\qquad - \mathbb{E}\left[\mathbb{1}(Y = 0)\left(\alpha + \frac{\mathbb{E}[\mathbb{1}(Y = 0, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 0)]}\right) w(j|\hat{Y}, Y, \hat{G})\right] \leqslant 0 \\ &\iff \mathbb{E}[\mathbb{1}(\hat{Y} = 1, Y = 0) w(j|\hat{Y}, Y, \hat{G})] - \alpha\,\mathbb{E}[\mathbb{1}(Y = 0) w(j|\hat{Y}, Y, \hat{G})] \\ &\qquad - \frac{\mathbb{E}[\mathbb{1}(Y = 0, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 0)]}\,\mathbb{E}[\mathbb{1}(Y = 0) w(j|\hat{Y}, Y, \hat{G})] \leqslant 0 \\ &\iff \frac{\mathbb{E}[\mathbb{1}(\hat{Y} = 1, Y = 0) w(j|\hat{Y}, Y, \hat{G})]}{\mathbb{E}[\mathbb{1}(Y = 0) w(j|\hat{Y}, Y, \hat{G})]} - \frac{\mathbb{E}[\mathbb{1}(Y = 0, \hat{Y} = 1)]}{\mathbb{E}[\mathbb{1}(Y = 0)]} - \alpha \leqslant 0\end{aligned}$$

■

3.13 Appendix: Optimality and feasibility for the *Ideal* algorithm

Optimality and feasibility guarantees

We provide optimality and feasibility guarantees for Algorithm 3 and optimality guarantees for Algorithm 5.

Theorem 3.13.1 (**Optimality and Feasibility for Algorithm 3**). *Let $\theta^* \in \Theta$ be such that it satisfies the constraints $\max_{w \in \mathcal{W}(\theta)} g_j(\theta^*, w) \leqslant 0,\ \forall j \in \mathcal{G}$ and $f_0(\theta^*) \leqslant f(\theta)$ for every $\theta \in \Theta$ that satisfies the same constraints. Let $0 \leqslant f_0(\theta) \leqslant B, \forall \theta \in \Theta$. Let the space of Lagrange multipliers be defined as $\Lambda = \{\lambda \in \mathbb{R}^m_+ \,|\, \|\lambda\|_1 \leqslant R\}$, for $R > 0$. Let $B_\lambda \geqslant \max_t \|\nabla_\lambda \mathcal{L}(\theta^{(t)}, \lambda^{(t)})\|_2$. Let $\bar{\theta}$ be the stochastic classifier returned by Algorithm 3 when run for T iterations, with the radius of the Lagrange multipliers $R = T^{1/4}$ and learning rate $\eta_\lambda = \frac{R}{B_\lambda \sqrt{T}}$ Then:*

$$\mathbf{E}_{\theta \sim \bar{\theta}} [f(\theta)] \leqslant f(\theta^*) + \mathcal{O}\left(\frac{1}{T^{1/4}}\right) + \rho$$

and

$$\mathbf{E}_{\theta \sim \bar{\theta}} \left[\max_{w \in \mathcal{W}(\theta)} g_j(\theta, w)\right] \leqslant \mathcal{O}\left(\frac{1}{T^{1/4}}\right) + \rho'$$

Thus for any given $\epsilon > 0$, by solving Steps 2 and 4 of Algorithm 3 to sufficiently small errors ρ, ρ', and by running the algorithm for a sufficiently large number of steps T, we can guarantee that the returned stochastic model is ϵ-optimal and ϵ-feasible.

Proof. Let $\bar{\lambda} = \frac{1}{T}\sum_{t=1}^T \lambda^{(t)}$. We will interpret the minimax problem in (3.8) as a zero-sum between the θ-player who optimizes $\mathcal{L}$ over θ, and the λ-player who optimizes $\mathcal{L}$ over λ. We first bound the average regret incurred by the players over T steps. The best response computation in Step 2 of Algorithm 3 gives us:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T \mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[\mathcal{L}(\theta, \lambda^{(t)})\right] &\leqslant \frac{1}{T}\sum_{t=1}^T \min_{\theta \in \Theta} \mathcal{L}(\theta, \lambda^{(t)}) + \epsilon \\
&\leqslant \min_{\theta \in \Theta} \frac{1}{T}\sum_{t=1}^T \mathcal{L}(\theta, \lambda^{(t)}) + \rho \\
&= \min_{\theta \in \Theta} \mathcal{L}(\theta, \bar{\lambda}) + \rho \\
&\leqslant \min_{\theta \in \Theta} \max_{\lambda \in \Lambda} \mathcal{L}(\theta, \lambda) + \rho \\
&\leqslant f(\theta^*) + \rho. \qquad (3.16)
\end{aligned}$$

We then apply standard gradient ascent analysis for the projected gradient updates to λ in Step 4 of the algorithm, and get:

$$\max_{\lambda \in \Lambda} \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{m} \lambda_j \delta_j^{(t)} \geqslant \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{m} \lambda_j^{(t)} \delta_j^{(t)} - \mathcal{O}\left(\frac{R}{\sqrt{T}}\right).$$

We then plug the upper and lower bounds for the gradient estimates $\delta_j^{(t)}$'s from Step 3 of the Algorithm 3 into the above inequality:

$$\begin{aligned}
&\max_{\lambda \in \Lambda} \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{m} \lambda_j \left(\mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[\max_{w \in \mathcal{W}(\theta)} g_j(\theta, w) \right] + \rho' \right) \\
&\geqslant \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{m} \lambda_j^{(t)} \mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[\max_{w \in \mathcal{W}(\theta)} g_j(\theta, w) \right] - \mathcal{O}\left(\frac{R}{\sqrt{T}}\right).
\end{aligned}$$

which further gives us:

$$\begin{aligned}
&\max_{\lambda \in \Lambda} \left\{ \sum_{j=1}^{m} \lambda_j \mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[\max_{w \in \mathcal{W}(\theta)} g_j(\theta, w) \right] + \|\lambda\|_1 \rho' \right\} \\
&\geqslant \sum_{j=1}^{m} \lambda_j^{(t)} \mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[\max_{w \in \mathcal{W}(\theta)} g_j(\theta, w) \right] - \mathcal{O}\left(\frac{R}{\sqrt{T}}\right).
\end{aligned}$$

Adding $\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\theta \sim \hat{\theta}^{(t)}} [f(\theta)]$ to both sides of the above inequality, we finally get:

$$\frac{1}{T} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[\mathcal{L}(\theta, \lambda^{(t)}) \right] \geqslant \max_{\lambda \in \Lambda} \left\{ \frac{1}{T} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[\mathcal{L}(\theta, \lambda) \right] + \|\lambda\|_1 \rho' \right\} - \mathcal{O}\left(\frac{R}{\sqrt{T}}\right). \tag{3.17}$$

Optimality. Now, substituting $\lambda = \mathbf{0}$ in (3.17) and combining with (3.16) completes the proof of the optimality guarantee:

$$\mathop{\mathbb{E}}_{\theta \sim \bar{\theta}} [f(\theta)] \leqslant f_0(\theta^*) + \mathcal{O}\left(\frac{R}{\sqrt{T}}\right) + \rho$$

Feasibility. To show feasibility, we fix a constraint index $j \in \mathcal{G}$. Now substituting $\lambda_j = R$ and $\lambda_{j'} = 0, \forall j' \neq j$ in (3.17) and combining with (3.16) gives us:

$$\frac{1}{T} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\theta \sim \hat{\theta}^{(t)}} \left[f(\theta) + R \max_{w \in \mathcal{W}(\theta)} g_j(\theta, w) \right] \leqslant f(\theta^*) + \mathcal{O}\left(\frac{R}{\sqrt{T}}\right) + \rho + R\rho'.$$

which can be re-written as:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\theta \sim \bar{\theta}} \left[\max_{w \in \mathcal{W}(\theta)} g_j(\theta, w) \right] &\leqslant \frac{f(\theta^*) - \mathbb{E}_{\theta \sim \bar{\theta}} [f(\theta)]}{R} + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right) + \frac{\rho}{R} + \rho'. \\
&\leqslant \frac{B}{R} + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right) + \frac{\rho}{R} + \rho',
\end{aligned}$$

which is our feasibility guarantee. Setting $R = \mathcal{O}(T^{1/4})$ then completes the proof. ■

Algorithm 5 Best response on θ of Algorithm 3

Require: λ', learning rate $\eta_{\mathbf{w}} > 0$, estimates of $P(G = j|\hat{G} = k)$ to specify constraints $r_{g,\hat{g}}$'s, κ

1: **for** $q = 1, \ldots, Q$ **do**

2: *Best response on (θ, $\boldsymbol{\mu}$)*: use an oracle to find find $\theta^{(q)} \in \Theta$ and $\boldsymbol{\mu}^{(q)} \in \mathcal{M}^m$ such that:

$$\ell(\theta^{(q)}, \boldsymbol{\mu}^{(q)}, \mathbf{w}^{(q)}; \lambda') \leqslant \min_{\theta \in \Theta, \boldsymbol{\mu} \in \mathcal{M}^m} \ell(\theta, \boldsymbol{\mu}, \mathbf{w}^{(q)}; \lambda') + \kappa,$$

for a small slack $\kappa > 0$.

3: *Ascent step on* $\mathbf{w}$:

$$w_j^{(q+1)} \leftarrow \Pi_{\mathcal{W}_\Delta} \left(w_j^{(q)} + \eta_{\mathbf{w}} \nabla_{w_j} \ell(\theta^{(q)}, \boldsymbol{\mu}^{(q)}, \mathbf{w}^{(q)}; \lambda') \right),$$

where $\nabla_{w_j} \ell(\cdot)$ is a sub-gradient of ℓ w.r.t. w_j.

4: **end for**

5: **return** A uniform distribution $\hat{\theta}$ over $\theta^{(1)}, \ldots, \theta^{(Q)}$

Best Response over θ

We next describe our procedure for computing a best response over θ in Step 2 of Algorithm 3. We will consider a slightly relaxed version of the best response problem where the equality constraints in $\mathcal{W}(\theta)$ are replaced with closely-approximating inequality constraints.

Recall that the constraint set $\mathcal{W}(\theta)$ contains two sets of constraints (3.5), the total probability constraints that depend on θ, and the simplex constraints that do not depend on θ. So to decouple these constraint sets from θ, we introduce Lagrange multipliers μ for the total probability constraints to make them a part of the objective, and obtain a nested *minimax* problem over θ, μ, and w, where w is constrained to satisfy the simplex constraints alone. We then jointly minimize the inner Lagrangian over θ and μ, and perform gradient ascent updates on w with projections onto the simplex constraints. The joint-minimization over θ and μ is not necessarily convex and is solved using a minimization oracle.

We begin by writing out the best-response problem over θ for a fixed λ':

$$\min_{\theta \in \Theta} \mathcal{L}(\theta, \lambda') \;=\; \min_{\theta \in \Theta} f(\theta) + \sum_{j=1}^{m} \lambda'_j \max_{w_j \in \mathcal{W}(\theta)} g_j(\theta, w_j), \tag{3.18}$$

where we use w_j to denote the maximizer over $\mathcal{W}(\theta)$ for constraint g_j explicitly. We separate out the the simplex constraints in $\mathcal{W}(\theta)$ (3.5) and denote them by:

$$\mathcal{W}_\Delta = \left\{ w \in \mathbb{R}_+^{\mathcal{G} \times \{0,1\}^2 \times \hat{\mathcal{G}}} \;\middle|\; \sum_{j=1}^{m} w(j \mid \hat{y}, y, k) = 1, \;\forall k \in \hat{\mathcal{G}}, y, \hat{y} \in \{0,1\} \right\},$$

where we represent each w as a vector of values $w(i|\hat{y}, y, k)$ for each $j \in \mathcal{G}, \hat{y} \in \{0,1\}, y \in \{0,1\}$, and $k \in \hat{\mathcal{G}}$. We then relax the total probability constraints in $\mathcal{W}(\theta)$ into a set of

inequality constraints:

$$P(G=j|\hat{G}=k) \;-\; \sum_{\hat{y},y\in\{0,1\}} w(j \mid \hat{y},y,k) P(\hat{Y}(\theta)=\hat{y}, Y=y|\hat{G}=k) \;-\; \tau \;\leqslant\; 0$$

$$\sum_{\hat{y},y\in\{0,1\}} w(j \mid \hat{y},y,k) P(\hat{Y}(\theta)=\hat{y}, Y=y|\hat{G}=k) \;-\; P(G=j|\hat{G}=k) \;-\; \tau \;\leqslant\; 0$$

for some small $\tau > 0$. We have a total of $U = 2 \times m \times \hat{m}$ relaxed inequality constraints, and will denote each of them as $r_u(\theta, w) \leqslant 0$, with index u running from 1 to U. Note that each $r_u(\theta, w)$ is linear in w.

Introducing Lagrange multipliers μ for the relaxed total probability constraints, the optimization problem in (3.18) can be re-written equivalently as:

$$\min_{\theta\in\Theta} \; f(\theta) \;+\; \sum_{j=1}^{m} \lambda'_j \max_{w_j\in\mathcal{W}_\Delta} \min_{\mu_j\in\mathcal{M}} \left\{ g_j(\theta, w_j) - \sum_{u=1}^{U} \mu_{j,u}\, r_u(\theta, w_j) \right\},$$

where note that each w_j is maximized over only the simplex constraints $\mathcal{W}_\Delta$ which are independent of θ, and $\mathcal{M} = \{\mu_j \in \mathbb{R}_+^{m\times\hat{m}} \,|\, \|\mu_j\|_1 \leqslant R'\}$, for some constant $R' > 0$. Because each w_j and μ_j appears only in the j-th term in the summation, we can pull out the max and min, and equivalently rewrite the above problem as:

$$\min_{\theta\in\Theta} \max_{\mathbf{w}\in\mathcal{W}_\Delta^m} \min_{\boldsymbol{\mu}\in\mathcal{M}^m} \underbrace{f(\theta) + \sum_{j=1}^{m} \lambda'_j \Bigg(\underbrace{g_j(\theta, w_j) - \sum_{u=1}^{U} \mu_{j,u}\, r_u(\theta, w_j)}_{\omega(\theta,\mu_j,w_j)} \Bigg)}_{\ell(\theta,\boldsymbol{\mu},\mathbf{w};\lambda')}, \tag{3.19}$$

where $\mathbf{w} = (w_1, \ldots, w_m)$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$. We then solve this nested minimax problem in Algorithm 5 by using an minimization *oracle* to perform a full optimization of ℓ over (θ, μ), and carrying out gradient ascent updates on ℓ over w_j.

We now proceed to show an optimality guarantee for Algorithm 5.

Theorem 3.13.2 (**Optimality Guarantee for Algorithm 5**). *Suppose for every $\theta \in \Theta$, there exists a $\widetilde{w}_j \in \mathcal{W}_\Delta$ such that $r_u(\theta, \widetilde{w}_j) \leqslant -\gamma$, $\forall u \in [U]$, for some $\gamma > 0$. Let $0 \leqslant g_j(\theta, w_j) \leqslant B'$, $\forall \theta \in \Theta, w_j \in \mathcal{W}_\Delta$. Let $B_{\mathbf{w}} \geqslant \max_q \|\nabla_{\mathbf{w}}\, \ell(\theta^{(q)}, \boldsymbol{\mu}^{(q)}, \mathbf{w}^{(q)}; \lambda'))\|_2$. Let $\hat{\theta}$ be the stochastic classifier returned by Algorithm 5 when run for a given λ' for Q iterations, with the radius of the Lagrange multipliers $R' = B'/\gamma$ and learning rate $\eta_{\mathbf{w}} = \frac{R'}{B_{\mathbf{w}}\sqrt{T}}$. Then:*

$$\mathop{\mathbb{E}}_{\theta\sim\hat{\theta}} [\mathcal{L}(\theta, \lambda')] \;\leqslant\; \min_{\theta\in\Theta} \mathcal{L}(\theta, \lambda') \;+\; \mathcal{O}\left(\frac{1}{\sqrt{Q}}\right) + \kappa.$$

Before proving Theorem 3.13.2, we will find it useful to state the following lemma.

Lemma 3.13.3 (**Boundedness of Inner Lagrange Multipliers in** (3.19))**.** *Suppose for every $\theta \in \Theta$, there exists a $\widetilde{w}_j \in \mathcal{W}$ such that $r_u(\theta, \widetilde{w}_j) \leqslant -\gamma$, $\forall u \in [U]$, for some $\gamma > 0$. Let $0 \leqslant g_j(\theta, w_j) \leqslant B'$, $\forall \theta \in \Theta, w_j \in \mathcal{W}_\Delta$. Let $\mathcal{M} = \{\mu_j \in \mathbb{R}_+^K \mid \|\mu_j\|_1 \leqslant R'\}$ with the radius of the Lagrange multipliers $R' = B'/\gamma$. Then we have for all $j \in \mathcal{G}$:*

$$\max_{w_j \in \mathcal{W}_\Delta} \min_{\mu_j \in \mathcal{M}} \omega\big(\theta, \mu_j, w_j\big) = \max_{w_j \in \mathcal{W}_\Delta : r_u(\theta, w_j) \leqslant 0, \, \forall u} g_j(\theta, w_j).$$

Proof. For a given $j \in \mathcal{G}$, let $w_j^* \in \operatorname*{argmax}_{w_j \in \mathcal{W}_\Delta : r_u(\theta, w_j) \leqslant 0, \, \forall u} g_j(\theta, w_j)$. Then:

$$g_j(\theta, w_j^*) \;=\; \max_{w_j \in \mathcal{W}_\Delta} \min_{\mu_j \in \mathbb{R}_+^K} \omega\big(\theta, \mu_j, w_j\big), \tag{3.20}$$

where note that μ_j is minimized over all non-negative values. Since the ω is linear in both μ_j and w_j, we can interchange the min and max:

$$g_j(\theta, w_j^*) \;=\; \min_{\mu_j \in \mathbb{R}_+^K} \max_{w_j \in \mathcal{W}_\Delta} \omega\big(\theta, \mu_j, w_j\big).$$

We show below that the minimizer μ^* in the above problem is in fact bounded and present in $\mathcal{M}$.

$$\begin{aligned}
g_j(\theta, w_j^*) \;&=\; \max_{w_j \in \mathcal{W}} \omega\big(\theta, \mu_j^*, w_j\big) \\
&=\; \max_{w_j \in \mathcal{W}} \left\{ g_j(\theta, w_j) \,-\, \sum_{k=1}^{K} \mu_{j,k}^* \, r_k(\theta, w_j) \right\} \\
&\geqslant\; g_j(\theta, \widetilde{w}_j) \,-\, \|\mu_j^*\|_1 \max_{k \in [K]} r_k(\theta, \widetilde{w}_j) \\
&\geqslant\; g_j(\theta, w_j) \,+\, \|\mu_j^*\|_1 \gamma \;\geqslant\; \|\mu_j^*\|_1 \gamma.
\end{aligned}$$

We further have:

$$\|\mu_j^*\|_1 \;\leqslant\; g_j(\theta, w_j)/\gamma \;\leqslant\; B'/\gamma. \tag{3.21}$$

Thus the minimizer $\mu_j^* \in \mathcal{M}$. So the minimization in (3.20) can be performed over only $\mathcal{M}$, which completes the proof of the lemma. ■

Equipped with the above result, we are now ready to prove Theorem 3.13.2.

Proof of Theorem 3.13.2. Let $\bar{w}_j = \frac{1}{Q} \sum_{q=1}^{Q} w_j^{(q)}$. The best response on θ and μ gives us:

$$\begin{aligned}
&\frac{1}{Q} \sum_{q=1}^{Q} \Big(f(\theta^{(q)}) \,+\, \sum_{j=1}^{m} \lambda_j' \, \omega\big(\theta^{(q)}, \mu_j^{(q)}, w_j^{(q)}\big) \Big) \\
&\leqslant\; \frac{1}{Q} \sum_{q=1}^{Q} \min_{\theta \in \Theta, \, \boldsymbol{\mu} \in \mathcal{M}^m} \Big(f(\theta) \,+\, \sum_{j=1}^{m} \lambda_j' \, \omega\big(\theta, \mu_j, w_j^{(q)}\big) \Big) \,+\, \kappa
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{Q}\sum_{q=1}^{Q}\Big(\min_{\theta\in\Theta} f(\theta) + \sum_{j=1}^{m}\lambda'_j \min_{\mu_j\in\mathcal{M}} \omega\big(\theta,\mu_j,w_j^{(q)}\big)\Big) + \kappa \quad (j\text{-th summation term depends on } \mu_j \text{ alone})\\
&\leqslant \min_{\theta\in\Theta}\frac{1}{Q}\sum_{q=1}^{Q}\Big(f(\theta) + \sum_{j=1}^{m}\lambda'_j \min_{\mu_j\in\mathcal{M}} \omega\big(\theta,\mu_j,w_j^{(q)}\big)\Big) + \kappa\\
&\leqslant \min_{\theta\in\Theta}\Big\{f(\theta) + \sum_{j=1}^{m}\lambda'_j \min_{\mu_j\in\mathcal{M}} \frac{1}{Q}\sum_{q=1}^{Q}\omega\big(\theta,\mu_j,w_j^{(q)}\big)\Big\} + \kappa\\
&= \min_{\theta\in\Theta}\Big\{f(\theta) + \sum_{j=1}^{m}\lambda'_j \min_{\mu_j\in\mathcal{M}} \omega\big(\theta,\mu_j,\bar{w}_j\big)\Big\} + \kappa\\
&\leqslant \min_{\theta\in\Theta}\Big\{f(\theta) + \sum_{j=1}^{m}\lambda'_j \max_{w_j\in\mathcal{W}}\min_{\mu_j\in\mathcal{M}} \omega\big(\theta,\mu_j,w_j\big)\Big\} + \kappa \quad (\text{by linearity of } \omega \text{ in } w_j)\\
&= \min_{\theta\in\Theta}\Big\{f(\theta) + \sum_{j=1}^{m}\lambda'_j \max_{w_j:\, r_u(\theta,w_j)\leqslant 0,\,\forall u} g_j(\theta,w_j)\Big\} + \kappa \quad (\text{from Lemma 3.13.3})\\
&= \min_{\theta\in\Theta}\mathcal{L}(\theta,\lambda') + \kappa. \qquad (3.22)
\end{aligned}
$$

Applying standard gradient ascent analysis to the gradient ascent steps on $\mathbf{w}$ (using the fact that ω is linear in $\mathbf{w}$)

$$
\begin{aligned}
&\frac{1}{Q}\sum_{q=1}^{Q}\Big(f(\theta^{(q)}) + \sum_{j=1}^{m}\lambda'_j\,\omega\big(\theta^{(q)},\mu_j^{(q)},w_j^{(q)}\big)\Big)\\
&\geqslant \max_{\mathbf{w}\in\mathcal{W}_\Delta^m}\frac{1}{Q}\sum_{q=1}^{Q}\Big(f(\theta^{(q)}) + \sum_{j=1}^{m}\lambda'_j\,\omega\big(\theta^{(q)},\mu_j^{(q)},w_j\big)\Big) - \mathcal{O}\left(\frac{1}{\sqrt{Q}}\right)\\
&= \frac{1}{Q}\sum_{q=1}^{Q}\Big(f(\theta^{(q)}) + \sum_{j=1}^{m}\lambda'_j \max_{w_j\in\mathcal{W}_\Delta}\omega\big(\theta^{(q)},\mu_j^{(q)},w_j\big)\Big) - \mathcal{O}\left(\frac{1}{\sqrt{Q}}\right) \quad (j\text{-th summation term depends on } w_j\\
&\geqslant \frac{1}{Q}\sum_{q=1}^{Q}\Big(f(\theta^{(q)}) + \sum_{j=1}^{m}\lambda'_j \max_{w_j\in\mathcal{W}_\Delta}\min_{\mu_j\in\mathcal{M}}\omega\big(\theta^{(q)},\mu_j,w_j\big)\Big) - \mathcal{O}\left(\frac{1}{\sqrt{Q}}\right) \quad (\text{by linearity of } \omega \text{ in } w_j \text{ and } \mu_j)\\
&= \mathop{\mathbb{E}}_{\theta\sim\hat{\theta}}\left[f(\theta) + \sum_{j=1}^{m}\lambda'_j \max_{w_j\in\mathcal{W}_\Delta}\min_{\mu_j\in\mathcal{M}}\omega\big(\theta,\mu_j,w_j\big)\right] - \mathcal{O}\left(\frac{1}{\sqrt{Q}}\right)\\
&= \mathop{\mathbb{E}}_{\theta\sim\hat{\theta}}\left[f(\theta^{(q)}) + \sum_{j=1}^{m}\lambda'_j \max_{w_j\in\mathcal{W}_\Delta:\, r_u(\theta,w_j)\leqslant 0,\,\forall u} g_j(\theta,w_j)\right] - \mathcal{O}\left(\frac{1}{\sqrt{Q}}\right) \quad (\text{from Lemma 3.13.3})\\
&= \mathop{\mathbb{E}}_{\theta\sim\hat{\theta}}\left[\mathcal{L}(\theta,\lambda')\right] - \mathcal{O}\left(\frac{1}{\sqrt{Q}}\right).
\end{aligned}
$$

Combining (3.22) and (3.23) completes the proof. ■

Algorithm 6 *Practical* Algorithm

Require: learning rates $\eta_\theta > 0$, $\eta_\lambda > 0$, estimates of $P(G = j|\hat{G} = k)$ to specify $\mathcal{W}(\theta)$

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

2: Solve for w given θ using linear programming or a gradient method: $w^{(t)} \leftarrow \max_{w \in \mathcal{W}(\theta^{(t)})} \sum_{j=1}^{m} \lambda_j^{(t)} g_j(\theta^{(t)}, w)$

3: *Descent step on θ:* $\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta_\theta \delta_\theta^{(t)}$, where $\delta_\theta^{(t)} = \nabla_\theta \left(f_0(\theta^{(t)}) + \sum_{j=1}^{m} \lambda_j^{(t)} g_j \left(\theta^{(t)}, w^{(t+1)} \right) \right)$

4: *Ascent step on λ:* $\tilde{\lambda}_j^{(t+1)} \leftarrow \lambda_j^{(t)} + \eta_\lambda g_j \left(\theta^{(t+1)}, w^{(t+1)} \right) \quad \forall j \in \mathcal{G}$, $\lambda^{(t+1)} \leftarrow \Pi_\Lambda(\tilde{\lambda}^{(t+1)})$,

5: **end for**

6: **return** $\theta^{(t^*)}$ where t^* denotes the *best* iterate that satisfies the constraints in (3.7) with the lowest objective.

3.14 Appendix: Discussions on the *Practical* algorithm

Here we provide the details of the *practical* Algorithm 6 to solve problem (3.8). We also further discuss how we arrive at Algorithm 6. Recall that in the minimax problem in (3.8), restated below, each of the m constraints contain a max over w:

$$\min_{\theta \in \Theta} \max_{\lambda \in \Lambda} f(\theta) + \sum_{j=1}^{m} \lambda_j \max_{w \in \mathcal{W}(\theta)} g_j(\theta, w).$$

We show below that this is equivalent to a minimax problem where the sum over j and max over w are swapped:

Lemma 3.14.1. *The minimax problem in* (3.8) *is equivalent to:*

$$\min_{\theta \in \Theta} \max_{\lambda \in \Lambda} \max_{w \in \mathcal{W}(\theta)} f(\theta) + \sum_{j=1}^{m} \lambda_j g_j(\theta, w). \tag{3.24}$$

Proof. Recall that the space of Lagrange multipliers $\Lambda = \{\lambda \in \mathbb{R}_+^m \,|\, \|\lambda\|_1 \leqslant R\}$, for $R > 0$. So the above maximization over Λ can be re-written in terms of a maximization over the

m-dimensional simplex Δ_m and a scalar $\beta \in [0, R]$:

$$\begin{aligned}
&\min_{\theta\in\Theta} \max_{\beta\in[0,R],\, \nu\in\Delta_m} f(\theta) + \beta \sum_{j=1}^{m} \nu_j \max_{w\in\mathcal{W}(\theta)} g_j(\theta, w) \\
&= \min_{\theta\in\Theta} \max_{\beta\in[0,R]} f(\theta) + \beta \max_{\nu\in\Delta_m} \sum_{j=1}^{m} \nu_j \max_{w\in\mathcal{W}(\theta)} g_j(\theta, w) \\
&= \min_{\theta\in\Theta} \max_{\beta\in[0,R]} f(\theta) + \beta \max_{j\in\mathcal{G}} \max_{w\in\mathcal{W}(\theta)} g_j(\theta, w) \\
&= \min_{\theta\in\Theta} \max_{\beta\in[0,R]} f(\theta) + \beta \max_{w\in\mathcal{W}(\theta)} \max_{j\in\mathcal{G}} g_j(\theta, w) \\
&= \min_{\theta\in\Theta} \max_{\beta\in[0,R]} f(\theta) + \beta \max_{w\in\mathcal{W}(\theta)} \max_{\nu\in\Delta_m} \sum_{j=1}^{m} \nu_j g_j(\theta, w) \\
&= \min_{\theta\in\Theta} f(\theta) + \max_{\beta\in[0,R],\, \nu\in\Delta_m} \max_{w\in\mathcal{W}(\theta)} \sum_{j=1}^{m} \beta\nu_j g_j(\theta, w) \\
&= \min_{\theta\in\Theta} f(\theta) + \max_{\lambda\in\Lambda} \max_{w\in\mathcal{W}(\theta)} \sum_{j=1}^{m} \lambda_j g_j(\theta, w),
\end{aligned}$$

which completes the proof. ■

The practical algorithm outlined in Algorithm 6 seeks to solve the re-written minimax problem in (3.24), and is similar in structure to the ideal algorithm in Algorithm 3, in that it has two high-level steps: an approximate best response over θ and gradient ascent updates on λ. However, the algorithm works with deterministic classifiers $\theta^{(t)}$, and uses a simple heuristic to approximate the best response step. Specifically, for the best response step, the algorithm finds the maximizer of the Lagrangian over w for a fixed $\theta^{(t)}$ by e.g. using linear programming:

$$w^{(t)} \leftarrow \max_{w\in\mathcal{W}(\theta^{(t)})} \sum_{j=1}^{m} \lambda_j^{(t)} g_j(\theta^{(t)}, w),$$

uses the maximizer $w^{(t)}$ to approximate the gradient of the Lagrangian at $\theta^{(t)}$:

$$\delta_\theta^{(t)} = \nabla_\theta \Big(f_0(\theta^{(t)}) + \sum_{j=1}^{m} \lambda_j^{(t)} f_j\left(\theta^{(t)}, w^{(t+1)}\right) \Big)$$

and performs a single gradient update on θ:

$$\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta_\theta \delta_\theta^{(t)}.$$

The gradient ascent step on λ is the same as the ideal algorithm, except that it is simpler to implement as the iterates $\theta^{(t)}$ are deterministic:

$$\tilde{\lambda}_j^{(t+1)} \leftarrow \lambda_j^{(t)} + \eta_\lambda f_j\Big(\theta^{(t+1)}, w^{(t+1)}\Big) \;\; \forall j \in \mathcal{G};$$

$$\lambda^{(t+1)} \leftarrow \Pi_\Lambda(\tilde{\lambda}^{(t+1)}).$$

3.15 Appendix: Additional experiment details and results

We provide more details on the experimental setup as well as further results.

Additional experimental setup details

This section contains further details on the experimental setup, including the datasets used and hyperparameters tuned. All categorical features in each dataset were binarized into one-hot vectors. All numerical features were bucketized into 4 quantiles, and further binarized into one-hot vectors.

For the naïve approach, we solve the constrained optimization problem (3.2) with respect to the noisy groups $\hat{G}$. For comparison, we also report the results of the unconstrained optimization problem and the constrained optimization problem (3.1) when the true groups G are known. For the DRO problem (3.3), we estimate the bound $\gamma_j = P(\hat{G} \neq G | G = j)$ in each case study. For the soft group assignments approach, we implement the *practical* algorithm (Algorithm 6).

In the experiments, we replace all expectations in the objective and constraints with finite-sample empirical versions. So that the constraints will be convex and differentiable, we replace all indicator functions with hinge upper bounds, as in Davenport et al. [61] and Eban et al. [78]. We use a linear model: $\phi(X;\theta) = \theta^T X$. The noisy protected groups $\hat{G}$ are included as a feature in the model, demonstrating that conditional independence between $\hat{G}$ and the model $\phi(X;\theta)$ is not required here, unlike some prior work [15]. Aside from being used to estimate the noise model $P(G = k|\hat{G} = j)$ for the soft group assignments approach[2], the true groups G are never used in the training or validation process.

Each dataset was split into train/validation/test sets with proportions 0.6/0.2/0.2. For each algorithm, we chose the *best* iterate $\theta^{(t^*)}$ out of T iterates on the train set, where we define *best* as the iterate that achieves the lowest objective value while satisfying all constraints. We select the hyperparameters that achieve the best performance on the validation set (details in Appendix 3.15). We repeat this procedure for ten random train/validation/test splits and record the mean and standard errors for all metrics[3].

Adult dataset

For the first case study, we used the Adult dataset from UCI [70], which includes 48,842 examples. The features used were *age*, *workclass*, *fnlwgt*, *education*, *education_num*, *marital_status*, *occupation*, *relationship*, *race*, *gender*, *capital_gain*, *capital_loss*, *hours_per_week*,

[2]If $P(G = k|\hat{G} = j)$ is estimated from an auxiliary dataset with a different distribution than test, this could lead to generalization issues for satisfying the true group constraints on test. In our experiments, we lump those generalization issues in with any distributional differences between train and test.

[3]When we report the "maximum" constraint violation, we use the mean and standard error of the constraint violation for the group j with the maximum mean constraint violation.

and *native_country*. Detailed descriptions of what these features represent are provided by UCI [70]. The label was whether or not *income_bracket* was above $50,000. The true protected groups were given by the *race* feature, and we combined all examples with race other than "white" or "black" into a group of race "other." When training with the noisy group labels, we did *not* include the true *race* as a feature in the model, but included the noisy race labels as a feature in the model instead. We set $\alpha = 0.05$ as the constraint slack.

The constraint violation that we report in Figure 3.1 is taken over a test dataset with n examples $(X_1, Y_1, G_1), ..., (X_n, Y_n, G_n)$, and is given by:

$$\max_{j \in \mathcal{G}} \quad \frac{\sum_{i=1}^n \mathbb{1}(\hat{Y}(\theta)_i = 1, Y_i = 1)}{\sum_{i=1}^n \mathbb{1}(Y_i = 1)} - \frac{\sum_{i=1}^n \mathbb{1}(\hat{Y}(\theta)_i = 1, Y_i = 1, G_i = j)}{\sum_{i=1}^n \mathbb{1}(Y_i = 1, G_i = j)} - \alpha,$$

where $\hat{Y}(\theta)_i = \mathbb{1}(\phi(\theta; X_i) > 0)$.

Section 3.12 shows how we specifically enforce equality of opportunity using the soft assignments approach, and Section 3.11 shows how we enforce equality of opportunity using DRO.

Credit dataset

For the second case study, we used default of credit card clients dataset from UCI [70] collected by a company in Taiwan [221], which contains 30000 examples and 24 features. The features used were *amount_of_the_given_credit*, *gender*, *education*, *education*, *marital_status*, *age*, *history_of_past_payment*, *amount_of_bill_statement*, *amount_of_previous_payment*. Detailed descriptions of what these features represent are provided by UCI [70]. The label was whether or not *default* was true. The true protected groups were given by the *education* feature, and we combined all examples with education level other than "graduate school" or "university" into a group of education level "high school and others". When training with the noisy group labels, we did *not* include the true *education* as a feature in the model, but included the noisy education level labels as a feature in the model instead. We set $\alpha = 0.03$ as the constraint slack.

The constraint violation that we report in Figure 3.1 is taken over a test dataset with n examples $(X_1, Y_1, G_1), ..., (X_n, Y_n, G_n)$, and is given by:

$$\max_{j \in \mathcal{G}} \quad \max(\Delta_j^{\text{TPR}}, \Delta_j^{\text{FPR}})$$

where

$$\Delta_j^{\text{TPR}} = \frac{\sum_{i=1}^n \mathbb{1}(\hat{Y}(\theta)_i = 1, Y_i = 1)}{\sum_{i=1}^n \mathbb{1}(Y_i = 1)} - \frac{\sum_{i=1}^n \mathbb{1}(\hat{Y}(\theta)_i = 1, Y_i = 1, G_i = j)}{\sum_{i=1}^n \mathbb{1}(Y_i = 1, G_i = j)} - \alpha$$

and

$$\Delta_j^{\text{FPR}} = \frac{\sum_{i=1}^n \mathbb{1}(\hat{Y}(\theta)_i = 1, Y_i = 0, G_i = j)}{\sum_{i=1}^n \mathbb{1}(Y_i = 0, G_i = j)} - \frac{\sum_{i=1}^n \mathbb{1}(\hat{Y}(\theta)_i = 1, Y_i = 0)}{\sum_{i=1}^n \mathbb{1}(Y_i = 0)} - \alpha$$

and $\hat{Y}(\theta)_i = \mathbb{1}(\phi(\theta; X_i) > 0)$.

Section 3.12 shows how we specifically enforce equalized odds using the soft assignments approach, and Section 3.11 shows how we enforce equalized odds using DRO.

Optimization code

For all case studies, we performed experiments comparing the naïve approach, the DRO approach (Section 3.5) and the soft group assignments approach (Section 3.6). We also compared these to the baselines of optimizing without constraints and optimizing with constraints with respect to the true groups. All optimization code was written in Python and TensorFlow [4]. All gradient steps were implemented using TensorFlow's Adam optimizer [5], though all experiments can also be reproduced using simple gradient descent without momentum. We computed full gradients over all datasets, but minibatching can also be used for very large datasets.

Table 3.1: Hyperparameters tuned for each approach

Hparam	Values tried	Relevant approaches	Description
η_θ	{0.001,0.01,0.1}	all approaches	learning rate for θ
η_λ	{0.25,0.5,1.0,2.0}	all except unconstrained	learning rate for λ
$\eta_{\tilde{p}_j}$	{0.001, 0.01, 0.1}	DRO	learning rate for $\tilde{p}_j$
η_w	{0.001, 0.01, 0.1}	soft assignments	learning rate using gradient methods for w

Hyperparameters

The hyperparameters for each approach were chosen to achieve the best performance on the validation set on average over 10 random train/validation/test splits, where "best" is defined as the set of hyperparameters that achieved the lowest error rate while satisfying all constraints relevant to the approach. The final hyperparameter values selected for each method were neither the largest nor smallest of all values tried. A list of all hyperparameters tuned and the values tried is given in Table 3.1.

For the naïve approach, the constraints used when selecting the hyperparameter values on the validation set were the constraints with respect to the noisy group labels given in Equation (3.2). For the DRO approach and the soft group assignments approach, the respective robust constraints were used when selecting hyperparameter values on the validation

[4] Abadi, M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. tensorflow.org.

[5] `https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/AdamOptimizer`

set. Specifically, for the DRO approach, the constraints used were those defined in Equation (3.3), and for the soft group assignments approach, the constraints used were those defined in Equation (3.7). For the unconstrained baseline, no constraints were taken into account when selecting the best hyperparameter values. For the baseline constrained with access to the true group labels, the true group constraints were used when selecting the best hyperparameter values.

Hinge relaxations of all constraints were used during training to achieve convexity. Since the hinge relaxation is an upper bound on the real constraints, the hinge-relaxed constraints may require some additional slack to maintain feasibility. This positive slack β was added to the original slack α when training with the hinge-relaxed constraints, and the amount of slack β was chosen so that the relevant hinge-relaxed constraints were satisfied on the training set.

All approaches ran for 750 iterations over the full dataset.

Additional experiment results

This section provides additional experiment results. All results reported here and in the main paper are on the test set (averaged over 10 random train/validation/test splits).

Case study 1 (Adult)

This section provides additional experiment results for case study 1 on the Adult dataset.

Figure 3.4 that the naïve approach, DRO approach, and soft assignments approaches all satisfied the fairness constraints for the noisy groups on the test set.

Figure 3.5 confirms that the DRO approach and the soft assignments approaches both managed to satisfy their respective robust constraints on the test set on average. For the DRO approach, the constraints measured in Figure 3.5 come from Equation (3.3), and for the soft assignments approach, the constraints measured in Figure 3.5 come from Equation (3.7). We provide the exact error rate values and maximum violations on the true groups for the Adult dataset in Table 3.2.

Case study 2 (Credit)

This section provides additional experiment results for case study 2 on the Credit dataset.

Figure 3.6 shows the constraint violations with respect to the true groups on test separated into TPR violations and FPR violations. For all noise levels, there were higher TPR violations than FPR violations. However, this does not mean that the FPR constraint was meaningless – the FPR constraint still ensured that the TPR constraints weren't satisfied by simply adding false positives.

Figure 3.7 confirms that the naïve approach, DRO approach, and soft assignments approaches all satisfied the fairness constraints for the noisy groups on the test set.

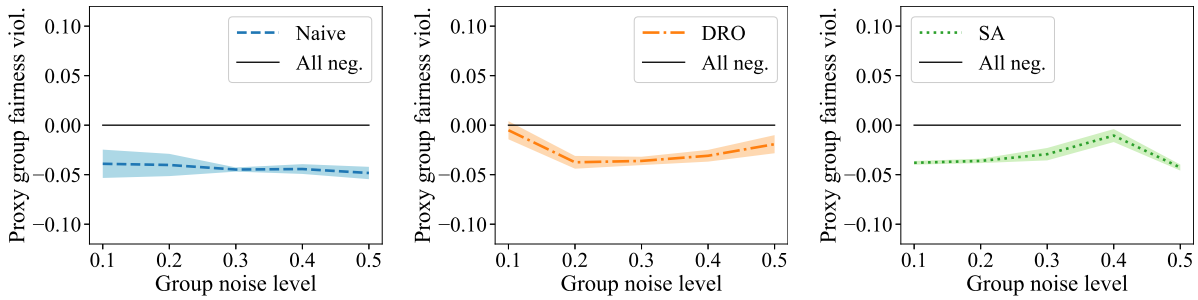


Figure 3.4: Maximum fairness constraint violations with respect to the noisy groups $\hat{G}$ on the test set for different group noise levels γ on the Adult dataset. For each noise level, we plot the mean and standard error over 10 random train/val/test splits. The black solid line illustrates a maximum constraint violation of 0. While the naïve approach (*left*) has increasingly higher fairness constraints with respect to the true groups as the noise increases, it always manages to satisfy the constraints with respect to the noisy groups $\hat{G}$

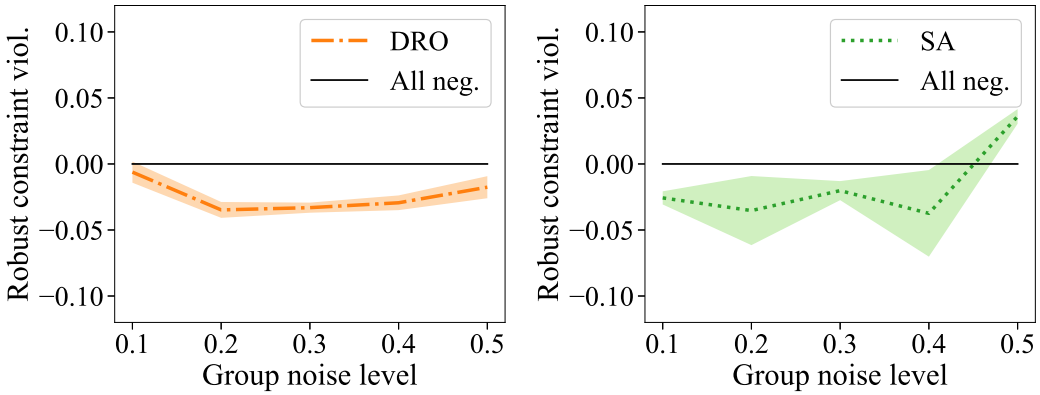


Figure 3.5: Maximum robust constraint violations on the test set for different group noise levels $P(\hat{G} \neq G)$ on the Adult dataset. For each noise level, we plot the mean and standard error over 10 random train/val/test splits. The black dotted line illustrates a maximum constraint violation of 0. Both the DRO approach (*left*) and the soft group assignments approach (*right*) managed to satisfy their respective robust constraints on the test set on average for all noise levels.

Figure 3.8 confirms that the DRO approach and the soft assignments approaches both managed to satisfy their respective robust constraints on the test set on average. For the DRO approach, the constraints measured in Figure 3.8 come from Equation (3.3), and for the soft assignments approach, the constraints measured in Figure 3.8 come from Equation

Table 3.2: Error rate and fairness constraint violations on the true groups for the Adult dataset (mean and standard error over 10 train/test/splits).

	DRO		Soft Assignments	
Noise	Error rate	Max G Viol.	Error rate	Max G Viol.
0.1	0.152 ± 0.001	0.002 ± 0.019	0.148 ± 0.001	-0.048 ± 0.002
0.2	0.200 ± 0.002	-0.045 ± 0.003	0.157 ± 0.003	-0.048 ± 0.002
0.3	0.216 ± 0.010	-0.044 ± 0.004	0.158 ± 0.005	0.002 ± 0.030
0.4	0.209 ± 0.006	-0.019 ± 0.031	0.188 ± 0.003	-0.016 ± 0.016
0.5	0.219 ± 0.012	-0.030 ± 0.032	0.218 ± 0.002	0.004 ± 0.006

(3.7).

We provide the exact error rate values and maximum violations on the true groups for the Credit dataset in Table 3.3.

Table 3.3: Error rate and fairness constraint violations on the true groups for the Credit dataset (mean and standard error over 10 train/test/splits).

	DRO		Soft Assignments	
Noise	Error rate	Max G Viol.	Error rate	Max G Viol.
0.1	0.206 ± 0.003	-0.006 ± 0.006	0.182 ± 0.002	0.000 ± 0.005
0.2	0.209 ± 0.002	-0.008 ± 0.008	0.182 ± 0.001	0.004 ± 0.005
0.3	0.212 ± 0.002	-0.006 ± 0.006	0.198 ± 0.001	-0.025 ± 0.007
0.4	0.210 ± 0.002	-0.017 ± 0.008	0.213 ± 0.001	-0.028 ± 0.005
0.5	0.211 ± 0.003	-0.015 ± 0.006	0.211 ± 0.001	-0.014 ± 0.004

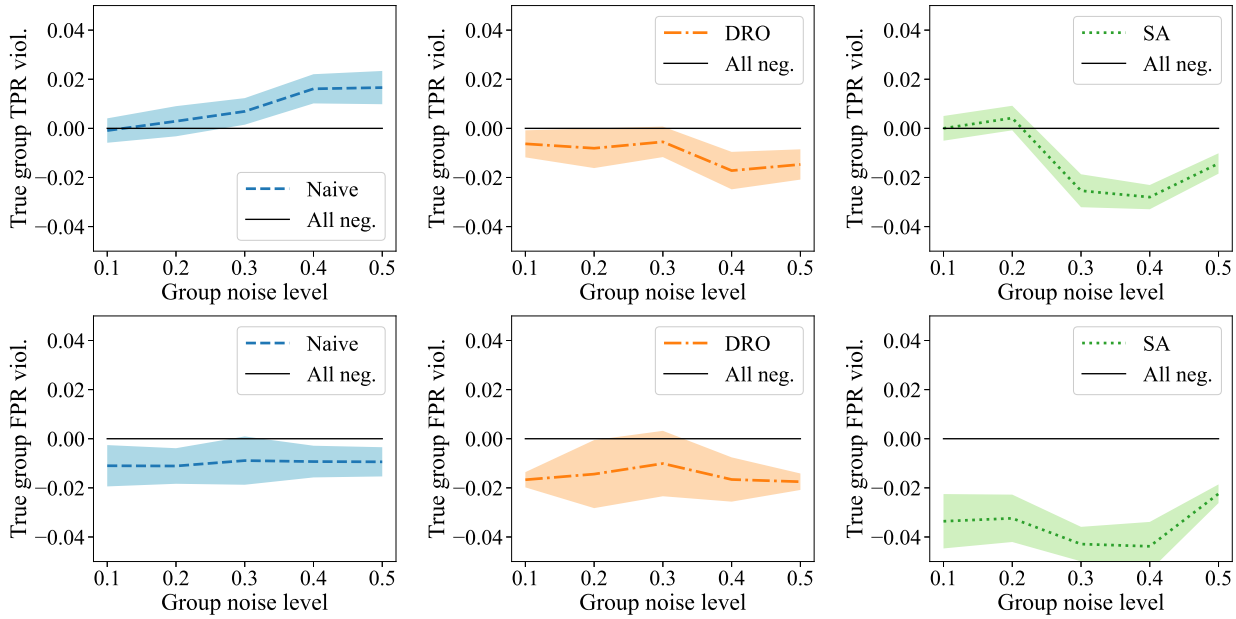


Figure 3.6: Case study 2 (Credit): Maximum true group TPR (top) and FPR (bottom) constraint violations for the Naive, DRO, and soft assignments (SA) approaches on test set for different group noise levels γ on the Credit dataset (mean and standard error over 10 train/val/test splits). The black solid line represents the performance of the trivial "all negatives" classifier, which has constraint violations of 0. A negative violation indicates satisfaction of the fairness constraints on the true groups.

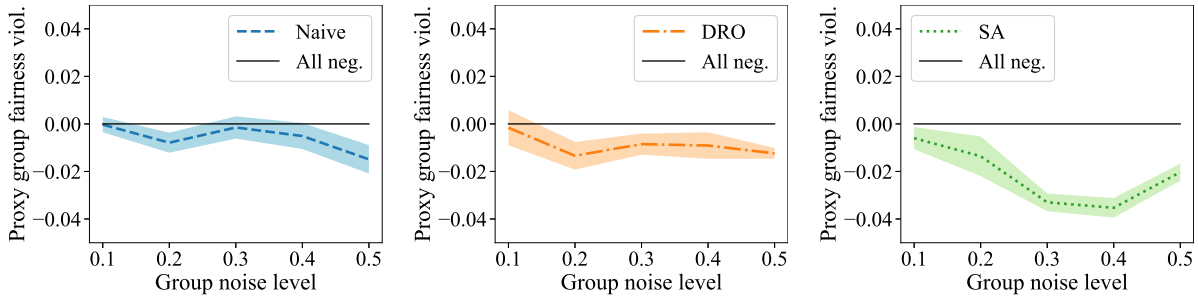


Figure 3.7: Maximum fairness constraint violations with respect to the noisy groups $\hat{G}$ on the test set for different group noise levels γ on the Credit dataset. For each noise level, we plot the mean and standard error over 10 random train/val/test splits. The black solid line illustrates a maximum constraint violation of 0. While the naïve approach (*left*) has increasingly higher fairness constraints with respect to the true groups as the noise increases, it always manages to satisfy the constraints with respect to the noisy groups $\hat{G}$

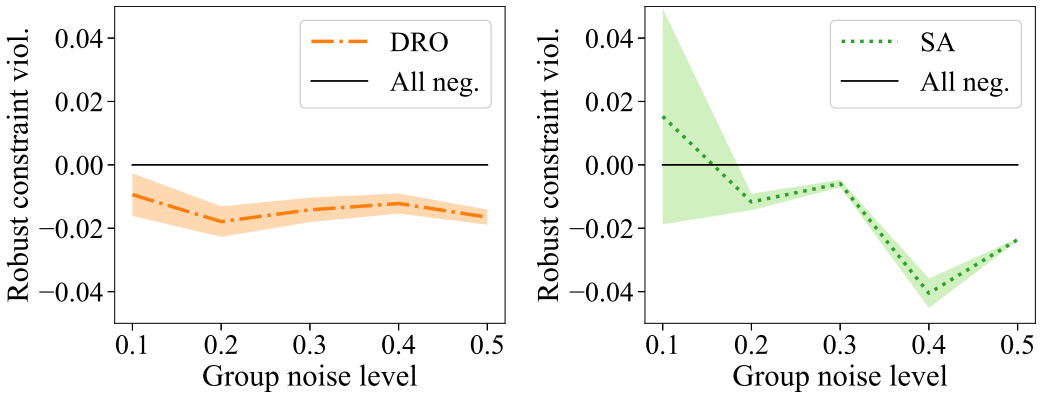


Figure 3.8: Maximum robust constraint violations on the test set for different group noise levels $P(\hat{G} \neq G)$ on the Credit dataset. For each noise level, we plot the mean and standard error over 10 random train/val/test splits. The black dotted line illustrates a maximum constraint violation of 0. Both the DRO approach (*left*) and the soft group assignments approach (*right*) managed to satisfy their respective robust constraints on the test set on average for all noise levels.

Chapter 4

Learning Competitive Equilibria in Exchange Economies with Bandit Feedback

4.1 Introduction

An exchange economy (EE) is a classical micro-economic construct used to model situations where multiple rational agents share a finite set of scarce resources. Such scenarios arise frequently for applications in operations management, urban planning, crowd sourcing, wireless networks, and sharing resources in data centers [51, 67, 89, 111, 118, 197]. In an EE, agents share a set of resources consisting of multiple resource types. They begin with an initial endowment and then exchange these resources among themselves based on a price system. This exchange process allows two agents to trade different resource types if they find it mutually beneficial to do so. Under certain conditions, continually trading in this manner results in a *competitive equilibrium* (CE), where the allocations have desirable Pareto-efficiency and fairness properties. EEs have attracted much research attention, historically since they are tractable models to study human behavior and price determination in real-world markets, and more recently for designing multi-resource fair division mechanisms [18, 19, 40, 58, 63, 204].

One of the most common use cases for fair division, which will be especially pertinent in this work, occurs in the context of shared computational resources. For instance, in a data center shared by an organization, we wish to allocate resources such as CPUs, memory, and GPUs to different users who wish to share this cluster in a way that is Pareto-efficient (so that the resources are put into good use) and fair (for long-term user satisfaction). Here, unlike in real world economies where agents might trade with each other until they reach an equilibrium, the equilibrium is computed using a central mechanism (e.g. a cluster manager) based on the preferences submitted by the agents to obtain an allocation with the above properties. Indeed, fair division mechanisms are a staple in many popular multi-

tenant cluster management frameworks used in practice, such as Mesos [113], Quincy [120], Kubernetes [42], and Yarn [209]. Due to this strong practical motivation, a recent line of work has studied such fair division mechanisms for resource sharing in a compute cluster [48, 91, 92, 178], with some of them based on exchange economies and their variants [107, 143, 207, 224].

However, prior work on EEs and fair division typically assumes knowledge of the agent preferences, in the form of a utility function which maps an allocation of the m resource types to the value the agent derives from the allocation. For instance, in the above example, an application developer needs to quantify how well her application performs for each allocation of CPU/memory/GPU she receives. At best, doing so requires the laborious and often erroneous task of profiling their application [64, 165], and at worst, it can be infeasible due to practical constraints [190, 210]. However, having received an allocation, application developers find it easier to report feedback about the utilities based on the performance they achieved. Moreover, in many real-world systems, this feedback scheme can often be automated [113].

Contributions & summary of results

We study a multi-round mechanism for computing CE in an exchange economy so as to generate fair and efficient allocations when the exact utilities are *unknown* a priori. A central mechanism is used to learn the user utilities over time via feedback from the agents. At the beginning of each round, the mechanism generates allocations; at the end of the round, agents report feedback on the allocation they received. The mechanism then uses this information to better learn the preferences. In particular, we focus on applications for fair division where a centralized mechanism can compute an allocation of these resources on each round, say, by estimating the utilities and finding their equilibria.

In this pursuit, we first formalize this online learning task and construct two loss functions: the first L^{CE} directly builds on the definition of a CE, while the latter L^{PE} is motivated by the fairness and Pareto-efficiency considerations that arise in fair division. To make the learning problem tractable, we focus on a parametric class of utilities which include the constant elasticity of substitution (CES) utilities which feature prominently in the econometric literature and other application-specific utilities used in the systems literature.

We develop a randomized online mechanism which efficiently learns utilities over rounds of allocations while simultaneously striving to achieve Pareto-efficient and fair allocations. We show that this mechanism achieves $\widetilde{O}(\sqrt{T})$ loss for the two loss functions with both in-expectation and high-probability upper bounds (Theorems 4.4.1 and 4.4.2), under a general family of utility functions. To the best of our knowledge, this is the first work that studies CE without knowledge of user utilities; as such different analysis techniques are necessary. For instance, finding a CE is distinctly different from a vanilla optimization task, and common strategies in bandit optimization such as upper-confidence-bound (UCB) based algorithms do not apply (details in 4.4). Instead, our algorithm uses a sampling procedure to balance the exploration-exploitation trade-off. We develop new techniques both to bound the losses and

to analyse the algorithm. Finally, we corroborate these theoretical insights with empirical simulations.

Related work

Our work builds on a rich line of literature at the intersection of microeconomics and machine learning. This richness is not surprising: many real world systems are economic and multi-agent in nature, where decisions taken by or for one agent are weighed against the considerations of others, especially when these agents have competing goals such as in resource allocation, matching markets, and in auction-like settings.

As in this work, several works have studied online learning formulations to handle situations where the agents' preferences are not known a priori, but can be learned from repeated interactions [12, 17, 21, 75, 124, 129]. Our setting departs from these as we wish to learn agent preferences in an exchange economy, with a focus on designing fair division mechanisms.

Since the seminal work of Varian [207], fair division of multiple resource types has received significant attention in the game theory, economics, and computer systems literature. One of the most common perspectives on this problem is as an exchange economy (or as a Fisher market, which is a special case of an EE). Moreover, fair allocation mechanisms have been deployed in many practical resource allocation tasks when compute resources are shared by multiple users. Due to space constraints, we defer a more detailed overview on this line of works in Appendix 4.11.

Notably, in all of the above cases, an important requirement for the mechanism is that agent utilities be known ahead of time. Some work has attempted to lift this limitation by making explicit assumptions on the utility, but it is not clear that if these assumptions hold in practice [147, 224]. Recently, Kandasamy et al. [130] provides a general method for learning agent utilities for fair division using feedback. However, they only study a *single-resource* setting and do not do not explore multiple resource types. Crucially, in the multi-resource setting, one agent can exchange a resource of one type for a different type of resource from another user, so that both are better off after the exchange. Thus, learning in a *multi-resource* setting is significantly more challenging than the single-resource case since there is no notion of exchange, and requires new analysis techniques.

4.2 Background

We first present some necessary background material on exchange economies, their competitive equilibria, and fair division mechanisms.

Exchange economies

In an exchange economy, we have n agents and m divisible resource types. Each agent $i \in [n]$ has an endowment, $e_i = (e_{i1}, \ldots, e_{im})$, where e_{ij} can be viewed the amount of resource j agent

i brings to the economy for trade. In the shared compute cluster example, e_i may represent agent i's contribution to this cluster. Without loss of generality we assume $\sum_{i\in[n]} e_{i1} = 1$ so that the space of resources is denoted by $[0,1]^m$.

We denote an allocation of these resources to the n agents by $x = (x_1, x_2, \ldots, x_n)$, where $x_i \in [0,1]^m$ and x_{ij} denote the amount of resource j that is allocated to agent i. The set of all feasible allocations is therefore $\mathcal{X} = \{x : \sum_{i=1}^m x_{ij} \leqslant \mathbf{1}, x_{ij} \geqslant 0, \forall i \in [n], j \in [m]\}$.

An agent's utility function is simply $u_i : [0,1]^m \to [0,1]$, where $u_i(x_i)$ represents her valuation for an allocation x_i she receives. Here u_i is non-decreasing, i.e., $u_i(x_i) \leqslant u_i(x_i')$ for all $x_i \leqslant x_i'$ element-wise (more allocations will not hurt).

In an exchange economy, agents exchange resources based on a price system. We denote a price vector by p, where $p \in \mathbb{R}_+^m$ and $1^\top p = 1$ (the normalization accounts for the fact that only relative prices matter). Here p_j denotes the price for resource j. Given a price vector p, an agent i has a *budget* $p^\top e_i$, which is the *monetary* value of her endowment according to the prices in p. As this is an economy, a rational agent will then seek to maximize her utility under her budget:

$$d_i(p) = \arg\max_{x_i \in [0,1]^m} u_i(x_i) \quad \text{subject to } p^\top x_i \leqslant p^\top e_i. \tag{4.1}$$

While generally, the preferred allocations $d_i(p)$ form a set, for simplicity we will assume it is a singleton and treat d_i as a function which outputs an allocation for agent i. This is justified under very general conditions [160, 208]. We refer to $d_i(p)$ chosen in the above manner as the agent i's demand for prices p.

Competitive equilibria – definition, existence and uniqueness: A natural way to allocate resources to agents is to set prices p for the resources, and have the agents maximize their utility under this price system. That is, we allocate $x(p) = (x_1, \ldots, x_n)$. Unfortunately, such an allocation may be infeasible, and even if it were, it may not result in an efficient allocation. However, under certain conditions, we can compute a *competitive equilibrium* (CE), where the prices have both of the desired properties:

Definition 4.2.1 (Competitive (Walrasian) Equilibrium)**.** A CE is a pair of allocations and prices $(x^\star, p^\star)$ such that ***(i)*** the allocations are feasible and ***(ii)*** all agents maximize their utilities under the budget induced by prices $p^\star$. Precisely,

$$\sum_{i\in[n]} x_{i,j}^\star \leqslant \sum_{i\in[n]} e_{ij} = 1, \quad \forall\, j \in [m],$$
$$x_i^\star = d_i(p^\star), \quad \forall\, i \in [n].$$

Some definitions of a CE require that the first condition above being an exact equality (e.g., [160]). However, when the utilities are strictly increasing (which will be the case in the sequel), both definitions coincide [208].

Utilities. In general, CEs do always exist but may not be unique. However, one important class of utilities that guarantee this condition with much attention in the fair division literature is the constant elasticity of substitution (CES) utility. Due to its favorable properties, CES utilities are widely-studied in many fair division works, and most of the existing algorithms that generate fair and efficient allocations assume CES utilities or its sub-classes [160, 208]. CES utilities are also ubiquitous in the microeconomics literature; due to this flexibility in interpolating between perfect substitutability and complementary, they are also able to approximate several real-world utility functions. Moreover, computationally, there are efficient methods for computing a CE in the CES and related classes [38, 224, 225]. In contrast, even when CE exist, they may be hard to find under more general classes of utilities [208].

Example 4.2.2 (CES utilities)**.** A CES utility takes the form $u_i(x) = \left(\sum_{j=1}^{m} \theta_{ij} x_i^{\rho}\right)^{1/\rho}$ where ρ is the elasticity of substitution, and $\theta_i = (\theta_{i1}, \ldots, \theta_{im})$ is an agent-specific parameter. When $\rho = 1$, this corresponds to linear utilities where goods are perfect substitutes. As $\rho \to \infty$, the utilities approach perfect complements.

Fair division

We describe exchange economies which are used in fair-division mechanisms. We first formally define the fair division problem.

In a standard mechanism for fair division when the utilities are inputs, each agent truthfully[1] submits her utility u_i to the mechanism. The mechanism then returns an allocation $x \in \mathcal{X}$ that are not only efficient but also fair, which satisfies the following two requirements: *sharing incentive* (SI) and *Pareto-efficiency* (PE). An allocation $x = (x_1, \ldots, x_m)$ satisfies SI if the utility an agent receives is at least as much as her utility when using her endowment, i.e. $u_i(x_i) \geqslant u_i(e_i)$. This simply states that she is never worse off than if she had kept her endowment to herself, so she has the incentive to participate in the fair division mechanism.

A feasible allocation x is said to be PE if the utility of one agent can be increased only by decreasing the utility of another. Rigorously, an allocation x dominates another x', if $u_i(x_i) \geqslant u_j(x_i')$ for all $i \in [n]$ and there exists some $i \in [n]$ such that $u_i(x_i) > u_i(x_i')$. An allocation is Pareto-efficient if it is not dominated by any other point. We denote the set of Pareto-efficient allocations by $\mathcal{PE}$. One advantage of the PE requirement, when compared to other formalisms which maximize social or egalitarian welfare, is that it does not compare the utility of one agent against that of another. The utilities are useful solely to specify an agent's preferences over different allocations.

EEs in fair division: The above problem description for fair division naturally renders itself to a solution based on EEs. By treating the resource allocation environment as an

[1]Unlike some previous works on fair division [91, 130, 178], we do not study strategic considerations, where agents may attempt to manipulate outcomes in their favor by falsely submitting their utilities.

exchange economy, we may compute its equilibrium to determine the allocations for each agent. Then, the SI property follows from the fact that each agent is maximizing her utility under her budget, and an agent's endowment (trivially) falls under her budget. The PE property follows from the first theorem of welfare economics [160, 208]. Several prior works have used this connection to design fair-division mechanisms for many practical applications [58, 207, 224].

Computing a CE: In order to realize a CE allocation in a fair division mechanism, the mechanism needs to compute a CE given a set of utilities. One way to do this is via tatonnement [208]. While there are general procedures, such as tatonnement [208], they are not guaranteed to converge to an equilibrium even when it exists; moreover, even when they do, the rate of convergence can be slow. This has led to the development of efficient procedures for special classes of functions. One such method is proportional response dynamics (PRD) [224, 225] which converges faster under CES utilities [225] and other classes of utilities [224] when $e_i = \alpha_i \mathbf{1}_m$ for all $i \in [n]$ (with $\sum_i \alpha_i = 1$). In fact, in our evaluations, we adopt PRD for computing a CE, which is a subroutine of the learning algorithm.

We note that in the context of fair division, the CE allocations are more pertinent than the CE prices. While the prices are used to compute fair allocations, they are not used directly in their own right.

4.3 Online learning formulation

We formalize online learning an equilibrium in an exchange economy under bandit feedback, when the exact agent utilities are unknown a priori. We consider a multi-round setting, where in each round t, the mechanism selects (x_t, p_t), where $x_t = (x_{t,1}, \ldots, x_{t,n}) \in \mathcal{X}$ are the allocations for each agent for the current round, and p_t are the prices for units of each resource.

The agents, having experienced their allocation, report stochastic feedback $\{y_{t,i}\}_{i\in[n]}$, where $y_{t,i}$ is σ sub-Gaussian and $\mathbb{E}[y_{t,i}|x_{t,i}] = u_i(x_{t,i})$. The mechanism then uses this information to compute allocations for the next round. As described in Section 4.1, this set up is motivated by use cases in data center resource allocations, where jobs (agents) cannot state their utility upfront, but can report feedback on their performance in an automated way.

Going forward, we slightly abuse notation when referring to the allocations. When $i \in [n]$ indexes an agent, $x_i = (x_{i1}, \ldots, x_{im}) \in [0,1]^m$ denotes the allocation to agent i. When t indexes a round, $x_t = (x_{t,1}, \ldots, x_{t,n}) \in \mathcal{X}$ will refer to an allocation to all agents, where $x_{t,i} = (x_{t,i,1}, \ldots, x_{t,i,m}) \in [0,1]^m$ denotes i's allocation in that round. The intended meaning should be clear from context.

Losses

We study two losses for this setting. The first loss is based directly on the definition of an equilibrium (Def. 4.2.1). For $a \in \mathbb{R}$, denote $a^+ = \max(0, a)$. We define the CE loss ℓ^{CE} of an allocation–price pair (x, p) as the sum, over all agents, of the difference between the maximum attainable utility under price p and the utility achieved by allocation x. The T-round loss L_T^{CE} is the sum of $\ell^{\mathrm{CE}}(x_t, p_t)$ losses over T rounds. We have:

$$\begin{aligned} \ell^{\mathrm{CE}}(x, p) &\stackrel{\text{def}}{=} \sum_{i=1}^{n} \left(\max_{x_i' : p^\top x_i' \leqslant p^\top e_i} u_i(x_i') - u_i(\mathbf{x}) \right)^+, \\ L_T^{\mathrm{CE}} &\stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell^{\mathrm{CE}}(x_t, p_t). \end{aligned} \tag{4.2}$$

It is straightforward to see that for a CE pair $(x^\star, p^\star)$, we have $\ell^{\mathrm{CE}}(x^\star, p^\star) = 0$. As this loss is based directly on the definition of a CE, it captures many of the properties of a CE.

Our second loss is motivated by the fair division use case. Recall from Sec. 4.2 that in fair division, while prices are useful in computing CE allocations, they have no value in their own right. Therefore, we will motivate our loss function based on the sharing incentive (SI) and Pareto-efficiency (PE) desiderata for fair division. It is composed of two parts. We define the SI loss ℓ^{SI} for an allocation x as the sum, over all agents, of how much they are worse off than their endowment utilities. We define the PE loss ℓ^{PE} for an allocation x as the minimum sum, over all agents, of how much they are worse off than some Pareto-efficient utilities. Next, we define the fair division loss ℓ^{FD} as the maximum of ℓ^{SI} and ℓ^{PE}. Finally, we define the T-round loss L_T^{FD} for the online mechanism as the sum of $\ell^{\mathrm{FD}}(x_t)$ losses over T rounds. We have:

$$\begin{aligned} \ell^{\mathrm{SI}}(x) &\stackrel{\text{def}}{=} \sum_{i=1}^{n} \left(u_i(e_i) - u_i(\mathbf{x}_i)\right)^+, \\ \ell^{\mathrm{PE}} &\stackrel{\text{def}}{=} \inf_{x' \in \mathcal{PE}} \sum_{i=1}^{n} \left(u_i(\mathbf{x}_i') - u_i(\mathbf{x}_i)\right)^+, \\ \ell^{\mathrm{FD}}(x) &\stackrel{\text{def}}{=} \max\left(\ell^{\mathrm{PE}}(x), \ell^{\mathrm{SI}}(x)\right), \\ L_T^{\mathrm{FD}} &\stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell^{\mathrm{FD}}(x_t). \end{aligned} \tag{4.3}$$

Note that individually achieving either small ℓ^{SI} or ℓ^{PE} is trivial: if an agent's utility is strictly increasing, then by allocating all the resources to this agent we have zero ℓ^{PE} as such an allocation is Pareto-efficient; moreover, by simply allocating each agent their endowment we have zero ℓ^{SI}. In ℓ^{FD}, we require both to be simultaneously small which necessitates a clever allocation that accounts for agents' endowments and utilities. One intuitive interpretation of the PE loss is that it can be bounded above by the L_1 distance to the Pareto-front in utility

space; i.e. denoting the set of Pareto-efficient utilities by $U_{\mathrm{PE}} = \{\{u_i(x_i)\}_{i\in[n]}; x \in \mathcal{PE}\} \subset \mathbb{R}^n$, and letting $u(x) = (u_1(x_1), \ldots, u_n(x_n)) \in \mathbb{R}^n$, we can write, $\ell^{\mathrm{PE}}(x) \leqslant \min_{u\in U_{\mathrm{PE}}} \|u - u(x)\|_1$.

The FD loss is more interpretable as it is stated in terms of the SI and PE requirements for fair division. On the other hand, the CE loss is less intuitive. Moreover, in EEs, while prices help us determine the allocations, they do not have value on their own. Given this, the CE loss has the somewhat undesirable property that it depends on the prices p_t. That said, since the CE loss is based directly on the definition of a CE, it captures other properties of a CE that are not considered in ℓ^{FD} (see an example in Appendix 4.11). It is also worth mentioning that either loss cannot be straightforwardly bounded in terms of the other.

Note that we have presented a basic version of the online learning framework as it provides a simplest platform to study the learning problem of efficient and fair allocations. For instance, one could consider richer settings where the utilities might change over time with certain contextual information. While these settings are beyond the scope of this work, we believe the analysis techniques and intuitions developed here are also insightful in analysing other variant settings.

Model and assumptions

To make the learning problem tractable, we make some additional assumptions on the problem. We consider the following parametric class of utility functions $\mathcal{P}$.

Let $\phi_j : [0,1] \to [0,1]$ be an increasing function which maps the allocation x_{ij} of resource j to agent i to some feature value. For brevity, we will write $\phi : [0,1]^m \to [0,1]^m$, such that $\phi(x_i) = (\phi_1(x_{i1}), \phi_2(x_{i2}), \ldots, \phi_m(x_{im}))$; Next, let $\mu : \mathbb{R}_+ \to [0,1]$ be an increasing function. Finally, let $\Theta \subset \mathbb{R}^m_+$ be a set of positive parameters. Then, we consider the following class of utilities $\mathcal{P}$:

$$\mathcal{P} = \Bigg\{ \{u_i\}_{i=1}^n; \;\; u_i(x_i) = \mu\left(\theta_i^\top \phi(x_i)\right) \\ \text{for some } \theta_i \in \Theta, \; \forall i \in [n] \Bigg\} \tag{4.4}$$

An agent's utility then takes the form $u_i(x_i) = \mu({\theta_i^*}^\top \phi(x_i))$ where the featurization ϕ and the function μ are known, but the true parameters $\theta_i^* \in \Theta$ are unknown and need to be learned by the mechanism.

We consider the above class of functions for the following reasons. First, observe that it represents a valid class of utilities in that for all positive θ, the utilities are increasing in the allocations. Second, a CE is guaranteed to exist uniquely in this class. Third, from a practical point of view, it subsumes a majority of utilities studied in the fair division literature, such as linear utilities, the CES utilities from Example 4.2.2 [18, 19, 40, 58, 204], and other application-specific utilities [210, 224], Fourth, also from a practical point of view, the CE can be efficiently computed on this class [225]. Finally, it also allows us to leverage techniques for estimating generalized linear models in our online learning mechanism [47, 84].

We will also assume the following regularity conditions on $\mathcal{P}$ to avoid some degenerate cases in our analysis. First, μ is continuously differentiable, it is Lipschitz-continuous with constant L_μ, and $C_\mu = \inf_{\theta\in\Theta,\mathbf{x}\in\mathcal{X}} \dot{\mu}\left(\theta^\top\phi(\mathbf{x})\right) > 0$. Second, $\Theta \subset [\theta_{min}, \infty)^m$, where $\theta_{min} > 0$. These assumptions can be relaxed (albeit with a more involved analysis), or replaced by other equivalent regularity conditions [47, 84], without affecting the main analysis ideas or take-aways in this paper. Our results also apply when μ, ϕ, and Θ can be defined separately for each agent, but we assume they are the same to simplify the exposition.

4.4 Algorithm and theoretical results

Algorithm 7 A Randomized Alg. for Learning in EEs

Input: number of initialization sub-phases $M \geqslant 1$, confidence parameters $\{\delta_t\}_{t\geqslant 1}$.
$t \leftarrow 0$
for $\ell = 1, \ldots, M$ **do** // Initialization phase
  for $k = 1, \ldots, \max(m, n)$ **do**
    $t \leftarrow t + 1$, $x_t \leftarrow (\mathbf{0}_m, \ldots, \mathbf{0}_m)$
    for $h = 1, \ldots, \min(m, n)$ **do**
      if $m < n$ **then**
        $x_{t,h+k-1,j} \leftarrow 1$ for all $j \in [m]$.
      else
        $x_{t,i,h+k-1} \leftarrow 1$ for all $i \in [n]$.
      end if
    end for
    Allocate x_t and observe rewards $\{y_{t,i}\}_{i\in[n]}$.
  end for
end for
while True do // Round for learning phase
  $t \leftarrow t + 1$
  for $i = 1, 2, \ldots, n$ **do**
    Compute $Q_{t,i} \stackrel{\text{def}}{=} \sum_{s=1}^{t-1} \phi(\mathbf{x}_{s,i})\phi(\mathbf{x}_{s,i})^\top$
    Compute $\bar{\theta}_{t,i} = \arg\min_{\theta\in\Theta} \left\| \sum_{s=1}^{t-1} \phi(\mathbf{x}_{s,i})\Big(\mu(\theta^\top\phi(\mathbf{x}_{s,i})) - y_{s,i}\Big)\right\|_{Q_{t,i}^{-1}}$
    Sample $\theta'_{t,i} \sim \mathcal{N}(\bar{\theta}_{t,i}, \alpha_t^2 Q_{t,i}^{-1})$. // See (4.5) for α_t.
    $\theta_{t,i} \leftarrow \arg\min_{\theta'\in\Theta} \|\theta'_{t,i} - \theta'\|$. // Projection
  end for
  Choose allocations and prices $x_t, p_t = \mathrm{CE}(\{u_{t,i}\}_{i=1}^n)$, where $u_{t,i}(\cdot) = \mu(\theta_{t,i}^\top\phi(\cdot))$
  Observe rewards $\{y_{t,i}\}_{i\in[n]}$.
end while

We present a randomized online learning algorithm for learning the agents' utilities and generating fair and efficient allocations. Note that this algorithm not only needs to learn the unknown utilities quickly, but should also simultaneoulsy find the CE allocation. This latter aspect introduces new challenges in our setting. For instance, the most popular approach for stochastic optimization under bandit feedback are based on upper-confidence-bounds (UCB). However, finding a CE cannot be straightforwardly framed as a vanilla optimization procedure and hence UCB procedures do not apply. Instead, our proposed algorithm uses a key randomized sampling step, which tradeoffs between exploration and exploration while maintaining the utilities' shape constraints in every round for computing the CE (details in proof sketch).

The algorithm, outlined in Algorithm 7, takes input parameters M and $\{\delta_t\}_{t\geqslant 1}$ whose values we will specify shortly. It begins with an initialization phase for M sub-phases (line 3), each of length $\min(n, m)$. During each sub-phase, we allocate each resource entirely to each user for at least one round. This initialization phase ensures that some matrices we define subsequently are well conditioned.

After the initialization phase, the algorithm operates on each of the remaining rounds as follows. For each user, it first computes quantities $Q_{t,i} \in \mathbb{R}^{m\times m}$ and $\bar{\theta}_{t,i} \in \mathbb{R}^m$ as defined in lines 19, and 20. As we explain shortly, $\bar{\theta}_{t,i}$ can be viewed as an estimate of θ^*_i based on the data from the first $t-1$ rounds. The algorithm then samples $\theta'_{t,i} \in \mathbb{R}^m$ from a normal distribution with mean $\bar{\theta}_{t,i}$ and co-variance $\alpha_t^2 Q_{t,i}$, where, α_t is defined as:

$$\begin{aligned} \alpha_t^2 &= 4\frac{\kappa^2\sigma^2}{C_\mu^2} m \log(t) \log\left(\frac{m}{\delta_t}\right), \\ \kappa &= 3 + 2\log\left(1 + 2\|\phi(\mathbf{1})\|_2^2\right). \end{aligned} \tag{4.5}$$

The sampling distribution, which is centered at our estimate $\bar{\theta}_{t,i}$, is designed to balance the exploration-exploitation trade-off on this problem. Next, it projects the sampled $\theta'_{i,t}$ onto Θ to obtain $\theta_{t,i}$.

In line (24), the algorithm obtains an allocation and price pair x_t, p_t by computing the CE on the $\theta_{t,i}$ values obtained above, i.e. by pretending that $u_{t,i}(\cdot) = \mu(\theta_{t,i}^\top \phi(\cdot))$ is the utility for user i.

It is important to note that the computation of the CE happens as a *subroutine* of the mechanism, and users will simply receive the allocations x_t. The mechanism collects the rewards $\{y_{t,i}\}_{i\in[n]}$ from each user and then repeats the same for the remaining rounds. As we discussed in Sec. 4.2, there are different ways to compute a CE efficiently in our setting, including tatonnement or the proportional response dynamics (PRD) algorithm [225] which we implemented. Given that our algorithm focus on learning the efficient and fair allocations, we do not focus on the computation complexity of CE in this work. Empirically, we find PRD converges quickly in the simulations.

Computation of $\bar{\theta}_{it}$: It is worth explaining steps 19–20 used to obtain the estimate $\bar{\theta}_{it}$ for user i's parameter θ^*_i. Recall that for each agent i, the mechanism receives stochastic rewards

$y_{t,i}$ where $y_{t,i}$ is a σ sub-Gaussian random variable with $\mathbb{E}[y_{t,i}] = u_i(\mathbf{x}_{t,i})$ in round t. Therefore, given the allocation-reward pairs $\{(\mathbf{x}_{s,i}, y_{s,i})\}_{s=1}^{t-1}$, the maximum quasi-likelihood estimator $\widehat{\theta}_{t,i}^{MLE}$ for θ_i is defined as the maximizer of the quasi-likelihood $\mathcal{L}(\theta) = \sum_{s=1}^{t-1} \log p_\theta(y_{s,i}|\mathbf{x}_{s,i})$, where $p_\theta(y_i|\mathbf{x}_i)$ is as defined below. Here, $\mu(\nu) = \frac{\partial b(\nu)}{\partial \nu}$ and $c(\cdot)$ is a normalising term. We have:

$$p_\theta(y_i|\mathbf{x}_i) = \exp\left(y_i \theta^\top \phi(\mathbf{x}_{si}) - b(\theta^\top \phi(\mathbf{x}_{si})) + c(y_i)\right). \tag{4.6}$$

Upon differentiating, we have that $\widehat{\theta}_{t,i}^{MLE}$ is the unique solution of the estimating equation:

$$\sum_{s=1}^{t-1} \phi(\mathbf{x}_{si}) \left(\mu\left(\left(\widehat{\theta}_{t,i}^{MLE}\right)^\top \phi(\mathbf{x}_{s,i})\right) - y_{si}\right) = 0.$$

In other words, $\theta_{t,i}^{MLE}$ would be the maximum likelihood estimate for θ_i^* if the rewards $y_{t,i}$ followed an exponential family likelihood as shown in (4.6). Our assumptions are more general; we only assume the rewards are sub-Gaussian centred at $\mu(\theta_i^{*\top} \phi(x_{t,i}))$. However, this estimate is known to be consistent under very general conditions, including when the rewards are sub-Gaussian [47, 84]. Since $\hat{\theta}_{t,i}^{MLE}$ might be outside of the set of feasible parameters Θ, this motivates us to perform the projection in the $Q_{t,i}^{-1}$ norm to obtain $\bar{\theta}_{t,i}$ as defined in line 20. Here, $Q_{t,i}$, defined in line (19), is the design matrix obtained from the data in the first $t-1$ steps.

On the algorithm design: It is worth comparing the design of our algorithm against prior work in the bandit literature under similar parametric assumptions [59, 84, 154, 188]. For instance, in a CE, each agent is maximizing their utility under a budget constraint. Therefore, a seemingly natural idea is to adopt a UCB based procedure, which is the most common approach for stochastic optimization under bandit feedback [14]. However, adopting a UCB-style method for our problem proved to be unfruitful. Consider using a UCB of the form $\mu(\bar{\theta}_{it}^\top \phi(\cdot)) + U_{it}(\cdot)$, where U_{it} quantifies the uncertainty in the current estimate. Unfortunately, a CE is not guaranteed to exist for utilities of the above form, which means that finding a suitable allocation can be difficult. An alternative idea is to consider UCBs of the form $\mu(\widehat{\theta}_{t,i}^\top \phi(\cdot))$ where $\widehat{\theta}_{t,i}$ is an upper confidence bound on θ_i^* (recall that both θ_i^* and ϕ are non-negative). While CEs are guaranteed to exist for such UCBs, $\widehat{\theta}_{t,i}$ is not guaranteed to uniformly converge to θ_i^*, resulting in linear loss.

Instead, our algorithm takes inspiration from classical Thompson sampling (TS) procedure for multi-armed bandits in the Bayesian paradigm [203]. The sampling step in line 21 is akin to sampling from the posterior beliefs in TS. It should be emphasized that the sampling distributions on each round cannot be interpreted as the posterior of some prior belief on θ_i^*. In fact, they were designed so as to put most of their mass inside a frequentist confidence set for θ_i^*.

Upper bounds on the loss

The following two theorems are the main results bounding the loss terms $L^{\mathrm{FD}}, L^{\mathrm{CE}}$ for Algorithm 7. In the first theorem, we are given a target failure probability of at most δ. By choosing δ_t appropriately, we obtain an infinite horizon algorithm for which both loss terms are $\widetilde{O}(\sqrt{T})$ with probability at least $1-\delta$. In the second theorem, with a given time horizon T, we obtain an algorithm whose expected losses are $\widetilde{O}(\sqrt{T})$.

Theorem 4.4.1. *Assume the conditions in Section 4.3. Let $\delta > 0$ be given. Choose $\delta_t = \frac{2\delta}{n\pi^2 t^2}$. Then, the following upper bounds on $L^{\mathrm{FD}}, L^{\mathrm{CE}}$ hold for Algorithm 7 with probability at least $1-\delta$.*

$$L^{\mathrm{FD}}(T),\, L^{\mathrm{CE}}(T) \;\in\; O\left(n\left(m + \tfrac{m^2}{\sqrt{M}}\right)\sqrt{T}\big(\log(nT/\delta) + \log(T)\big)\right).$$

Theorem 4.4.2. *Assume the conditions in Section 4.3. Let $T > M\max(m^2, n)$ be given. Choose $\delta_t = \frac{1}{T}$. Then, the follow upper bounds on $L^{\mathrm{FD}}, L^{\mathrm{CE}}$ hold for Algorithm 7.*

$$\mathbb{E}[L^{\mathrm{FD}}(T)],\, \mathbb{E}[L^{\mathrm{CE}}(T)] \;\in\; O\left(n\left(m + \tfrac{m^2}{\sqrt{M}}\right)\sqrt{T}\,(\log(T))\right),$$

Above, probabilities and expectations are with respect to both the randomness in the observations and the sampling procedure. Both theorems show that we can learn with respect to both losses at $\sqrt{T}$ rate. Note that the rates depend on the number of initialization subphases M. By choosing $M = m^2$, we get a $\widetilde{O}(nm\sqrt{T})$ bound. However, this also requires a large initialization phase, which may not be feasible in practice. We can instead choose M to be small, but this leads to correspondingly worse asymptotic bounds.

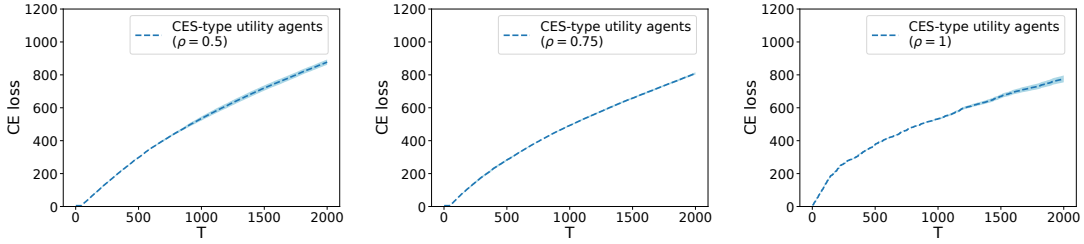


Figure 4.1: The CE loss L_T^{CE} vs the number of rounds T, evaluated with $m = 3$ resource types and $n = 5$ agents with CES utilities. We present results for $\rho = 0.5$, $\rho = 0.75$, and $\rho = 1$ respectively (see Example 4.2.2). All figures show results which are averaged over 10 runs, and the shaded region shows the standard error at each time T.

Proof sketch. Our proof uses some prior martingale concentration results from the bandit literature [84, 188], and additionally, we use some high level intuitions from prior frequentist analyses of Thompson sampling [7, 133, 170]. At the same time, we also require novel

techniques, both to bound the loss terms, and analyse the algorithm. Our proof for bounding L_T^{CE} first defines high probability events $A_{t,i}, B_{t,i}$ for each agent i and round t. $A_{t,i}$ captures the event that the estimated $\bar{\theta}_{t,i}$ is close to θ_i^* in $Q_{t,i}$ norm. We upper bound $\mathbb{P}(A_{t,i}^c)$ using the properties of the maximum quasi-likelihood estimator on GLMs [47, 84] and a martingale argument. $B_{t,i}$ captures the event that the sampled $\theta_{t,i}$ is close to $\bar{\theta}_{t,i}$ in $Q_{t,i}$ norm. Given these events, we then bound the instantaneous losses $\ell^{\mathrm{CE}}(x_t, p_t)$ by a super martingale with bounded differences. The final bound is obtained by an application of the Azuma inequality.

Another key ingredient in this proof is to show that the sampling step also explores sufficiently–the B_{it} event only captures exploitation; since the sampling distribution is a multi-variate Gaussian, this can be conveniently argued using an upper bound on the standard normal tail probability. While bounding L^{FD} uses several results and techniques as above, it cannot be directly related to L^{CE}, and requires a separate analysis.

4.5 Experiments

We evaluated Algorithm 7 with simulations. To the best of our knowledge, this is the first online algorithm studying fair and efficient allocations with unknown utilities with multiple heterogeneous resource types, and there are no existing natural baselines. There is also no straightforward adaptation of the method described in Kandasamy et al. [130] for single resource types since they do not consider the exchange of resources. We evaluated based on two types of utilities.

1. CES utilities: Described in Example 4.2.2.

2. Amdahl's utilities: The Amdahl's utility function, described in Zahedi et al. [224], is used to model the performance of jobs distributed across heterogeneous machines in a data center. This utility is motivated by Amdahl's Law [8], which models a job's speed up in terms of the fraction of work that can be parallelized. Let $0 < f_{ij} < 1$ denote the parallel fraction of user i's job on machine type j. Then, an agent's Amdahl utility is: $u_i(x) = \sum_{j=1}^{m} \theta_{ij}\phi_{ij}(x_{ij})$, where $\phi_{ij}(x_{ij}) = {}^{x_{ij}}/_{f_i+(1-f_i)x_{ij}}$. $\phi_{ij}(x_{ij})$ is the relative speedup produced by allocation x_{ij}. Both CES and Amdahl utilities belong to our class $\mathcal{P}$ given in (4.4).

We focus our evaluation on the CE loss; computing the FD loss is computationally expensive as it requires taking an infimum over the Pareto-front (more details in Appendix 4.11). Our first set of experiments consider an environment with $m = 3$ resource types and $n = 5$ agents, all of whom have CES utilities. We conduct three experiments with different values for the elasticity of substitution ρ. Our second set of experiments consider an environment with $m = 2$ resource types and $n = 8$ agents, all of whom have Amdahl's utilities, where the results are similar and thus included in Appendix 4.10. We conduct three experiments with different values for the parallel fraction f_{ij}. All experiments are run for $T = 2000$ rounds, where we set $\delta = \frac{1}{T}$. The results are given in Figure 4.1. They show that the CE loss grows

sublinearly with T which indicates that the algorithm is able to learn utilities and compute a CE.

To compute the CE at line 24 of Algorithm 7, we use the proportional response dynamics procedure from [38, 225] with 20 iterations. To compute L^{CE}, we need to maximize each agent's utility subject to a budget. Full experimental details and additional results are included in Appendix 4.10.

4.6 Conclusion and future directions

We introduced and studied the problem of online learning a competitive equilibrium in an exchange economy, without a priori knowledge of agents' utilities. We quantify the learning performance via two losses, the first motivated from the definition of an equilibrium, and the second by fairness and Pareto-efficiency considerations in fair division. We develop a randomized algorithm which achieves $\tilde{O}(nm\sqrt{T})$ loss after T rounds under both losses, and corroborate these theoretical results with simulations. While our work takes the first step towards sequentially learning a market equilibrium in exchange economies, an interesting avenue for future work would be to study learning approaches in broader classes of agent utilities and market dynamics.

Our work addresses the technical challenge of efficient and fair allocations when agents' utilities are unknown, with a specific fairness notion on sharing incentives. However, we emphasize that there are other notions of fairness in the fair division literature, with varying connections to prior sociopolitical framing. On a broader societal level, this work, as with many other fair allocation algorithms, if being applied to scenarios where the choice of fairness criteria is not appropriate can lead to potential negative impact. Thus, we emphasize that whether our model is applicable to certain applications should be carefully evaluated by domain experts, along with awareness of the tradeoffs involved.

4.7 Appendix: Technical lemmas

We first provide some useful technical lemmas.

Lemma 4.7.1. *Suppose that Z is a χ^2_m random variable, i.e. $Z = \sum_{k=1}^m Z_k^2$, where for all k, Z_k are i.i.d. random variables drawn from $\mathcal{N}(0,1)$. Then,*

$$P(Z > m + \alpha) = \begin{cases} e^{-\frac{\alpha}{8}} & \alpha > m \\ e^{-\frac{\alpha^2}{8m}} & \alpha \leqslant m. \end{cases}$$

Proof. Suppose that X is sub-exponential random variable with parameters (ν, b) and expectation μ. Applying well known tail bounds for sub-exponential random variables (e.g.

[214]) yields:

$$P(X > \mu + \alpha) = \begin{cases} e^{-\frac{\alpha^2}{2\nu^2}} & 0 \leqslant \alpha \leqslant \frac{\nu^2}{b}, \text{ and} \\ e^{-\frac{\alpha}{2b}} & \alpha > \frac{\nu^2}{b}. \end{cases}$$

The lemma follows from the fact that a χ^2_m random variable is sub-exponential with parameters $(\nu, b) = (2, 4)$. ■

Lemma 4.7.2. *(Lower bound for normal distributions) Let Z be a random variable $Z \sim \mathcal{N}(0,1)$, then $P(Z > t) \geqslant \frac{1}{t+\sqrt{t^2+4}}\sqrt{\frac{2}{\pi}}e^{-\frac{t^2}{2}}$.*

Proof. First, from Abramowitz et al. [2] (7.1.13) we have,

$$e^{x^2}\int_x^\infty e^{-t^2}dt \geqslant \frac{1}{x+\sqrt{x^2+2}}.$$

Set $t = \sqrt{2}x$, then the above equation yields:

$$P(Z > t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-\frac{x^2}{2}}dx \geqslant \frac{1}{t+\sqrt{t^2+4}}\sqrt{\frac{2}{\pi}}e^{-\frac{t^2}{2}},$$

which completes the proof. ■

Lemma 4.7.3. *(Azuma-Hoeffding inequality[214]) Let $(Z_s)_{s\geqslant 0}$ be a super martingale w.r.t. a filtration $(\mathcal{F}_t)_{t\geqslant 0}$. Let $(B_t)_{t\geqslant 0}$ be predictable processes w.r.t. $(\mathcal{F}_t)_{t\geqslant 0}$, such that $|Z_s - Z_{s-1}| \leqslant B_s$ for all $s \geqslant 1$ almost surely. Then for any $\delta > 0$,*

$$P\left(Z_T - Z_0 \leqslant \sqrt{2\log\left(\frac{1}{\delta}\right)\sum_{t=1}^T B_t^2}\right) \geqslant 1 - \delta.$$

Lemma 4.7.4. *$\forall x \in [0, c], c > 0$, we have $x \leqslant \frac{c}{\log(1+c)}\log(1+x)$.*

Proof. The result follows immediately from the fact that the function $f(x) \stackrel{\text{def}}{=} \frac{x}{\log(1+x)}$ is non-decreasing on $(0, \infty)$. ■

Lemma 4.7.5. *(Lemma 1, Filippi et al. [84]) Let $(\mathcal{F}_k, k \geqslant 0)$ be a filtration, $(m_k; k \geqslant 0)$ be an $\mathbb{R}^d$-valued stochastic process adapted to $(\mathcal{F}_k)$. Assume that η_k is conditionally sub-Gaussian in the sense that there exists some $R > 0$ such that for any $\gamma \geqslant 0$, $k \geqslant 1$, $\mathbb{E}[\exp(\gamma\eta_k)|\mathcal{F}_{k-1}] \leqslant \exp\left(\frac{\gamma^2R^2}{2}\right)$ almost surely. Then, consider the martingale $\xi_t = \sum_{k=1}^t m_{k-1}\eta_k$ and the process $M_t = \sum_{k=1}^t m_{k-1}m_{k-1}^\top$. Assume that with probability one, the smallest eigenvalue of M_d is lower bounded by some positive constant λ_0, and that $\|m_k\|_2 \leqslant c_m$ almost surely for any $k \geqslant 0$. Then, the following holds true: for any $0 < \delta < \min(1, d/e)$ and $t > \max(d, 2)$, with probability at least $1 - \delta$,*

$$\|\xi_t\|_{M_t^{-1}} \leqslant \kappa R\sqrt{2d\log(t)\log(d/\delta)},$$

where $\kappa = \sqrt{3 + 2\log(1 + 2\frac{c_m^2}{\lambda_0})}$.

4.8 Appendix: Bounding L^{CE}

First, consider any round t. We will let $\mathcal{F}_t \stackrel{\text{def}}{=} \sigma\left(\{(\mathbf{x}_{is}, y_{is})_{i=1,s=1}^{n,t-1}\}\right)$ denote the σ-algebra generated by the observations in the first $t-1$ rounds. Clearly, $\{\mathcal{F}_t\}_{t\geqslant 0}$ is a filtration. We will denote $\mathbb{E}_t[\cdot|\mathcal{F}_t] = \mathbb{E}_t[\cdot]$ to be the expectation when conditioning on the past observations up to round $t-1$. Similarly, define $P_t(\cdot) \stackrel{\text{def}}{=} P(\cdot|\mathcal{F}_t) = \mathbb{E}[\mathbb{1}(\cdot)|\mathcal{F}_t]$.

Recall that $\{\delta_t\}_{t\geqslant 0}$ are inputs to the algorithm. Similarly, let $\{\delta_{2t}\}_{t\geqslant 0}$ be a sequence. We will specify values for both sequences later in this proof. Given these, further define the following quantities on round t:

$$\beta_{1t} \stackrel{\text{def}}{=} \frac{2}{C_\mu}\kappa\sigma\sqrt{2m\log(t)}\sqrt{\log\left(\frac{m}{\delta_t}\right)}$$

$$\beta_{2t} \stackrel{\text{def}}{=} \sqrt{\alpha_t(m+\gamma_{2t})} \qquad \text{where,} \qquad \gamma_{2t} \stackrel{\text{def}}{=} \max\left(8\log\left(\frac{1}{\delta_{2t}}\right), \sqrt{8m\log\left(\frac{1}{\delta_{2t}}\right)}\right)$$

$$\beta_{3t} \stackrel{\text{def}}{=} L_\mu(\beta_{1t}+\beta_{2t}).$$

Here, recall that L_μ is the Lipschitz constant of $\mu(\cdot)$, C_μ is such that $C_\mu \stackrel{\text{def}}{=} \inf_{\theta\in\Theta,\mathbf{x}\in\mathcal{X}} \dot{\mu}\left(\theta^\top\phi(\mathbf{x})\right)$, and α_t is a sequence that is defined and used in Algorithm 7.

Next, we consider the following two events:

$$A_{it} \stackrel{\text{def}}{=} \{\|\theta_i^* - \bar{\theta}_{it}\|_{Q_{it}} \leqslant \beta_{1t}\},$$
$$B_{it} \stackrel{\text{def}}{=} \{\|\bar{\theta}_{it} - \theta_{it}\|_{Q_{it}} \leqslant \beta_{2t}\}.$$

where $Q_{it} \stackrel{\text{def}}{=} \sum_{s=1}^{t-1}\phi(\mathbf{x}_{is})\phi(\mathbf{x}_{is})^\top$ is a design matrix that corresponding to the first $t-1$ steps.

Lastly, define

$$\rho_{it}(\mathbf{x}) \stackrel{\text{def}}{=} \|\phi(\mathbf{x})\|_{Q_{it}^{-1}} = \sqrt{\phi^\top(\mathbf{x})Q_{it}^{-1}\phi(\mathbf{x})},$$

and

$$S_{it} \stackrel{\text{def}}{=} \left\{\mathbf{x}\in\mathcal{X} : u_i(\mathbf{x}_{it}^*) - u_i(\mathbf{x}) \geqslant \beta_{3t}\rho_{it}(\mathbf{x})\right\},$$

where $\mathbf{x}_{it}^* = \arg\max_{\mathbf{y}\in\mathcal{X},\mathbf{p}_t^\top\mathbf{y}\leqslant\mathbf{p}_t^\top e_i} u_i(\mathbf{y})$. Here, we used $\mathcal{X}$ to denote the set of feasible allocations for one agent: $\{\mathbf{x}\in\mathbb{R}^m : 0\leqslant\mathbf{x}\leqslant\mathbf{1}\}$.

Intuitively, $\tilde{\mathbf{x}}_{it}$ is the best true optimal affordable allocation for agent i in round t under the price function $\mathbf{p}_t$. Since the set $\{\mathbf{y}\in\mathcal{X}, \mathbf{p}_t^\top\mathbf{y}\leqslant\mathbf{p}_t^\top e_i\}$ is a compact set, the maximum is well defined.

Now we begin our analysis with the following lemmas.

Lemma 4.8.1. *For any round $t > t_0$, $P_t(A_{it}) \geqslant 1-\delta_{1t}$.*

Proof. Define function $g_{it}(\theta) = \sum_{s=1}^{t-1} \mu\left(\theta^\top \phi(\mathbf{x}_{is})\right) \phi(\mathbf{x}_{is})$. Then by the fundamental theorem of calculus, we have

$$g_{it}(\theta_{it}) - g_{it}(\theta_{it}) = G_{it}(\theta_i^* - \bar{\theta}_{it}),$$

where $G_{it} = \int_0^1 \nabla g_{it}\left(s\theta_i^* + (1-s)\bar{\theta}_{it}\right) ds$, and

$$\nabla g_{it}(\theta) = \sum_{s=1}^{t-1} \phi(\mathbf{x}_{is})\phi(\mathbf{x}_{is})^\top \mu'(\theta^\top \phi(\mathbf{x}_{is})).$$

By the definition of C_μ and Q_{it}, we have that $G_{it} \succeq C_\mu Q_{it} \succeq M \cdot I$, where the last inequality follows due to the initialisation scheme. Therefore, G_{it} is invertible and moreover,

$$G_{it}^{-1} \preceq \frac{1}{C_\mu} Q_{it}^{-1}. \tag{4.7}$$

We can write,

$$\theta_i^* - \bar{\theta}_{it} = G_{it}^{-1}\left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it})\right). \tag{4.8}$$

Therefore, we have,

$$\begin{aligned}
&\left(\theta_i^* - \bar{\theta}_{it}\right)^\top Q_{it} \left(\theta_i^* - \bar{\theta}_{it}\right) \\
&= \left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it})\right)^\top G_{it}^{-1} Q_{it} G_{it}^{-1} \left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it})\right) \\
&\leqslant \frac{1}{C_\mu^2} \left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it})\right) Q_{it}^{-1} \left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it})\right) \\
&= \frac{1}{C_\mu^2} \left\| \left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it})\right) \right\|_{Q_{it}^{-1}},
\end{aligned}$$

where the first equality follows from Eq (4.7), and the inequality follows from Eq (4.8).

Therefore,

$$\begin{aligned}
\|\theta_i^* - \bar{\theta}_{it}\|_{Q_{it}} &\leqslant \frac{1}{C_\mu} \left\| \left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it})\right) \right\|_{Q_{it}^{-1}} \\
&\leqslant \frac{2}{C_\mu} \left\| \left(g_{it}(\theta_{it}) - g_{it}(\bar{\theta}_{it}^{MLE})\right) \right\|_{Q_{it}^{-1}} \\
&= \frac{2}{C_\mu} \left\| \sum_{s=1}^{t-1} \phi(\mathbf{x}_{is})(Y_{is} - \mu\left(\phi(\mathbf{x}_{is})^\top \theta_i^*\right)) \right\|_{Q_{it}^{-1}}.
\end{aligned}$$

where the second inequality is from the triangle inequality, and the last equality is from the definition of θ_{it}^{MLE} and g_{it}.

Let A_{it} denote the event that

$$\left\| \sum_{s=1}^{t-1} \phi(\mathbf{x}_{is})(Y_{is} - \mu\left(\phi(\mathbf{x}_{is})^\top \theta_i^*\right)) \right\|_{Q_{it}^{-1}} \leqslant \kappa\sigma\sqrt{2m\log(t)} \sqrt{\log\left(\frac{d}{\delta_t}\right)},$$

then we have A_{it} holds with probability at least δ_t by Lemma 4.7.5.

■

Lemma 4.8.2. *For any round $t > t_0$, $P_t(B_{it}) \geqslant 1 - \delta_{2t}$.*

Proof. First, recall that $B_{it} = \{\|\bar{\theta}_{it} - \theta_{it}\|_{Q_{it}} \leqslant \beta_{2t}\}$. We can now write,

$$\begin{aligned} P(B_{it}^c) &= P(\|\theta_{it} - \bar{\theta}_{it}\|_{Q_{it}} > \beta_{2t}) \\ &\leqslant P(\|\theta'_{it} - \bar{\theta}_{it}\|_{Q_{it}} > \beta_{2t}) \\ &= P(\|\theta'_{it} - \bar{\theta}_{it}\|_{\alpha_t^{-1}Q_{it}} > \alpha_t^{-\frac{1}{2}}\beta_{2t}) \\ &= P(\sqrt{Z} > \sqrt{M_t\gamma_{2t}}), \end{aligned}$$

where $Z = \left(\theta_{it} - \bar{\theta}_{it}\right)^\top \alpha_t^{-2} Q_{it} \left(\theta_{it} - \bar{\theta}_{it}\right)$. The first step simply uses the fact that since $\bar{\theta}_{it}$ is already inside Θ (see line 17 in Algorithm 1), projecting θ'_{it} to be inside Θ after sampling only brings it *even closer* to $\bar{\theta}_{it}$.

Note that Z is a χ^2_m random variable. This follows from the fact that

$$\theta_{it} \sim \mathcal{N}(\bar{\theta}_{it}, \alpha_t^2 Q_{it}^{-1}),$$

therefore we have

$$\alpha_t^{-1} Q_{it}^{1/2} \left(\theta_{it} - \bar{\theta}_{it}\right) \sim \mathcal{N}(0, I_m).$$

Denote $y = \alpha_t^{-1} Q_{it}^{1/2} \left(\theta_{it} - \bar{\theta}_{it}\right)$, then $Z = y^\top y$ is a χ^2_m random variable.

Therefore, by Lemma 4.7.1, and the definition that $\gamma_{2t} = \max\left(8\log(\frac{1}{\delta_{2t}}), \sqrt{\log(\frac{1}{\delta_{2t}})}\right)$, we have

$$P(B_{it}^c) = P(Z > n + \gamma_{2t}) \leqslant \delta_{2t},$$

which completes the proof. ■

Lemma 4.8.3. *Let $\mathbf{x}$ be arbitrary such that $\mathbf{x} \in \mathcal{X}$. Then,*

$$P_t(u_{it}(\mathbf{x}) > u_i(x) | A_{it}) \geqslant q_0,$$

with $q_0 = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2}+\sqrt{6}} \frac{1}{e} \approx 0.075$.

Proof. First, notice that

$$\begin{aligned} u_{it}(\mathbf{x}) > u_i(\mathbf{x}) &\iff \mu\left(\theta_{it}^\top \phi(\mathbf{x})\right) > \mu\left((\theta_i^*)^\top \phi(\mathbf{x})\right) \\ &\iff \theta_{it}^\top \phi(\mathbf{x}) > (\theta_i^*)^\top \phi(\mathbf{x}) \\ &\iff \frac{\left(\theta_{it} - \bar{\theta}_{it}\right)^\top \phi(\mathbf{x})}{\alpha_t \rho_{it}(\mathbf{x})} > \frac{\left(\theta_i^* - \bar{\theta}_{it}\right)^\top \phi(\mathbf{x})}{\alpha_t \rho_{it}(\mathbf{x})}. \end{aligned}$$

Since $\theta_{it} \sim \mathcal{N}(\bar{\theta}_{it}, \alpha_t^2 Q_{it}^{-1})$, we have

$$\left(\theta_{it} - \bar{\theta}_{it}\right)^\top \phi(\mathbf{x}) \sim \mathcal{N}(0, \alpha_t^2 \phi(x)^\top Q_{it}^{-1} \phi(x)), \implies \left(\theta_{it} - \bar{\theta}_{it}\right)^\top \phi(\mathbf{x}) \sim \mathcal{N}(0, \alpha_t^2 \rho_{it}^2(\mathbf{x})).$$
$$\implies \frac{\left(\theta_{it} - \bar{\theta}_{it}\right)^\top \phi(\mathbf{x})}{\alpha_t \rho_{it}(\mathbf{x})} \sim \mathcal{N}(0,1).$$

From the above we have that

$$P_t\left(u_{it}(\mathbf{x}) > u_i(\mathbf{x}) | A_{it}\right) = P_t\left(Z > \frac{\left(\theta_{it} - \bar{\theta}_{it}\right)^\top \phi(\mathbf{x})}{\alpha_t \rho_{it}(\mathbf{x})} \middle| A_{it}\right),$$

where $Z \sim \mathcal{N}(0,1)$ is sampled independently of the observations, since the randomness in Algorithm 7 can be assumed to be independent of the randomness in the observations. Therefore, under the event A_{it},

$$\begin{aligned}\left|\frac{\left(\theta_{it} - \bar{\theta}_{it}\right)^\top \phi(\mathbf{x})}{\alpha_t \rho_{it}(\mathbf{x})}\right| &= \left|\frac{\left(\theta_{it} - \bar{\theta}_{it}\right)^\top Q_{it}^{\frac{1}{2}} Q_{it}^{-\frac{1}{2}} \phi(\mathbf{x})}{\alpha_t \rho_{it}(\mathbf{x})}\right| \\ &\leqslant \frac{\|\theta_{it} - \bar{\theta}_{it}\|_{Q_{it}} \|\phi(\mathbf{x})\|_{Q_{it}^{-1}}}{\alpha_t \rho_{it}(\mathbf{x})} \\ &\leqslant \frac{\beta_{it}}{\alpha_t} = \frac{\sqrt{8}}{\alpha_0}.\end{aligned}$$

Here, the first inequality follows from the definition of the matrix norm and the definition of A_{it}, and the second inequality follows from the definition of β_{1t}.

Therefore, by Lemma 4.7.2, we have

$$\begin{aligned}&P_t(u_{it}(\mathbf{x}) > u_i(x) | A_{it}) \\ &= P_{Z \sim \mathcal{N}(0,1)}(Z > \frac{\sqrt{8}}{\alpha_0}) \\ &\geqslant \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{8/\alpha_0^2} + \sqrt{4 + \sqrt{8/\alpha_0^2}}} e^{-\frac{4}{\alpha_0^2}}.\end{aligned}$$

Setting $\alpha_0^2 = 4$, we have

$$P_t\left(u_{it}(\mathbf{x}) > u_i(x) | A_{it}\right) \geqslant \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2} + \sqrt{6}} \frac{1}{e} \approx 0.075,$$

which completes the proof. ■

Lemma 4.8.4. *Let $\theta_1, \theta_2 \in \Theta \in \mathbb{R}^m$. Let $Q \succeq 0, Q \in \mathbb{R}^{m \times m}$ be a positive semi-definite matrix, and $\rho_Q(\mathbf{x}) = \sqrt{\phi(\mathbf{x})^\top Q^{-1} \phi(\mathbf{x})}$. Then,*

$$\left|\mu(\theta_1^\top \phi(\mathbf{x})) - \mu(\theta_2^\top \phi(\mathbf{x}))\right| \leqslant L_\mu \|\theta_1 - \theta_2\|_Q \cdot \rho_Q(\mathbf{x}).$$

Proof. This follows from the Lipschitz properties of μ and the following simple calculations:

$$\begin{aligned}\left|\mu(\theta_1^\top\phi(\mathbf{x})) - \mu(\theta_2^\top\phi(\mathbf{x}))\right| &\leqslant L_\mu|(\theta_1-\theta_2)^\top\phi(\mathbf{x})| \\ &= L_\mu|(\theta_1-\theta_2)^\top Q^{\frac{1}{2}}Q^{-\frac{1}{2}}\phi(\mathbf{x})| \\ &\leqslant L_\mu\|\theta_1-\theta_2\|_Q\|\phi(\mathbf{x})\|_{Q^{-1}} \\ &= L_\mu\|\theta_1-\theta_2\|_Q\cdot\rho_Q(\mathbf{x}).\end{aligned}$$

■

Lemma 4.8.5. *For any round $t > t_0$, $P_t(\mathbf{x}_{it} \notin \mathcal{S}_{it}) \geqslant q_0(1-\delta_t) - \delta_{2t}$.*

Proof. First, when event B_{it} holds, by lemma 4.8.4, we have that for all $\mathbf{x}$,

$$|u_{it}(\mathbf{x}) - \bar{u}_{it}(\mathbf{x})| \leqslant L_\mu\beta_{2t}\rho_{it}(\mathbf{x}).$$

Note that by definition, $u_{it}(\mathbf{x}) = \mu\left((\theta_{it})^\top\phi(\mathbf{x})\right)$, and $\bar{u}_{it}(\mathbf{x}) = \mu\left((\bar{\theta}_{it})^\top\phi(\mathbf{x})\right)$. Therefore,

$$\bar{u}_{it}(\mathbf{x}) - u_{it}(\mathbf{x}) > -L_\mu\beta_{2t}\rho_{it}(\mathbf{x}). \tag{4.9}$$

On the other side, under event A_{it}, by lemma 4.8.4, we have that for all $\mathbf{x}$,

$$|u_i(\mathbf{x}) - \bar{u}_{it}(\mathbf{x})| \leqslant L_\mu\beta_{1t}\rho_{it}(\mathbf{x}). \tag{4.10}$$

Moreover, recall that by definition for any $\mathbf{x} \in \mathcal{S}_{it}$,

$$u_i(\mathbf{x}^*_{it}) - u_i(\mathbf{x}) \geqslant \beta_{3t}\rho_{it}(\mathbf{x}). \tag{4.11}$$

Therefore, consider any $\mathbf{x} \in \mathcal{S}_{it}$, and under the condition that $A_{it} \cap B_{it} \cap \{u_{it}(\mathbf{x}^*_{it}) > u_i(\mathbf{x}^*_{it})\}$, we have

$$\begin{aligned}u_{it}(\mathbf{x}^*_{it}) - u_{it}(\mathbf{x}) &> u_i(\mathbf{x}^*_{it}) - u_{it}(\mathbf{x}) \\ &= (u_i(\mathbf{x}^*_{it}) - u_i(\mathbf{x})) + (u_i(\mathbf{x}) - \bar{u}_{it}(\mathbf{x})) + (\bar{u}_{it}(\mathbf{x}) - u_{it}(\mathbf{x})) \\ &> 0,\end{aligned} \tag{4.12}$$

where the last inequality follows from combining equations Eq (4.9), Eq (4.10), Eq (4.11) and the definition of β_{3t}. Hence, Eq (4.12) implies that, under the same condition, $\mathbf{x}_{it} \notin \mathcal{S}_{it}$ since by construction, $\mathbf{x}_{it}$ maximizes u_{it} under the budget, thus

$$u_{it}(\mathbf{x}_{it}) \geqslant u_{it}(\mathbf{x}^*_{it}),$$

This further implies that,

$$\begin{aligned}P_t(x_{it} \notin \mathcal{S}_{it}) &\geqslant P_t\left(u_{it}(\mathbf{x}^*_{it}) > u_{it}(x), \forall x \in \mathcal{S}_{it}\right) \\ &\geqslant P_t\left(u_{it}(\mathbf{x}^*_{it}) > u_{it}(x), \forall x \in \mathcal{S}_{it} | A_{it} \cap B_{it}\{u_i(\mathbf{x}^*_{it}) > u_i(\mathbf{x})\}\right)\end{aligned}$$

$$
\begin{aligned}
&\times P(A_{it} \cap B_{it} \cap \{u_i(\mathbf{x}^*_{it}) > u_i(\mathbf{x})\}) \\
&= P(A_{it} \cap B_{it} \cap \{u_i(\mathbf{x}^*_{it}) > u_i(\mathbf{x})\}) \\
&\geqslant P(A_{it} \cap \{u_i(\mathbf{x}^*_{it}) > u_i(\mathbf{x})\}) - P(B^c_{it}) \\
&= P(\{u_i(\mathbf{x}^*_{it}) > u_i(\mathbf{x})\}|A_{it})P(A_{it}) - P(B^c_{it}) \\
&\geqslant q_0(1-\delta_t) + \delta_{2t}.
\end{aligned}
$$

Here, the second and third inequality both from the law of total probability and rearranging terms, and the last inequality follows from Lemma 4.8.1, Lemma 4.8.2 and Lemma 4.8.3, which completes the proof. ■

Lemma 4.8.6. *For $t \geqslant \max(t_0, t'_0)$,*

$$\mathbb{E}_t[\ell_{it}] \leqslant \frac{5}{q_0}\beta_{3t}\mathbb{E}_t[\rho_{it}(\mathbf{x}_{it})] + \delta_t + \delta_{2t}.$$

Here, $\ell_{it} = (u_i(\mathbf{x}^*_{it}) - u_i(\mathbf{x}_{it}))^+$, and t'_0 is chosen such that, $\forall t > t'_0$, $\delta_t < \frac{1}{4}$, $\delta_{2t} < \frac{q_0}{4}$.

Proof. First, define

$$\mathbf{x}'_{it} = \underset{\mathbf{x}: p_t^\top \leqslant p_t^\top e_i, \mathbf{x} \notin \mathcal{S}_{it}}{\arg\min} \rho_{it}(\mathbf{x}).$$

This implies that,

$$\underset{t}{\mathbb{E}}[\rho_{it}(\mathbf{x}_{it})] \geqslant \underset{t}{\mathbb{E}}[\rho_{it}(\mathbf{x}_{it})|\mathbf{x}_{it} \notin \mathcal{S}_{it}]P(\mathbf{x} \notin \mathcal{S}_{it}) \geqslant \rho_{it}(\mathbf{x}'_{it})P(\mathbf{x} \notin \mathcal{S}_{it}).$$

Therefore, by Lemma 4.8.5, we have

$$\rho_{it}(\mathbf{x}'_{it}) \leqslant \frac{\mathbb{E}_t[\rho_{it}(\mathbf{x}_{it})]}{q_0(1-\delta_t) - \delta_{2t}}.$$

Select t'_0 such that, $\forall t \geqslant t'_0$, $\delta_t = \frac{1}{4}$, and $\delta_{2t} \leqslant \frac{q_0}{4}$, then we have:

$$\rho_{it}(\mathbf{x}'_{it}) \leqslant \frac{2}{q_0}\underset{t}{\mathbb{E}}[\rho_{it}(\mathbf{x}_{it})].$$

Also, under $A_{it} \cap B_{it}$,

$$\|\theta^*_i - \theta_{it}\|_{Q_{it}} \leqslant \|\theta^*_i - \bar{\theta}_{it}\|_{Q_{it}} + \|\bar{\theta}_{it} - \theta_{it}\|_{Q_{it}} \leqslant \beta_{1t} + \beta_{2t},$$

where the first inequality follows from triangle inequality, and the second one follows from the definitions of A_{it} and B_{it}. Hence,

$$|\mu_{it}(\mathbf{x}) - \mu_i(\mathbf{x}) \leqslant L_\mu(\beta_{1t} + \beta_{2t})\rho_{it}(\mathbf{x}) = \beta_{3t}\rho_{it}(\mathbf{x}).$$

Therefore, we have

$$\ell_{it} = u_i(\mathbf{x}^*_{it}) - u_i(\mathbf{x}_{it})$$

$$
\begin{aligned}
&= u_i(\mathbf{x}^*_{it}) - u_i(\mathbf{x}'_{it}) + u_i(\mathbf{x}'_{it}) - u_i(\mathbf{x}_{it}) \\
&\leqslant 2\beta_{3t}\rho_{it}(\mathbf{x}'_{it}) + \beta_{3t}\rho_{it}(\mathbf{x}_{it}) \\
&\leqslant \frac{4}{q_0}\beta_{3t}\,\mathbb{E}[\rho_{it}(\mathbf{x}_{it})] + \beta_{3t}\rho_{it}(\mathbf{x}_{it}).
\end{aligned}
$$

which further yields

$$
\begin{aligned}
\mathbb{E}_t[\ell_{it}] &\leqslant \underset{t}{\mathbb{E}}[\ell_{it}|A_{it}\cap B_{it}] + \underset{t}{\mathbb{E}}[\ell_{it}|A^c_{it}\cup B^c_{it}]P(A^c_{it}\cup B^c_{it}) \\
&\leqslant \frac{4}{q_0}\beta_{3t}\,\mathbb{E}[\rho_{it}(\mathbf{x}_{it})] + \beta_{3t}\,\underset{t}{\mathbb{E}}[\rho_{it}(\mathbf{x}_{it})] + \delta_t + \delta_{2t} \\
&\leqslant \frac{5}{q_0}\beta_{3t}\mathbb{E}_t[\rho_{it}(\mathbf{x}_{it})] + \delta_t + \delta_{2t},
\end{aligned}
\tag{4.13}
$$

which completes the proof. ■

Lemma 4.8.7. *Let $\delta' > 0$. Define $L_{iT} = \sum_{t=1}^{T}\ell_{it}$. Then, with probability at least $1-\delta'$,*

$$
L_{iT} \leqslant \sum_{t=1}^{T}(\delta_t + \delta_{2t}) + \tilde{O}\left(\frac{m^2}{\sqrt{M}}\sqrt{T}\left(\log(T) + \log\left(\frac{1}{\delta'}\right)\right)\right)
$$

Proof. First, define for $s > 1$,

$$
u_{is} = \ell_{is} - \frac{5\beta_{3s}}{q_0}\rho_{is}(\mathbf{x}_{is}) - (\delta_s + \delta_{2s}),
$$

and $v_{it} = \sum_{s=1}^{t} u_{is}$, with $v_{i0} = 0$ and $u_{i0} = 0$. We show that $\{v_{it}\}, t \geqslant 0$ is a super-martingale with respect to the filtration $(\mathcal{F}_t)_{t\geqslant 0}$.

First,

$$
\underset{t}{\mathbb{E}}[u_{it}] = \underset{t}{\mathbb{E}}[\ell_{it}] - \frac{5\beta_{3t}}{q_0}\,\underset{t}{\mathbb{E}}[\rho_{it}(\mathbf{x}_{it})] - (\delta_t + \delta_{2t}) \leqslant 0.
$$

Moreover,

$$
\begin{aligned}
|v_{it} - v_{i,t-1}| &\leqslant |\ell_{it}| + \frac{5\beta_{3t}}{q_0}|\rho_{it}(\mathbf{x}_{it})| + (\delta_t + \delta_{2t}) \\
&\leqslant 1 + \frac{5\beta_{3t}}{q_0}\frac{\|\phi(1)\|_2}{\sqrt{M}} + 1 \\
&\leqslant \frac{7\beta_{3t}}{q_0}\frac{\|\phi(\mathbf{1})\|_2}{\sqrt{M}} \triangleq D_t.
\end{aligned}
$$

Therefore, by Lemma 4.7.3, with probability at least $1-\delta'$, we have that

$$
v_{iT} - v_{i,0} \leqslant \sqrt{s\log\left(\frac{1}{\delta'}\sum_{t-1}^{T}D_t^2\right)} \leqslant \frac{7\beta_{3t}}{q_0}\|\phi(1)\|_2\sqrt{\frac{2\log(\frac{1}{\delta'})}{M}T}.
$$

Therefore, we have

$$L_{iT} = \sum_{s=1}^{T} \ell_{is} \leqslant \sum_{t=1}^{T} (\delta_t + \delta_{2t}) + \frac{5\beta_{3t}}{q_0} \sum_{t=1}^{T} \rho_{it}(\mathbf{x}_t) + \frac{7\beta_{3t}}{q_0} \|\phi(\mathbf{1})\|_2 \sqrt{\frac{2\log(\frac{1}{\delta'})}{M} T}. \tag{4.14}$$

Now it remains to bound $\sum_{t=1}^{T} \rho_{it}(\mathbf{x}_t)$. Since $Q_{it} \succeq MI$, by the definition of $\rho_{it}(\mathbf{x}_{it})$, we have

$$\rho_{it} x_{it} \leqslant \frac{\|\phi(\mathbf{1})\|_2}{\sqrt{M}}.$$

Hence, by Lemma 4.7.4 and rearranging terms, we have

$$\sum_{t=t_0}^{T} \rho_{it}^2(\mathbf{x}_{it}) \leqslant \frac{\|\phi(\mathbf{1})\|_2^2}{M} \frac{1}{\log\left(1 + \frac{\|\phi(\mathbf{1})\|_2}{\sqrt{M}}\right)} \sum_{t=t_0}^{T} \log\left(1 + \phi^\top(\mathbf{x}_{it}) Q_{it}^{-1} \phi(\mathbf{x}_{it})\right). \tag{4.15}$$

Also notice that,

$$\sum_{t=t_0}^{T} \log\left(1 + \phi^\top(\mathbf{x}_{it}) Q_{it}^{-1} \phi(\mathbf{x}_{it})\right) = \log \Pi_{t=t_0}^{T} \left(1 + \|\phi(\mathbf{x}_{it})\|_{Q_{it}^{-1}}^2\right) = \log \frac{\det(Q_{iT})}{\det(Q_{i,t_0})}.$$

Note that the trace of $Q_{i,t+1}$ is upper-bounded by $t \cdot \|\phi(\mathbf{1})\|_2$, then given that the trace of the positive definite matrix Q_{iT} is equal to the sum of its eigenvalues, we have that $\det(Q_{iT}) \leqslant \left(t \|\phi(\mathbf{1})\|_2^2\right)^m$. Moreover, $\det(Q_{i,t_0}) \geqslant (M)^m$, therefore,

$$\sum_{t=t_0}^{T} \log\left(1 + \phi^\top(\mathbf{x}_{it}) Q_{it}^{-1} \phi(\mathbf{x}_{it})\right) \leqslant m \log\left(\frac{\|\phi(\mathbf{1})\|_2^2 T}{M}\right).$$

Combining with Eq (4.15), and applying Cauchy-Schwartz inequality, we have

$$\sum_{t=t_0}^{T} \rho_{it}(\mathbf{x}_{it}) \leqslant \sqrt{T \sum_{t=t_0}^{T} \rho_{it}^2(\mathbf{x}_{it})} \leqslant \sqrt{T} \sqrt{\frac{\|\phi(\mathbf{1})\|_2^2}{M} \frac{m}{\log\left(1 + \frac{\|\phi(\mathbf{1})\|_2}{\sqrt{M}}\right)} \log\left(\frac{\|\phi(\mathbf{1})\|_2^2 T}{M}\right)}. \tag{4.16}$$

Putting together Eq (4.14) and Eq (4.16), with the fact that $\|\phi(\mathbf{1})\|_2 = O(\sqrt{m})$, e.g. $\|\phi(\mathbf{1})\|_2 \leqslant c_\phi \sqrt{m}$, we have that with probability at least $1 - \delta'$,

$$\begin{aligned} L_{iT} &\leqslant \sum_{t=1}^{T} (\delta_t + \delta_{2t}) + \frac{\beta_{3t}}{q_0} \frac{\sqrt{T}}{\sqrt{M}} \left(5 c_\phi m \sqrt{\frac{1}{\log\left(1 + \frac{\|\phi(\mathbf{1})\|_2}{m}\right)} \log\left(\frac{\|\phi(\mathbf{1})\|_2^2 T}{M}\right)} + 7 c_\phi \sqrt{m} \sqrt{2\log(\frac{1}{\delta'})}\right) \\ &= \sum_{t=1}^{T} (\delta_t + \delta_{2t}) + \tilde{O}\left(\frac{m^2}{\sqrt{M}} \sqrt{T} \left(\log(T) + \log\left(\frac{1}{\delta'}\right)\right)\right), \end{aligned}$$

where the last step comes from the fact that $\beta_{3t} = \tilde{O}(m)$. This completes the proof. ■

Proof of Theorem 4.4.1 for L^{CE}

Corollary 4.8.8. *With probability at least* $1-\delta'$,

$$L_T^{CE} \leqslant n\sum_{t=1}^{T}(\delta_{1t}+\delta_{2t}) + \tilde{O}\left(n\frac{m^2}{\sqrt{M}}\sqrt{T}\left(\log(T)+\log(\frac{1}{\delta'})\right)\right)$$

Proof. This a direct result from Lemma 4.8.7 and the definition of L_T^{CE}: With probability at least $1-\delta'$,

$$\begin{aligned} L_T^{CE} &= \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\max_{\mathbf{y}:p(\mathbf{y})\leqslant p(e_i)} u_i(\mathbf{y}) - u_i(\mathbf{x}_{it})\right)^+ \\ &\leqslant n\sum_{t=1}^{T}(\delta_t+\delta_{2t}) + \tilde{O}\left(n\frac{m^2}{\sqrt{M}}\sqrt{T}\left(\log(T)+\log\left(\frac{1}{\delta'}\right)\right)\right). \end{aligned}$$

■

Proof. Choose $\delta_t = \delta_{2t} = \frac{2\delta}{n\pi^2 t^2}$. Then,

$$\sum_{t=1}^{T}(\delta_t+\delta_{2t}) \leqslant \frac{2}{3}\delta.$$

Also choose $\delta' = \frac{\delta}{3}$, then by Corollary 4.8.8, with probability at least $1-\delta$,

$$L_T^{CE} = O\left(n\frac{m^2}{\sqrt{M}}\sqrt{T}\left(\log\left(\frac{\delta}{3}\right)+\log(T)\right)\right).$$

■

Proof of Theorem 4.4.2 for L^{CE}

Proof. Choose $\delta_{1t} = \delta_{2t} = \delta' = \frac{1}{T}$. Denote the event where $L_T^{CE} = O\left(n\frac{m^2}{\sqrt{M}}\sqrt{T}\left(\log\left(\frac{\delta}{3}\right)+\log(T)\right)\right)$ holds as $\mathcal{E}$.

Then, by Lemma 4.8.7,

$$\mathbb{E}[L_T^{CE}] = \mathbb{E}[L_T^{CE}|\mathcal{E}] + \mathbb{E}[L_T^{CE}|\mathcal{E}^c]P(\mathcal{E}^c) \leqslant 2 + O\left(n\frac{m^2}{\sqrt{M}}\sqrt{T}\left(\log\left(\frac{\delta}{3}\right)+\log(T)\right)\right).$$

where $M \geqslant m$. This completes the proof. ■

4.9 Appendix: Bounding L^{FD}

Recall that the definition of ℓ^{FD} is directly based on the requirements of Pareto efficiency and fair share: $\ell^{\mathrm{FD}}(x) \stackrel{\text{def}}{=} \max\left(\ell^{\mathrm{PE}}(x), \ell^{\mathrm{SI}}(x)\right)$, where $\ell^{\mathrm{PE}}(x) = \min_{x' \in \mathcal{PE}} \sum_{i=1}^{n} \left(u_i(\mathbf{x}'_i) - u_i(\mathbf{x}_i)\right)^+$; and $\ell^{\mathrm{SI}}(x) = \sum_{i=1}^{n} \left(u_i(e_i) - u_i(\mathbf{x}_i)\right)^+$.

To bound L^{FD}, we first provide a useful lemma which shows that ℓ^{SI} is a weaker notion than ℓ^{CE}.

Lemma 4.9.1. *For any allocation x and price p, $\ell^{\mathrm{SI}}(x) \leqslant \ell^{\mathrm{CE}}(x, p)$.*

Proof. This simply uses the fact that an agent's endowment is always affordable under any price vector p. Therefore,

$$\begin{aligned}
\ell^{\mathrm{SI}}(x) &= \sum_{i=1}^{n} \left(u_i(e_i) - u_i(\mathbf{x}_i)\right)^+ \\
&\leqslant \sum_{i=1}^{n} \left(\max_{y : p^\top y \leqslant p^\top e_i} u_i(y) - u_i(x) \right)^+, \forall p \\
&= \ell^{\mathrm{CE}}(x, p),
\end{aligned}$$

■

Having lemma 4.9.1 at hand, the key remaining task is to bound ℓ^{PE}. We will show that this can be achieved by an analogous analysis as in Section 4.8, but with some key differences.

First, we define $\tilde{\mathcal{S}}_{it}$ (in comparison to $\mathcal{S}_{it}$ used in Section 4.8):

$$\tilde{\mathcal{S}}_{it} \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathcal{X} : u_i(\mathbf{x}^*_i) - u_i(\mathbf{x}) \geqslant \beta_{3t} \rho_{it}(\mathbf{x}) \right\},$$

where $x^* \in \mathbb{R}^{n \times m}$ is the unique equilibrium allocation. Note that $\tilde{\mathcal{S}}_{it}$ shares a similar spirit as $\mathcal{S}_{it}$, which is used in Section 4.8, but with a different referencing point x^*.

We show a key lemma which provides a lower bound on $P(x \notin \tilde{\mathcal{S}}_{it})$.

Lemma 4.9.2. *For any round $t > t_0$, $P_t(\mathbf{x}_{it} \notin \tilde{\mathcal{S}}_{it}) \geqslant q_0(1 - \delta_t) - \delta_{2t}$.*

Proof. First, when event B_{it} holds, by lemma 4.8.4, we have that for all $\mathbf{x}$,

$$|u_{it}(\mathbf{x}) - \bar{u}_{it}(\mathbf{x})| \leqslant L_\mu \beta_{2t} \rho_{it}(\mathbf{x}).$$

Note that by definition, $u_{it}(\mathbf{x}) = \mu\left((\theta_{it})^\top \phi(\mathbf{x})\right)$, and $\bar{u}_{it}(\mathbf{x}) = \mu\left((\bar{\theta}_{it})^\top \phi(\mathbf{x})\right)$. Therefore,

$$\bar{u}_{it}(\mathbf{x}) - u_{it}(\mathbf{x}) > -L_\mu \beta_{2t} \rho_{it}(\mathbf{x}). \tag{4.17}$$

On the other side, under event A_{it}, by lemma 4.8.4, we have that for all $\mathbf{x}$,

$$|u_i(\mathbf{x}) - \bar{u}_{it}(\mathbf{x})| \leqslant L_\mu \beta_{1t} \rho_{it}(\mathbf{x}). \tag{4.18}$$

Moreover, recall that by definition for any $\mathbf{x} \in \tilde{\mathcal{S}}_{it}$,

$$u_i(\mathbf{x}_i^*) - u_i(\mathbf{x}) \geqslant \beta_{3t}\rho_{it}(\mathbf{x}). \tag{4.19}$$

Therefore, consider any $\mathbf{x} \in \tilde{\mathcal{S}}_{it}$, and under the condition that $A_{it} \cap B_{it} \cap \{u_{it}(\mathbf{x}_i^*) > u_i(\mathbf{x}_i^*)\}$, we have

$$\begin{aligned} u_{it}(\mathbf{x}_i^*) - u_{it}(\mathbf{x}) &> u_i(\mathbf{x}_i^*) - u_{it}(\mathbf{x}) \\ &= (u_i(\mathbf{x}_i^*) - u_i(\mathbf{x})) + (u_i(\mathbf{x}) - \bar{u}_{it}(\mathbf{x})) + (\bar{u}_{it}(\mathbf{x}) - u_{it}(\mathbf{x})) \\ &> 0, \end{aligned} \tag{4.20}$$

where the last inequality follows from combining equations Eq (4.17), Eq (4.18), Eq (4.19) and the definition of β_{3t}. Moreover, recall that $\mathbf{x}_{it}$ maximizes u_{it} under the budget, thus

$$u_{it}(\mathbf{x}_{it}) \geqslant u_{it}(\mathbf{x}_i^*),$$

Therefore, Eq (4.20) implies that, $\mathbf{x}_{it} \notin \tilde{\mathcal{S}}_{it}$. This further implies,

$$\begin{aligned} P_t(x_{it} \notin \tilde{\mathcal{S}}_{it}) &\geqslant P_t\left(u_{it}(\mathbf{x}_i^*) > u_{it}(x), \forall x \in \tilde{\mathcal{S}}_{it}\right) \\ &\geqslant P_t\left(u_{it}(\mathbf{x}_i^*) > u_{it}(x), \forall x \in \tilde{\mathcal{S}}_{it} | A_{it} \cap B_{it}\{u_i(\mathbf{x}_{it}^*) > u_i(\mathbf{x})\}\right) \\ &\quad \cdot P(A_{it} \cap B_{it} \cap \{u_i(\mathbf{x}_{it}^*) > u_i(\mathbf{x})\}) \\ &= P(A_{it} \cap B_{it} \cap \{u_i(\mathbf{x}_i^*) > u_i(\mathbf{x})\}) \\ &\geqslant P(A_{it} \cap \{u_i(\mathbf{x}_i^*) > u_i(\mathbf{x})\}) - P(B_{it}^c) \\ &= P(\{u_i(\mathbf{x}_i^*) > u_i(\mathbf{x})\}|A_{it})P(A_{it}) - P(B_{it}^c) \\ &\geqslant q_0(1 - \delta_t) + \delta_{2t}. \end{aligned}$$

Here, the second and third inequality both from the law of total probability and rearranging terms, and the last inequality follows from Lemma 4.8.1, Lemma 4.8.2 and Lemma 4.8.3, which completes the proof. ■

Lemma 4.9.3. *At any round $t > t_0$, define $\mathbf{x}''_{it} \stackrel{\text{def}}{=} \arg\min_{x \notin \tilde{\mathcal{S}}_{it}, \mathbf{p}_t^\top \mathbf{y} \leqslant \mathbf{p}_t^\top e_i} \rho_{it}(x_{it})$, then we have*

$$\ell^{\mathrm{PE}}(x_t) \leqslant \sum_{i \in [n]} b_{it}^{\mathrm{PE}},$$

with probability at least $1 - \delta_t - \delta_{2t}$, and $b_{it}^{\mathrm{PE}} = 2\beta_{3t}\rho_{it}(x''_{it}) + \beta_{3t}\rho_{it}(x_{it})$.

Proof. Begin with the definition of ℓ^{PE}, we have

$$\ell^{\mathrm{PE}}(x_t) = \min_{x \in \mathcal{PE}} \sum_{i \in [n]} (u_i(x_i) - u_i(x_{it}))$$

$$\begin{aligned}
&\leqslant \sum_{i\in[n]} \left(u_i(x_i^*) - u_i(x_{it})\right) \\
&= \sum_{i\in[n]} \left(u_i(x_i^*) - u_i(x_{it}'')\right) + \sum_{i\in[n]} \left(u_i(x_i'') - u_i(x_{it})\right) \\
&\leqslant \beta_{3t} \sum_{i\in[n]} \rho_{it}(x_{it}'') + \sum_{i\in[n]} \left(u_i(x_i'') - u_i(x_{it})\right),
\end{aligned}$$

where the last inequality follows from this definition. Moreover, we have

$$u_i(x_{it}'') \leqslant u_{it}(x_{it}'') + \beta_{3t}\rho_{it}(x_{it}''),$$

and

$$u_i(x_{it}) \geqslant u_{it}(x_{it}) - \beta_{3t}\rho_{it}(x_{it}).$$

under the event $A_{it} \cap B_{it}$, by Eq (4.17) and Eq (4.18). Putting these together yields

$$\ell^{PE}(x_t) \leqslant 2\beta_{3t} \sum_{i\in[n]} \rho_{it}(x_{it}'') + \beta_{3t} \sum_{i\in[n]} \rho_{it}(x_{it}) \stackrel{\text{def}}{=} \sum_{i\in[n]} b_{it}^{PE},$$

which completes the proof. ■

Now we show that the above result leads to the lemma below, which shows a analogous guarantee as we obtained in lemma 4.8.6.

Lemma 4.9.4. *For $t \geqslant \max(t_0, t_0')$,*

$$\mathbb{E}_t[\sum_{i\in[n]} \ell_{it}^{FD}] \leqslant \frac{5}{q_0}\beta_{3t}\mathbb{E}_t[\sum_{i\in[n]} \rho_{it}(\mathbf{x}_{it})] + n(\delta_t + \delta_{2t}).$$

Here, t_0' is chosen such that, $\forall t > t_0'$, $\delta_t < \frac{1}{4}$, $\delta_{2t} < \frac{q_0}{4}$.

Proof. First, note that, by the definition of $\mathbf{x}_{it}''$,

$$\mathop{\mathbb{E}}_t[\rho_t(\mathbf{x}_{it})] \geqslant \mathop{\mathbb{E}}_t[\rho_t(\mathbf{x}_{it})|\mathbf{x}_{it} \notin \mathcal{S}_{it}]P(\mathbf{x} \notin \mathcal{S}_{it}) \geqslant \rho_{it}(\mathbf{x}_{it}'')P(\mathbf{x} \notin \mathcal{S}_{it}).$$

Moreover, combining the above with Lemma 4.9.2, we have

$$\rho_{it}(\mathbf{x}_{it}'') \leqslant \frac{\mathbb{E}_t[\rho_{it}(\mathbf{x}_{it})]}{q_0(1-\delta_t) - \delta_{2t}}.$$

Select t_0' such that, $\forall t \geqslant t_0'$, $\delta_t = \frac{1}{4}$, and $\delta_{2t} \leqslant \frac{q_0}{4}$, then we have:

$$\rho_{it}(\mathbf{x}_{it}'') \leqslant \frac{2}{q_0} \mathop{\mathbb{E}}_t[\rho_{it}(\mathbf{x}_{it})].$$

Also, under $A_{it} \cap B_{it}$,

$$\|\theta_i^* - \theta_{it}\|_{Q_{it}} \leqslant \|\theta_i^* - \bar{\theta}_{it}\|_{Q_{it}} + \|\bar{\theta}_{it} - \theta_{it}\|_{Q_{it}} \leqslant \beta_{1t} + \beta_{2t},$$

where the first inequality follows from triangle inequality, and the second one follows from the definitions of A_{it} and B_{it}. Hence,

$$|\mu_{it}(\mathbf{x}) - \mu_i(\mathbf{x}) \leqslant L_\mu(\beta_{1t} + \beta_{2t})\rho_{it}(\mathbf{x}) = \beta_{3t}\rho_{it}(\mathbf{x}).$$

Therefore, we have

$$\begin{aligned}\sum_{i\in[n]} b_{it}^{\mathrm{PE}} &\leqslant 2\beta_{3t}\sum_{i\in[n]}\rho_{it}(\mathbf{x}''_{it}) + \beta_{3t}\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it}) \\ &\leqslant \frac{4}{q_0}\beta_{3t}\,\mathbb{E}[\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it})] + \beta_{3t}\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it}).\end{aligned}$$

Moreover, by Lemma 4.9.1 and eq (4.13) which holds under the same condition of $A_{it}\cap B_{it}$, we have

$$\begin{aligned}\sum_{i\in[n]} \ell_{it}^{\mathrm{FD}} &\leqslant \sum_{i\in[n]}\max\{\ell_{it}^{\mathrm{PE}}, \ell_{it}^{\mathrm{SI}}\} \\ &\leqslant 2\beta_{3t}\sum_{i\in[n]}\rho_{it}(\mathbf{x}''_{it}) + \beta_{3t}\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it}) \\ &\leqslant \frac{4}{q_0}\beta_{3t}\,\mathbb{E}[\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it})] + \beta_{3t}\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it}).\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathbb{E}_t[\sum_{i\in[n]}\ell_{it}^{\mathrm{FD}}] \leqslant \mathbb{E}_t[\sum_{i\in[n]} b_{it}^{\mathrm{PE}}] &\leqslant \frac{4}{q_0}\beta_{3t}\,\mathbb{E}[\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it})] + \beta_{3t}\,\mathbb{E}_t[\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it})] + n(\delta_t + \delta_{2t}) \\ &\leqslant \frac{5}{q_0}\beta_{3t}\mathbb{E}_t[\sum_{i\in[n]}\rho_{it}(\mathbf{x}_{it})] + n(\delta_t + \delta_{2t}),\end{aligned}$$

which completes the proof. ■

With the above lemmas at hand, we are now ready to provide a proof of Theorem 4.4.1 for L_T^{FD}.

Proof of Theorem 4.4.1 for L^{FD}

Proof. Lemma 4.9.4 shows a analog guarantee as we obtained in lemma 4.8.6 for the L^{FD} loss function. Therefore, following the same steps in lemma 4.8.7, we have that with probability

at least $1-\delta'$ where δ' will be specified momentarily,

$$\begin{aligned} L_T^{\mathrm{FD}} &= \sum_{t=t_0}^{T}\sum_{i=1}^{n} \ell_{it}^{\mathrm{FD}} \\ &\leqslant n\sum_{t=1}^{T}(\delta_t+\delta_{2t}) + \frac{n\beta_{3t}}{q_0}\frac{\sqrt{T}}{\sqrt{M}}\Bigg(5c_\phi m\sqrt{\frac{1}{\log\left(1+\frac{\|\phi(\mathbf{1})\|_2}{\sqrt{M}}\right)}\log\left(\frac{\|\phi(\mathbf{1})\|_2^2 T}{M}\right)} \\ &\quad + 7c_\phi^2\sqrt{m}\sqrt{2\log(\frac{1}{\delta'})}\Bigg) \\ &= n\sum_{t=1}^{T}(\delta_t+\delta_{2t}) + \tilde{O}\left(n\frac{m^2}{\sqrt{M}}\sqrt{T}\left(\log(T)+\log\left(\frac{1}{\delta'}\right)\right)\right), \end{aligned} \tag{4.21}$$

Choose $\delta_t = \delta_{2t} = \frac{2\delta}{n\pi^2 t^2}$. Then,

$$\sum_{t=1}^{T}(\delta_t+\delta_{2t}) \leqslant \frac{2}{3}\delta.$$

Also choose $\delta' = \frac{\delta}{3}$, then by Eq (4.21), with probability at least $1-\delta$,

$$L_T^{\mathrm{FD}} = O\left(\frac{nm^2}{\sqrt{M}}\sqrt{T}\left(\log\left(\frac{\delta}{3}\right)+\log(T)\right)\right).$$

■

Proof of Theorem 4.4.2 for L^{FD}

Proof. The theorem results follow from eq (4.21). Choose $\delta_{1t} = \delta_{2t} = \delta' = \frac{1}{T}$ and denote the event where $L_T^{\mathrm{FD}} = O\left(\frac{nm^2}{\sqrt{M}}\sqrt{T}\left(\log\left(\frac{\delta}{3}\right)+\log(T)\right)\right)$ holds as $\mathcal{E}$. Then,

$$\mathbb{E}[L_T^{\mathrm{FD}}] = \mathbb{E}[L_T^{\mathrm{PE}}|\mathcal{E}] + \mathbb{E}[L_T^{\mathrm{FD}}|\mathcal{E}^c]P(\mathcal{E}^c) \leqslant O\left(\frac{nm^2}{\sqrt{M}}\sqrt{T}\left(\log\left(\frac{\delta}{3}\right)+\log(T)\right)\right),$$

where $M \geqslant m$. This completes the proof. ■

4.10 Appendix: Additional experimental details and results

In this section, we present the simulation results with the Amdahl utilities, as described in Section 3.7, and additional implementation details.

Figure 4.2 presents the simulation results for agents with the Amdahl's utilities.

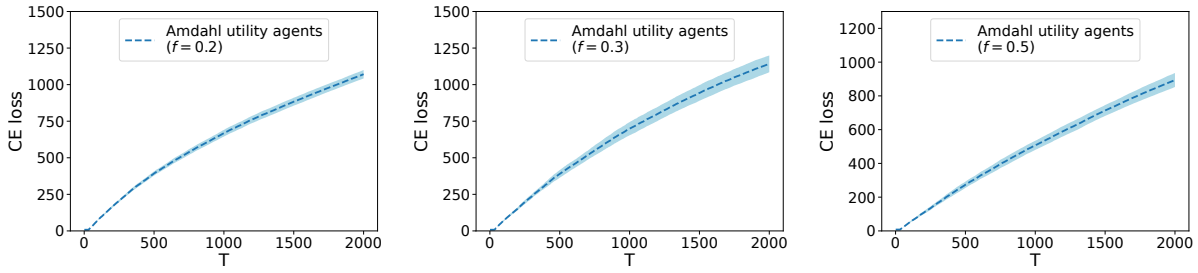


Figure 4.2: The CE loss L_T^{CE} vs the number of rounds T. evaluated with $m = 2$ resource types and $n = 8$ agents and Amdahl's utilities. The three figures correspond to $f_i = 0.2$, $f_i = 0.3$, and $f_i = 0.5$ (as in Section 3.7). All figures show results which are averaged over 10 runs, and the shaded region shows the standard error at each time T.

Empirically, we compute L^{CE} by maximizing each agent's utility subject to the budget constraint. We approximate this by randomly sampling feasible allocations y from a simplex, accept those that cost no more price than the agent's endowment, lastly take the maximum. We sampled up to 50 accepted samples in each round. All experiments are run on a AWS EC2 p3.2xlarge machine.

4.11 Appendix: Further discussions

Further Background on Fair Division and Exchange Economies

Since the seminal work of Varian [207], fair division of multiple resource types has received significant attention in the game theory, economics, and computer systems literature. We provide more background on the related works in the fair division literature and their applications in this section.

Among the theoretical works in fair division, one of the most common perspectives on this problem is as an exchange economy (or as a Fisher market, which is a special case of an EE) [18, 19, 40, 58, 107, 160, 204, 207].

Fair allocation mechanisms have been deployed in many practical resource allocation tasks when compute resources are shared by multiple users [224]. There have also been applications of other market-based resource allocation schemes for data centers and power grids [42, 113, 143, 209, 219]. One line of work in this setting studied fair division when the resources in question are perfect complements; some examples include dominant resource fairness and its variants [68, 91, 92, 107, 152, 178]. Although the assumption of perfect complement resource types leads to computationally simple mechanisms, in many practical applications, there is ample substitutability between resources, and hence the above mechanisms can be inappropriate. For example, in compute clusters, CPUs and GPUs are often interchangeable for many jobs, albeit with different performance characteristics. Indeed, in this work, we

in particular focused on the applications of EE and fair division mechanisms for computing resource allocation.

Computation of the FD loss

We note that one main challenge for computing the FD loss is that we need to approximate the Pareto Front and then take a minimum over it. To approximate it, even in the simplest two agent two resource set up, this requires grid search on a 4D space which can be computationally prohibitive. In our experiments, the dimensions are 15 (3 resources by 5 agents) and 16 (2 resources by 8 agents), for which grid search is not feasible. Given that efficient computations over the Pareto frontier have remain a technical challenge, we focused the evaluations on the CE loss in this work.

On the loss functions

We provide an example to demonstrate that the FD loss (4.3), while more interpretable than the CE loss (4.2), may not capture all properties of an equilibrium.

For this consider the following example with $n = 3$ agents and $m = 2$ resources where the endowments of agent 1, agent 2, and agent 3 are $e_1 = (0.45, 0.05)$, $e_2 = (0.45, 0.05)$, and $e_3 = (0.1, 0.9)$ respectively. Their utilities are:

$$u_1(x_1) = 0.1x_{11} + x_{12}, \qquad u_2(x_2) = 0.2x_{21} + x_{22}, \qquad u_3(x_3) = x_{31} + 0.1x_{32}.$$

The utilities of the three users if they were to simply use their endowment is, $u_1(e_1) = 0.1 \times 0.45 + 0.05 = 0.095$, $u_2(e_2) = 0.14$, and $u_3(e_3) = 0.19$. We find that while agents 1 and 2 benefit more from the second resource, they have more of the first resource in their endowments and vice versa for agent 3. By exchanging resources, we can obtain a more efficient allocation.

The unique equilibrium prices for the two goods are $p^\star = (1/2, 1/2)$ and the allocations are $x_1^\star = (0, 0.5)$ for agent 1, $x_2^\star = (0, 0.5)$ for agent 2, and $x_3^\star = (1.0, 0.0)$ for agent 3. The utilities of the agents under the equilibrium allocations are $u_1(x_1) = 0.5$, $u_2(x_2) = 0.5$, and $u_3(x_3) = 1.0$. Here, we find that by the definition of CE, $\ell^{\mathrm{PE}}(x^\star, p^\star) = 0$. It can also be verified that $\ell^{\mathrm{FD}}(x^\star, p^\star) = 0$.

In contrast, consider the following allocation for the 3 users: $x_1 = (0.35, 0.49)$ for agent 1, $x_2 = (0.35, 0.49)$ for agent 2, and $x_3 = (0.3, 0.02)$ for agent 3. Here, the utilities are $u_1(x_1) = 0.1 \times 0.35 + 0.49 = 0.525$, $u_2(x_2) = 0.56$, and $u_3(x_3) = 0.3002$. This allocation is both PE (as the utility of one user can only be increased by taking resources from someone else), and SI (as all three users are better off than having their endowments). Therefore, $\ell^{\mathrm{FD}}((x_1, x_2, x_3)) = 0$. However, user 3 might complain that their contribution of resource 2 (which was useful for users 1 and 2) has not been properly accounted for in the allocation. Specifically, there do not exist a set of prices p for which $\ell^{\mathrm{PE}}(x, p) = 0$. This example illustrates the role of prices in this economy: it allows us to value the resources relative to each other based on the demand.

Chapter 5

No-Regret Learning in Partially-Informed Auctions

5.1 Introduction

Selling mechanisms play a crucial role in economic theory and have a wide range of applications across many industries [11, 79, 163, 164, 181]. Under the canonical mechanism design model, buyers choose whether or not to buy items for sale based on their true values for those items. This fundamental model, however, assumes that the buyers know exactly how much they value the items for sale, which is often not the case.

One of the overriding reasons that a buyer may not know their true values is information asymmetry: the seller may purposefully obfuscate information about an item for sale. For example, the seller may hide information about the item in the hopes of better revenue [90]. Alternatively, information about the item may be private, and thus the seller may wish to protect this sensitive information by only revealing partial information about the item. For instance, in online advertising auctions, bids represent how much advertisers are willing to pay to display their ad to a particular user. Historically, advertisers have bid based on uniquely identifying information about users, but there has been a growing effort to protect users' privacy by obfuscating this sensitive information [80, 99, 123].

In these scenarios, the buyer only has partial information about the item for sale but still must decide whether to make a purchase. This raises the question: **how should a buyer determine their purchase strategy with only incomplete item information?**

We study posted-price auctions—a fundamental mechanism family that is appealingly interpretable—with incomplete item information. In particular, the seller reveals obfuscated ("masked") information about the item using a fixed, unknown masking function. We study an online setting where, at each round, a fresh item is drawn from an unknown distribution (for example, a distribution over users visiting a webpage). The seller sets a price and the buyer chooses whether to buy the item based on the incomplete information that the seller provides. We propose no-regret learning algorithms for the buyer that achieve sub-linear

Item distribution	**Prices**	**Masking function h**	**Regret**
Known	Adversarial	SimHash $h : [0,1]^d \to \{0,1\}^\ell$	$\mathcal{O}(\sqrt{Td\ell \log(T\ell/\delta)})$ (Theorem 5.3.5)
Unknown	Stochastic	Arbitrary $h : \mathcal{X} \to [n]$	$\mathcal{O}(\sqrt{T(n \log T/n + \log 1/\delta)})$ (Theorem 5.4.3)
Unknown	Adversarial	Arbitrary $h : \mathcal{X} \to [n]$	$\tilde{\mathcal{O}}(T^{2/3}n^{1/3})$ (Remark 5.4.5)

Table 5.1: Summary of regret bounds which hold with probability at least $1 - \delta$.

regret compared to an oracle buyer who has perfect knowledge of the item distribution as well as the seller's masking function.

Our results

We study no-regret learning with incomplete item information in two settings:

1. First, we propose an algorithm for a setting where the item distribution is known to the buyer and the mask is a SimHash function mapping $[0,1]^d$ to $\{0,1\}^\ell$. In other words, each item is defined by d real-valued features and the seller reveals ℓ bits about the item to the buyer, as defined by a function that is unknown to the buyer. This model has been studied from an applied perspective in the context of ad auctions [80]. We provide an algorithm with regret $\mathcal{O}(\sqrt{Td\ell \log(T\ell/\delta)})$.
2. Next, we study a setting where the masking function is an arbitrary mapping from the set of all items, denoted $\mathcal{X}$, to a finite set of size n. We propose an online learning algorithm with regret $\mathcal{O}(\sqrt{T(n \log T/n + \log 1/\delta)})$ when the prices are stochastic, where T is the length of the horizon.

In the first setting where the masking function is a SimHash function mapping $[0,1]^d$ to $\{0,1\}^\ell$, the domain of the masking function is of size $n = 2^\ell$, so our regret bound of $\tilde{\mathcal{O}}((Td\ell)^{1/2})$ is exponentially better than the latter regret bound.

We summarize these results in Table 5.1.

Related work

This work draws on several threads of research on designing auctions with incomplete information, learning to bid, and privacy-preserving simple auctions.

Auction design with incomplete information. Auctions with incomplete value information have attracted much research attention. Several prior works have explored auction design where the information about the item may be incomplete to the buyer or the seller [10, 31, 32, 82, 88, 149, 150, 185]. In particular, Ganuza [88] studied the incentives of the auctioneer to release signals about the item to the buyers that refine their private

valuations before a second-price auction. In their model, the seller reveals a noisy item feature vector which is an unbiased estimator of the true one. Bergemann and Pesendorfer [31] considered single-item multi-bidder auctions where the seller decides how accurately the bidders can learn their valuations. However, all of this prior work is limited to offline settings; they did not explore online purchase strategies that the buyer can adopt. To the best of our knowledge, this work constitutes the first analysis of no-regret learning algorithms in partially-informed posted-price auctions.

Learning to bid with an unknown value. Instead of focusing the problem from the side of the auctioneer who aims to maximize revenue over repeated rounds, another active line of research studies bidding strategies for the bidders when they do not know their values [23, 66, 83, 218]. Feng et al. [83] considered a single-item multi-bidder setting where the bidder learns to bid via partial feedback, and provide algorithms with regret rates against the best fixed bid in hindsight. Dikkala and Tardos [66] explored a setting where bidders need to experiment in order to learn their valuations. The key difference between this work and ours is that our algorithms exploit the available "partial" information instead of having zero knowledge about the item being sold. This partial information model allows us to trade off between the amount of information revealed about the item and the regret. Moreover, we compete with the best purchase policy, rather than the best fixed bid in hindsight.

Private auctions. Our theoretical model is motivated by recent work on designing privacy-enhanced auctions for practical usage. An important application of these auctions is online advertising, where the items being auctioned are user queries and the auctioner must trade off between user privacy and revenue maximization [62, 80, 99, 123, 183]. Epasto et al. [80] present a detailed exploration of Chrome's Federated Learning of Cohorts (FLoC) API, where user information is masked. Our contribution is to provide a formal treatment of such auctions, including providing algorithms that have theoretical guarantees in the setting of arbitrary masking functions that are unknown to the buyer.

5.2 Preliminaries

We begin by defining formally the setting we study for auctions with partial information.

Setup

We consider a setting where there is a single seller and a single buyer. There is a distribution $\mathcal{P}$ over items which are elements of an abstract set $\mathcal{X}$. The buyer has a bounded valuation, $v^*(\mathbf{x}) \in [0, H]$ for every item $\mathbf{x} \in \mathcal{X}$ and some $H \in \mathbb{R}_+$[1]. The seller and buyer interact over

[1]The assumption that there is a bound on the maximum amount that the agents value the item is widely made in prior research. In our main application—ad auctions—the value of an impression is typically very cheap, so this assumption is mild.

a series of T rounds. At each round $t \in [T]$, the seller draws an item $\mathbf{x}_t \overset{iid}{\sim} \mathcal{P}$ and sets a price $p_t(\mathbf{x}_t) \in [0, H]$ for the item. The seller does not reveal $\mathbf{x}_t$ to the buyer, but rather reveals some partial information $h(\mathbf{x}_t) \in \mathcal{Y}$ where $h : \mathcal{X} \to \mathcal{Y}$ is a fixed "masking" function that maps to an abstract set $\mathcal{Y}$.

We assume that the price function p_t does not give any more information about $\mathbf{x}_t$ than that is provided by the masking function h. In other words, if $h(\mathbf{x}) = h(\mathbf{x}')$, then $p_t(\mathbf{x}) = p_t(\mathbf{x}')$, which implies

$$\mathbb{E}[v^*(\mathbf{x}) \mid h(\mathbf{x}), p(\mathbf{x})] = \mathbb{E}[v^*(\mathbf{x}) \mid h(\mathbf{x})]. \tag{5.1}$$

The buyer uses a strategy $s_t : \mathcal{Y} \times \mathbb{R} \to \{0, 1\}$ to decide whether to buy the item given the partially revealed item information and the price. Here $s_t(h(\mathbf{x}_t), p_t(\mathbf{x}_t)) = 1$ if and only if the buyer buys the item. Letting $b_t = s_t(h(\mathbf{x}_t), p_t(\mathbf{x}_t))$, the buyer's utility is $u_t = b_t(v^*(\mathbf{x}_t) - p_t(\mathbf{x}_t)) \in [-H, H]$. We summarize this process below.

Procedure 2 Online model

for $t = 1, 2, \ldots, T$ **do**
  Seller selects a price function $p_t : \mathcal{X} \to [0, H]$.
  Item $\mathbf{x}_t$ is sampled from $\mathcal{P}$.
  Seller publishes item information $h(\mathbf{x}_t)$ and a price $p_t(\mathbf{x}_t)$.
  Buyer decides whether or not to buy: $b_t = s_t(h(\mathbf{x}_t), p_t(\mathbf{x}_t)) \in \{0, 1\}$.
  Buyer obtains reward $u_t = (v^*(\mathbf{x}_t) - p_t(\mathbf{x}_t)) \cdot b_t$.
  if b_t **then** // item is purchased
    Buyer observes $\mathbf{x}_t$.
  end if
end for

Example 5.2.1 (Advertising auctions)**.** We instantiate our model in the context of advertising auctions, taking inspiration from Epasto et al. [80]. The seller is a platform and the buyer is an advertiser. Each item $\mathbf{x} \in \mathcal{X}$ describes a user who visits the platform. For example, $\mathcal{X} = \mathbb{R}^d$ might denote features that uniquely identify each user. On round t, the advertiser has a value $v^*(\mathbf{x}_t)$ for the opportunity to show the user $\mathbf{x}_t$ an ad. In order to preserve user privacy, the platform does not reveal $\mathbf{x}_t$ to the advertiser, but rather some summary $h(\mathbf{x}_t)$. For example, Epasto et al. [80] study a setting where h is a SimHash function, so $\mathbf{x}_t \in \mathbb{R}^d$ and $h(\mathbf{x}_t) \in \mathbb{R}^\ell$ for some $\ell < d$. The platform sets a price $p_t(\mathbf{x}_t)$ which the advertiser pays to show the user an ad.

We study this problem from the perspective of the buyer: how should they select the strategy s_t at each round to maximize their utility? We study two settings: a setting where the distribution $\mathcal{P}$ is unknown to the buyer and the prices are stochastic (Section 5.4) and a model where $\mathcal{P}$ is known to the buyer with adversarial prices (Section 5.3).

Regret and the optimal strategy

We measure the regret of the buyer in our online model with regard to the optimal strategy s^* of a myopic buyer[2] who has perfect knowledge of the distribution $\mathcal{P}$ and the masking function h, but not the realized item $\mathbf{x}_t$. To make this dependence on the environment clear, in any single round, we use the notation $s^*(h(\mathbf{x}), p(\mathbf{x}), h, \mathcal{P})$ to denote the optimal strategy (we drop the subscript t for simplicity). More formally, s^* maximizes the expected utility:

$$\arg\max_{s \in \mathcal{S}} \mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}}[(v^*(\mathbf{x}) - p(\mathbf{x}))s(h(\mathbf{x}), p(\mathbf{x}), h, \mathcal{P}) \mid h(\mathbf{x})],$$

where $\mathcal{S}$ represents the set of all decision functions $s(\cdot) : \mathcal{Y} \times \mathbb{R} \times \{h(\cdot), \mathcal{P}\} \to \{0, 1\}$.

Definition 5.2.2. The buyer's (expected) regret with respect to the optimal strategy s^*, denoted R_T, is defined as

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(v^*(\mathbf{x}_t) - p_t(\mathbf{x}_t)\right)s^*\left(h(\mathbf{x}_t), p_t(\mathbf{x}_t), h(\cdot), \mathcal{P}\right) - \left(v^*(\mathbf{x}_t) - p_t(\mathbf{x}_t)\right)s_t\left(h(\mathbf{x}_t), p_t(\mathbf{x}_t)\right)\right]. \tag{5.2}$$

In the following proposition, we identify the form of the optimal strategy s^*. The proof is in Appendix 5.6.

Proposition 5.2.3. *The strategy s^* that maximizes $\mathbb{E}_{\mathbf{x} \sim \mathcal{P}}[(v^*(\mathbf{x}) - p(\mathbf{x}))s^*(h(\mathbf{x}), p(\mathbf{x}), h, \mathcal{P}) \mid h(\mathbf{x})]$ is*

$$s^*(h(\mathbf{x}), p(\mathbf{x}), h, \mathcal{P}) = \mathbb{1}\left(\mathop{\mathbb{E}}_{x \sim \mathcal{P}}[v^*(\mathbf{x}) \mid h(\mathbf{x})] > p(\mathbf{x})\right).$$

Even when the buyer has no information about the distribution $\mathcal{P}$ (Section 5.4), we show that he can guarantee low regret with respect to s^* with either stochastic or adversarial prices, in polynomial per-round runtime. When the distribution $\mathcal{P}$ is known (Section 5.3), we provide an algorithm with exponentially better regret.

5.3 Known item distribution

First, we focus on a specific class of masking functions, SimHash, motivated by recent practical applications in ad auctions [80]. Here, $\mathcal{X}$ is a feature space $[0, 1]^d$ and the masking function h is a SimHash function that is unknown to the buyer. In other words, there are ℓ unknown vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_\ell \in \mathbb{R}^d$ such that the masking function, denoted as $h_{\mathbf{w}}$ with $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_\ell)$, is $h_{\mathbf{w}}(\boldsymbol{x}) = (\text{sgn}(\boldsymbol{w}_1 \cdot \boldsymbol{x}), \ldots, \text{sgn}(\boldsymbol{w}_\ell \cdot \boldsymbol{x}))^\top$.

We consider the setting where the distribution of the items is known to the buyer (for example, via historical data). We provide an algorithm that achieves a regret $\tilde{\mathcal{O}}(\sqrt{Td\ell})$

[2] A myopic buyer optimizes his utility separately in each round.

even under adversarial prices. Since the masking function maps to a set of size $n = 2^\ell$, the regret only depends logarithmically on n. As we detail in the subsequent Section 5.4, this algorithm achieves exponentially better regret compared to the algorithm we present where the distribution $\mathcal{P}$ over items is unknown and the masking function is arbitrary.

Algorithm 8 Explore-then-Commit (Known Distribution)

Input: horizon T, distribution $\mathcal{P}$, $d, \ell \in \mathbb{N}_+$, and $\delta \in (0,1)$.
Compute $t' = \sqrt{4Td\ell \log(\ell/\delta)}$.
for $t = 1, 2, \ldots, t'$ **do** // Exploration phase
  Receive $h(\mathbf{x}_t)$, price $p_t(\mathbf{x}_t)$ where $\mathbf{x}_t \overset{iid}{\sim} \mathcal{P}$.
  Make decision $b_t = 1$ and observe $\mathbf{x}_t$.
end for
Use linear programming to compute $\hat{\mathbf{w}} = (\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_\ell)$ such that $h_{\hat{\mathbf{w}}}(\mathbf{x}_i) = h_{\mathbf{w}}(\mathbf{x}_i)$ for all $i \in [t']$.
for $t = t'+1, t'+2, \ldots, T$ **do** // Exploitation phase
  Receive $h_{\mathbf{w}}(\mathbf{x}_t)$, price $p_t(\mathbf{x}_t)$ where $\mathbf{x}_t \overset{iid}{\sim} \mathcal{P}$.
  Obtain an estimate $\hat{Z}_t$ of $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h^{-1}_{\hat{\mathbf{w}}}(h_{\mathbf{w}}(\mathbf{x}_t))]$ using the Integration Algorithm by Lovász and Vempala [157].
  Make decision $b_t = \mathbb{1}(\hat{Z}_t \geqslant p_t(\mathbf{x}_t))$.
end for

Algorithm 8 begins with an exploration phase of length $t' = \tilde{O}(\sqrt{Td\ell})$, during which the buyer buys the item in each round. The algorithm then uses linear programming to solve for separators $\hat{\boldsymbol{w}} = (\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_\ell)$, $\hat{\mathbf{w}}_j \in \mathbb{R}^d$ for all $j \in [\ell]$, such that $\operatorname{sgn}(\mathbf{w}_j \cdot \mathbf{x}_i) = \operatorname{sgn}(\hat{\mathbf{w}}_j \cdot \mathbf{x}_i)$ for all $j \in [\ell]$ and $i \in [t']$. During the rest of the rounds $t \in \{t'+1, \ldots, T\}$, the algorithm exploits. Since the optimal strategy is to buy if $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[v^*(\mathbf{x}) \mid h_{\mathbf{w}}(\mathbf{x}) = h_{\mathbf{w}}(\mathbf{x}_t)] \geqslant p(\mathbf{x}_t)$ (Prop. 5.2.3), the algorithm uses $h_{\mathbf{w}}(\mathbf{x}_t)$ and $\hat{\mathbf{w}} = (\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_\ell)$ to compute an estimate of $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[v^*(\mathbf{x}) \mid h_{\mathbf{w}}(\mathbf{x}) = h_{\mathbf{w}}(\mathbf{x}_t)]$.

The intuition behind the estimate is the following. Letting $h^{-1}_{\mathbf{w}}(\mathbf{x}_t) = \{\boldsymbol{x} : h_{\mathbf{w}}(\mathbf{x}) = h_{\mathbf{w}}(\mathbf{x}_t)\}$ (a convex polytope), we have that $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[v^*(\mathbf{x}) \mid h_{\mathbf{w}}(\mathbf{x}) = h_{\mathbf{w}}(\mathbf{x}_t)] = \mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h^{-1}_{\mathbf{w}}(h_{\mathbf{w}}(\mathbf{x}_t))]$. Since the buyer does not know $\mathbf{w}$, we cannot compute the set $h^{-1}_{\mathbf{w}}(h_{\mathbf{w}}(\mathbf{x}_t))$, but we can compute the set $h^{-1}_{\hat{\mathbf{w}}}(h_{\mathbf{w}}(\mathbf{x}_t))$ using the estimated separators that we have obtained after the exploration phase. Even still, the conditional expectation $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h^{-1}_{\hat{\mathbf{w}}}(h_{\mathbf{w}}(\mathbf{x}_t))\right]$ may be challenging to compute in high dimensions. Therefore, we use a sampling algorithm by Lovász and Vempala [157] to compute an estimate Z_t of $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h^{-1}_{\hat{\mathbf{w}}}(h_{\mathbf{w}}(\mathbf{x}_t))]$. The buyer buys the item if $Z_t \geqslant p_t(\mathbf{x}_t)$.

To compute the estimate Z_t, Lovász and Vempala [157] require that if π is the density function of $\mathcal{P}$, then $v^*(\mathbf{x})\pi(\mathbf{x})$ is log-concave and "well-rounded." Many well-studied distributions are log-concave, including the normal, exponential, uniform, and beta distributions, among many others. Moreover, every concave function that is nonnegative on its domain

is log-concave. If v^* and π are log-concave, then $v^*(\mathbf{x})\pi(\mathbf{x})$ is also log-concave. For example, the Cartesian product of single-dimensional log-concave distributions (exponential, logistic, extreme value, Laplace, and beta distributions, among many others) is log-concave. Log-concavity has also been widely-assumed in prior works in machine learning and high-dimensional statistics [e.g., 20, 191].

The function $v^*(\mathbf{x})\pi(\mathbf{x})$ is well-rounded if for any $\mathcal{A} \subseteq \mathcal{X}$, the distribution defined by $f_\pi(\mathcal{A}) = \frac{\int_{\mathcal{A}} v^*(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{X}} v^*(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}}$ is neither too spread out nor too concentrated. We include the formal definition in Appendix 5.7 (Def. 5.7.1). Every log-concave function can be brought to a well-rounded position by an affine transformation of the space in polynomial time [157].

Regret analysis. We now prove that the regret of Algorithm 8 is $\tilde{O}(\sqrt{Td\ell})$. To do so, we must contend with two sources of error: the fact that we use the learned linear separators $\hat{\mathbf{w}}$ instead of $\mathbf{w}$ and the estimation error introduced by the sampling algorithm.

We begin our analysis by proving that for any $\boldsymbol{y} \in \{0,1\}^\ell$ in the image of $h_{\mathbf{w}} : \mathcal{X} \to \{0,1\}^\ell$, the agent's expected value conditioned on $\mathbf{x} \in h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})$ is close to its true expected value conditioned on $\mathbf{x} \in h_{\mathbf{w}}^{-1}(\boldsymbol{y})$. In this section, we use the notation $\epsilon = \frac{\ell}{t'}\left(d \ln \frac{2et'}{d} + \ln \frac{2\ell}{\delta}\right)$.

Lemma 5.3.1. *For any $\boldsymbol{y} \in \{0,1\}^\ell$, with probability at least $1-\delta$ over $\mathbf{x}_1, \ldots, \mathbf{x}_{t'} \sim \mathcal{P}$,*

$$\left|\mathbb{E}\left[v^*(\mathbf{x}) | \mathbf{x} \in h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right] - \mathbb{E}\left[v^*(\mathbf{x}) | \mathbf{x} \in h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right]\right| \leqslant H\epsilon.$$

Proof. We can decompose the set $h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})$ as

$$h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) = \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \cap h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right) \cup \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right).$$

We can therefore write

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{P}}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right] \\
=&\,\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \cap h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \\
&\cdot \Pr\left[\mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \cap h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \\
+&\,\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \\
&\cdot \Pr\left[\mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right].
\end{aligned}$$

Similarly, we can write

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{P}}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right] \\
=&\,\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \cap h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \\
&\cdot \Pr\left[\mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \cap h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \\
+&\,\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\mathbf{w}}^{-1}(\boldsymbol{y}) \setminus h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right)\right] \\
&\cdot \Pr\left[\mathbf{x} \in \left(h_{\mathbf{w}}^{-1}(\boldsymbol{y}) \setminus h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right)\right].
\end{aligned}$$

Matching terms, we have that

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{P}}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right] - \mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right]\right| \\
=&\left|\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right]\right. \\
&\cdot \Pr\left[\mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \\
&- \mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\mathbf{w}}^{-1}(\boldsymbol{y}) \setminus h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right)\right] \\
&\left.\cdot \Pr\left[\mathbf{x} \in \left(h_{\mathbf{w}}^{-1}(\boldsymbol{y}) \setminus h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right)\right]\right|.
\end{aligned}$$

We know that $\mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)$ if and only if $h_{\widehat{\mathbf{w}}}(\mathbf{x}) = \boldsymbol{y}$ and $h_{\mathbf{w}}(\mathbf{x}) \neq \boldsymbol{y}$, which means that $h_{\widehat{\mathbf{w}}}(\mathbf{x}) \neq h_{\mathbf{w}}(\mathbf{x})$. The following claim bounds $\Pr[h_{\widehat{\mathbf{w}}}(\mathbf{x}) = \boldsymbol{y} \text{ and } h_{\mathbf{w}}(\mathbf{x}) \neq \boldsymbol{y}] \leqslant \Pr[h_{\widehat{\mathbf{w}}}(\mathbf{x}) \neq h_{\mathbf{w}}(\mathbf{x})]$.

Claim 5.3.2. *With probability $1-\delta$, $\Pr_{\mathbf{x}\sim\mathcal{P}}[h_{\widehat{\mathbf{w}}}(\mathbf{x}) \neq h_{\mathbf{w}}(\mathbf{x})] \leqslant \epsilon$.*

Proof of Claim 5.3.2. For a fixed $i \in [\ell]$, by the standard PAC learning generalization bound in the realizable setting (e.g., Theorem 4.8 by Anthony and Bartlett [9]), we have that with probability $1 - \frac{\delta}{\ell}$,

$$\Pr[\operatorname{sgn}(\mathbf{w}_i \cdot \boldsymbol{x}) \neq \operatorname{sgn}(\widehat{\mathbf{w}}_i \cdot \boldsymbol{x})] \leqslant \frac{1}{t'}\left(d \ln \frac{2et'}{d} + \ln \frac{2\ell}{\delta}\right).$$

Therefore, with probability $1-\delta$,

$$\begin{aligned}
&\Pr_{\mathbf{x}\sim\mathcal{P}}[h_{\widehat{\mathbf{w}}}(\mathbf{x}) \neq h_{\mathbf{w}}(\mathbf{x})] \\
=&\Pr[\exists i \in [\ell] \text{ such that } \operatorname{sgn}(\mathbf{w}_i \cdot \boldsymbol{x}) \neq \operatorname{sgn}(\widehat{\mathbf{w}}_i \cdot \boldsymbol{x})] \leqslant \epsilon,
\end{aligned}$$

as claimed. ∎

By Claim 5.3.2 and the fact that $v^*(\boldsymbol{x}) \in [0, H]$, we therefore know that $\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \cdot \Pr_{\mathbf{x}\sim\mathcal{P}}\left[\mathbf{x} \in \left(h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y}) \setminus h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right)\right] \in [0, H\epsilon]$. By a symmetric argument, $\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in \left(h_{\mathbf{w}}^{-1}(\boldsymbol{y}) \setminus h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right)\right] \cdot \Pr_{\mathbf{x}\sim\mathcal{P}}\left[\mathbf{x} \in \left(h_{\mathbf{w}}^{-1}(\boldsymbol{y}) \setminus h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right)\right] \in [0, H\epsilon]$. Therefore, the lemma statement holds. ∎

Lemma 5.3.1 guarantees that $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\widehat{\mathbf{w}}}^{-1}(\boldsymbol{y})\right]$ is a good approximation of $\mathbb{E}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(\boldsymbol{y})\right]$—which is the key quantity needed to compute the optimal policy (see Prop. 5.2.3). However, this estimate may be difficult to compute when $\mathbf{x}$ is high dimensional, despite the fact that $\mathcal{P}$ is known. The integration algorithm of Lovász and Vempala [157] allows us to estimate it in polynomial time, as we summarize in the following lemma.

Lemma 5.3.3. *Suppose that $v^*(\mathbf{x})\pi(\mathbf{x})$ is log-concave and well-rounded. Then for any $\mathbf{y} \in \{0,1\}^\ell$, with probability at least $1-\delta$, we can compute a constant A in polynomial time such that $|A - \mathbb{E}_{\mathbf{x}\sim\mathcal{P}}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(\mathbf{y})\right]| \leqslant H\epsilon$.*

Proof. By Lemma 5.3.1, with probability at least $1 - \delta/2$,

$$\left| \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathcal{P}} \left[v^*(\tilde{\boldsymbol{x}}) \mid \tilde{\boldsymbol{x}} \in h_{\widehat{\mathbf{w}}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x})) \right] - \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathcal{P}} \left[v^*(\tilde{\boldsymbol{x}}) \mid \tilde{\boldsymbol{x}} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x})) \right] \right| \leqslant \epsilon H.$$

By definition

$$\mathop{\mathbb{E}}_{\tilde{\mathbf{x}} \sim \mathcal{P}} [v^*(\tilde{\mathbf{x}}) \mid \tilde{\boldsymbol{x}} \in h_{\widehat{\mathbf{w}}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x}))] = \int_{\tilde{\mathbf{x}} \in h_{\widehat{\mathbf{w}}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x}))} v^*(\tilde{\mathbf{x}}) \pi(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}.$$

Then, by Lovász and Vempala [157, Theorem 1.3], in polynomial runtime with probability of at least $1 - \delta/2$, we can compute a constant A, such that

$$\left| A - \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathcal{P}} \left[v^*(\tilde{\boldsymbol{x}}) \mid \tilde{\boldsymbol{x}} \in h_{\widehat{\mathbf{w}}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x})) \right] \right| \leqslant \epsilon \cdot \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathcal{P}} \left[v^*(\tilde{\boldsymbol{x}}) \mid \tilde{\boldsymbol{x}} \in h_{\widehat{\mathbf{w}}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x})) \right] \leqslant \epsilon H.$$

By the triangle inequality and a union bound, we have that with probability of at least $1 - \delta$, we can compute a value A such that

$$\left| A - \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathcal{P}} \left[v^*(\tilde{\boldsymbol{x}}) \mid \tilde{\boldsymbol{x}} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x})) \right] \right| \leqslant 2\epsilon H,$$

which completes the proof. ■

Lemma 5.3.4. *In each round $t \in \{t' + 1, \ldots, T\}$ of the exploitation phase in Algorithm 8, with probability at least $1 - \delta$, the expected instantaneous regret incurred in round t is at most*

$$\frac{2H\ell}{t'} \left(d \ln \frac{2et'}{d} + \ln \frac{\ell \cdot 2^{\ell+2}}{\delta} \right).$$

Proof. Given $h_{\mathbf{w}}(\mathbf{x}_t)$ and price $p(\mathbf{x}_t)$, denote the estimated value of $\mathbb{E}_{\mathbf{x} \sim \mathcal{P}} \left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\widehat{\mathbf{w}}}^{-1}(h_{\mathbf{w}}(\boldsymbol{x}_t)) \right]$ obtained using the sampling algorithm (Alg 8, step 10) as $A(h_{\mathbf{w}}(\mathbf{x}_t))$. For simplicity of notation, we denote the decision of the oracle policy as s^* and the decision of the learned policy as s_t:

$$s^* = \mathbb{1} \left(\mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}} \left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\boldsymbol{w}}^{-1}(h_{\boldsymbol{w}}(\boldsymbol{x}_t)) \right] > p(\boldsymbol{x}_t) \right),$$
$$s_t = \mathbb{1} \left(A(h_{\mathbf{w}}(\mathbf{x}_t)) > p(\boldsymbol{x}_t) \right).$$

Now we bound the expected instantaneous regret in round t:

$$\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[(v^*(\mathbf{x}_t) - p(\mathbf{x}_t)) \, s^* - (v^*(\mathbf{x}_t) - p(\mathbf{x}_t)) \, s_t \right]$$

$$= \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[\left(v^*(\mathbf{x}_t) - p(\mathbf{x}_t) \right) \left(s^* - s_t \right) \right].$$

Let Δ denote the difference $\Delta = s^* - s_t$, so $\Delta \in \{-1, 0, 1\}$ is a random variable that depends on $\mathbf{x}_t$. By the law of total expectation,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[\left(v^*(\mathbf{x}_t) - p(\mathbf{x}_t) \right) s^* - \left(v^*(\mathbf{x}_t) - p(\mathbf{x}_t) \right) s_t \right] \\
&= \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[\mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}} \left[\left(v^*(\mathbf{x}) - p(\mathbf{x}_t) \right) \Delta \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\mathbf{x}_t)) \right] \right] \\
&= \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[\left(\mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}} \left[v^*(\mathbf{x}) | \mathbf{x} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\mathbf{x}_t)) \right] - p(\mathbf{x}_t) \right) \Delta \right].
\end{aligned}$$

The variable Δ is only nonzero when $s^* \neq s_t$. Let E denote the event where $s^* \neq s_t$ and let $p_E = \mathbb{P}_{\mathbf{x}_t \sim \mathcal{P}}[E]$. Then

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[\left(v^*(\mathbf{x}_t) - p(\mathbf{x}_t) \right) s^* - \left(v^*(\mathbf{x}_t) - p(\mathbf{x}_t) \right) s_t \right] \\
&\leqslant \mathop{\mathbb{E}}_{\mathbf{x}_t} \left[\left| \mathop{\mathbb{E}}_{\mathbf{x}} \left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\mathbf{x}_t)) \right] - p(\mathbf{x}_t) \right| \,\middle|\, E \right] \cdot p_E \\
&\leqslant \mathop{\mathbb{E}}_{\mathbf{x}_t} \left[\left| \mathop{\mathbb{E}}_{\mathbf{x}} \left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\mathbf{x}_t)) \right] - p(\mathbf{x}_t) \right| \,\middle|\, E \right].
\end{aligned}$$

By definition, when event E happens, we know that

$$\begin{aligned}
&A(h_{\mathbf{w}}(\mathbf{x}_t)) < p(\boldsymbol{x}_t) \leqslant \mathop{\mathbb{E}}_{\mathbf{x}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\boldsymbol{w}}^{-1}(h_{\boldsymbol{w}}(\boldsymbol{x}_t))] \text{ or} \\
&\mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\boldsymbol{w}}^{-1}(h_{\boldsymbol{w}}(\mathbf{x}_t))] < p(\mathbf{x}_t) \leqslant A(h_{\mathbf{w}}(\mathbf{x}_t)),
\end{aligned}$$

where in either case we have that

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\boldsymbol{w}}^{-1}(h_{\boldsymbol{w}}(\mathbf{x}_t))] - p(\mathbf{x}_t) \right| \\
&\leqslant \left| \mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\boldsymbol{w}}^{-1}(h_{\boldsymbol{w}}(\mathbf{x}_t))] - A(h_{\mathbf{w}}(\mathbf{x}_t)) \right|.
\end{aligned}$$

Given $\delta' \in (0, 1)$, let $\epsilon' = \frac{\ell}{t'} \left(d \ln \frac{2et'}{d} + \ln \frac{4\ell}{\delta'} \right)$. By Lemma 5.3.3, for any value of $h_{\mathbf{w}}(\mathbf{x}) \in \mathcal{Y}$, with probability at least $1 - \delta'$,

$$\left| \mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}}[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\boldsymbol{w}}^{-1}(h_{\boldsymbol{w}}(\mathbf{x}_t))] - A(h_{\mathbf{w}}(\mathbf{x}_t)) \right| \leqslant 2\epsilon' H.$$

Setting $\delta' = \delta / |\mathcal{Y}| = \delta / 2^\ell$, by a union bound over elements in $\mathcal{Y}$, we have that with probability at least $1 - \delta$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[\left(v^*(\mathbf{x}_t) - p(\mathbf{x}_t) \right) s^* - \left(v^*(\mathbf{x}_t) - p(\mathbf{x}_t) \right) s_t \right] \\
&\leqslant \mathop{\mathbb{E}}_{\mathbf{x}_t \sim \mathcal{P}} \left[\left| \mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}} \left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\mathbf{x}_t)) \right] - p(\mathbf{x}_t) \right| \,\middle|\, E \right]
\end{aligned}$$

$$\leqslant \mathop{\mathbb{E}}_{\mathbf{x}_t}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}}\left[v^*(\mathbf{x}) \mid \mathbf{x} \in h_{\mathbf{w}}^{-1}(h_{\mathbf{w}}(\mathbf{x}_t))\right] - A(h_{\mathbf{w}}(\mathbf{x}_t))\right| \,\middle|\, E\right]$$
$$\leqslant \frac{2H\ell}{t'}\left(d\ln\frac{2et'}{d} + \ln\frac{\ell\cdot 2^{\ell+2}}{\delta}\right),$$

which completes the proof. ■

This instantaneous regret bound then implies a regret bound for Algorithm 8, the proof of which is in Appendix 5.7.

Theorem 5.3.5. *With probability at least $1-\delta$, the regret of Algorithm 8 is $R_T = \mathcal{O}(\sqrt{Td\ell\log(T\ell/\delta)})$.*

In Algorithm 8, we assumed that we knew the horizon T. This assumption can be lifted via a doubling trick; see Appendix 5.7.

5.4 General masking functions

We next consider a more general setting where in each round, an item $\mathbf{x}_t$ is drawn from an unknown distribution $\mathcal{P}$ and a published price p_t is drawn from some fixed unknown distribution. We study the case where the masking function is an arbitrary mapping $h : \mathcal{X} \to [n]$. In this setting, is there a no-regret strategy that the buyer can use?

We answer this question in the affirmative, building on the classical `Exp4.VC` algorithm [36]. Out of the box, `Exp4.VC` has per-round runtime that is exponential in n, but we exploit the structure of our problem setting to obtain a polynomial per-round runtime. We prove that this algorithm has a regret bound of $\mathcal{O}(\sqrt{T(n\log T/n + \log 1/\delta)})$ with probability at least $1-\delta$.

`Exp4.VC` algorithm

Based on the optimal strategy from Proposition 5.2.3, we define an infinite set of policies that take as input $(p_t, h(\mathbf{x}_t))$ and return decisions in $\{0,1\}$ indicating whether or not the buyer should buy the item. Each policy is defined by a vector $\boldsymbol{v} \in [0,H]^n$ as follows: $\pi_{\boldsymbol{v}}(p_t, h(\mathbf{x}_t)) = \mathbb{1}(\boldsymbol{v}[h(\mathbf{x}_t)] \geqslant p_t)$. The optimal strategy from Proposition 5.2.3 corresponds to the strategy $\pi_{\boldsymbol{v}}$ with $\boldsymbol{v} = (\mathbb{E}[v^*(\mathbf{x}) \mid h(\mathbf{x}) = 1], \ldots, \mathbb{E}[v^*(\mathbf{x}) \mid h(\mathbf{x}) = n])$. We use the notation $\Pi = \{\pi_{\boldsymbol{v}} \mid \boldsymbol{v} \in [0,H]^n\}$ to denote the set of all such policies.

A key observation is that our problem can be framed as a contextual bandit problem with an oblivious adversary and an infinite set of contexts. At each round $t = 1, \ldots, T$, the buyer observes stochastic context $(p_t, h(\mathbf{x}_t))$ and makes a purchase decision. This is a contextual bandit problem with two arms: the first arm corresponds to the decision "no purchase" and the second arm corresponds to the decision "purchase." The reward of the first arm is always zero, while the reward of pulling the second arm depends on the item $\mathbf{x}_t$ and the price p_t.

We prove that the class of policies Π has a VC dimension of n, which allows us to adapt `Exp4.VC` from Beygelzimer et al. [36], a generic contextual bandits algorithm for

Algorithm 9 `Exp4.VC` with an unknown distribution

Input: $T \geqslant 0$, $\delta \in (0,1)$.
Set $\tau = \sqrt{Tn \log \frac{eT}{n} + \log \frac{2}{\delta}}$.
for $t = 1, 2, \ldots, \tau$ **do** // Initialization phase
  Receive $h(\mathbf{x}_t)$, price p_t where $\mathbf{x}_t \overset{iid}{\sim} \mathcal{P}$.
  Make decision $b_t \sim Bern(0.5)$ at random.
end for
for $i = 1, 2, \ldots, n$ **do** // Extract a finite subset of policies
  Let $i_1, i_2, \ldots, i_{m_i} \in [\tau]$ be the set of indices where $h(\mathbf{x}_{i_j}) = i$ for all $j \in \{1, 2, \ldots, m_i\}$
  Define $\mathcal{V}_i = \{0, p_{i_1}, p_{i_2}, \ldots, p_{i_{m_i}}\}$
end for
Define $\mathcal{V} = \times_{i=1}^{n} \mathcal{V}_i$
Set $\gamma = \sqrt{\frac{\log |\mathcal{V}|}{2(T-\tau)}}$ and $w_{\tau+1}^{\boldsymbol{v}} = 1$ for all $\boldsymbol{v} \in \mathcal{V}$.
for $t = \tau + 1, \ldots, T$ **do** // `Exp.4` subroutine
  Receive $h(\mathbf{x}_t)$, price p_t where $\mathbf{x}_t \overset{iid}{\sim} \mathcal{P}$.
  Get advice vectors $\boldsymbol{\xi}_t^{\boldsymbol{v}} \in \{0,1\}^2$ for all $\boldsymbol{v} \in \mathcal{V}$ where $\boldsymbol{\xi}_t^{\boldsymbol{v}} = (1 - \pi_{\boldsymbol{v}}(p_t, h(\mathbf{x}_t)), \pi_{\boldsymbol{v}}(p_t, h(\mathbf{x}_t)))$.
  Set $W_t = \sum_{\boldsymbol{v} \in \mathcal{V}} w_t^{\boldsymbol{v}}$ and define $\bar{\boldsymbol{\xi}}_t \in (0,1)^2$ as

$$\bar{\boldsymbol{\xi}}_t[0] = (1 - 2\gamma) \sum_{\boldsymbol{v} \in \mathcal{V}} \frac{w_t^{\boldsymbol{v}} \boldsymbol{\xi}_t^{\boldsymbol{v}}[0]}{W_t} + \gamma$$

$$\bar{\boldsymbol{\xi}}_t[1] = (1 - 2\gamma) \sum_{\boldsymbol{v} \in \mathcal{V}} \frac{w_t^{\boldsymbol{v}} \boldsymbol{\xi}_t^{\boldsymbol{v}}[1]}{W_t} + \gamma.$$

  Draw decision $b_t \sim Bern(\bar{\boldsymbol{\xi}}_t[1])$ and receive reward $u_t = (v^*(\mathbf{x}_t) - p_t) b_t$.
  Set $\hat{\mathbf{r}}_t = (b_t u_t / \bar{\boldsymbol{\xi}}_t[1], 0)^\top$.
  for $\boldsymbol{v} \in \mathcal{V}$ **do**
    Set

$$C_t^{\boldsymbol{v}} = \frac{\gamma}{2} \left(\boldsymbol{\xi}_t^{\boldsymbol{v}} \cdot \hat{\mathbf{r}}_t + \sum_{b=0}^{1} \frac{\boldsymbol{\xi}_t^{\boldsymbol{v}}[b]}{\bar{\boldsymbol{\xi}}_t^{\boldsymbol{v}}[b]} \sqrt{\frac{\log |\mathcal{V}|/\delta}{2(T-\tau)}} \right)$$

    Set $w_{t+1}^{\boldsymbol{v}} = w_t^{\boldsymbol{v}} \exp C_t^{\boldsymbol{v}}$
  end for
end for

policy classes with finite VC dimension. Algorithm 9 begins with an initialization phase of length τ. In this phase, the buyer chooses their action uniformly at random and collects

tuples $\{(h(\mathbf{x}_t), p_t)\}_{t=1}^{\tau}$. The algorithm then uses these tuples to identify a finite (though exponentially large), representative subset of policies in Π. In particular, using the collected tuples, the algorithm partitions Π into a *finite* set of equivalence classes where two policies π, π' are equivalent if they agree on the set of τ collected tuples. Then the buyer constructs a finite set of policies Π' by selecting one policy from each equivalence class. Algorithm 9 does this by first defining for each $i \in [n]$ a set $\mathcal{V}_i \subset \mathbb{R}$ which is the set of all prices from the first τ rounds for items $\mathbf{x}_t$ with $h(\mathbf{x}_t) = i$. The finite set Π' of policies is then defined as $\Pi' = \{\pi_{\boldsymbol{v}} : \boldsymbol{v} \in \times_{i=1}^{n} \mathcal{V}_i\}$. We prove that Π' contains a policy from each equivalence class in Lemma 5.4.2.

In the remaining rounds, Algorithm 9 follows the `Exp4.P` strategy [36] which runs multiplicative weight updates on each of the selected policies $\pi_{\boldsymbol{v}}$ with $\boldsymbol{v} \in \times_{i=1}^{n} \mathcal{V}_i$. Out-of-the-box, `Exp4.VC` would therefore have a per-round runtime that is exponential in n since $\times_{i=1}^{n} \mathcal{V}_i$ is exponentially large. However, with a careful analysis, we show that in our setting these multiplicative weight updates can be computed in polynomial time.

Regret

The key first step is to show that although the set of all policies we need to consider Π is infinite, it has a *finite* VC dimension. The full proof is in Appendix 5.8.

Lemma 5.4.1. *The VC dimension of* Π *is* n*.*

Proof sketch. First, we show that the functions in Π cannot be used to label $n+1$ contexts in all possible ways. Given $n+1$ contexts $(h(\mathbf{x}_1), p_1), \ldots, (h(\mathbf{x}_{n+1}), p_{n+1})$, by the pigeonhole principle there must exist at least two items $\mathbf{x}_i$ and $\mathbf{x}_j$ that have the same index: $h(\mathbf{x}_i) = h(\mathbf{x}_j)$. Therefore, for any policy $\pi_{\boldsymbol{v}}$, the decisions for these two items are determined by the same threshold $\boldsymbol{v}[h(\mathbf{x}_i)] = \boldsymbol{v}[h(\mathbf{x}_j)]$. Without loss of generality, assume that $p_i < p_j$. There is no policy $\pi_{\boldsymbol{v}}$ where the decision is to purchase item j but not purchase item i because this would imply that $p_j \leqslant \boldsymbol{v}[h(\mathbf{x}_j)] = \boldsymbol{v}[h(\mathbf{x}_i)] < p_i$. However, with fewer than $n+1$ items, since all items can use a different threshold, their decisions do not interfere with each other. ■

Next, in order to invoke the regret bound of `Exp4.VC` , we verify that Π' is a representative set of policies. Formally, suppose we partition Π into a set of equivalence classes where policies π and π' are equivalent if they agree on the set of τ tuples collected in the initalization phase. We prove that Π' contains a policy from each equivalence class.

Lemma 5.4.2. *Let* $\mathcal{V}$ *be defined as in Algorithm 9. Then*

$$\begin{aligned} &\{(\pi_{\boldsymbol{v}}(h(\mathbf{x}_1), p_1), \ldots, \pi_{\boldsymbol{v}}(h(\mathbf{x}_\tau), p_\tau)) : \boldsymbol{v} \in [0, H]^n\} \\ = &\{(\pi_{\boldsymbol{v}}(h(\mathbf{x}_1), p_1), \ldots, \pi_{\boldsymbol{v}}(h(\mathbf{x}_\tau), p_\tau)) : \boldsymbol{v} \in \mathcal{V}\}. \end{aligned}$$

The proof of this lemma can be found in Appendix 5.8.

Lemmas 5.4.1 and 5.4.2 imply the following regret bound:[3]

Theorem 5.4.3. *With probability $1-\delta$, Algorithm 9 achieves a regret rate that is $R_T = \mathcal{O}(\sqrt{T(n\log T/n + \log 1/\delta)})$.*

Proof. This theorem follows directly from Lemmas 5.4.1 and 5.4.2 and Theorem 5 of Beygelzimer et al. [36]. ■

Computational Complexity

The key challenge in applying `Exp4.VC` out-of-the-box is that it computes multiplicative weight updates over every policy $\pi_{\boldsymbol{v}}$ with $\boldsymbol{v} \in \mathcal{V}$—an exponential number of policies. We show that by exploiting our problem structure, we can perform these multiplicative weight updates in each round in polynomial time. In particular, we show that we can efficiently compute the purchase probabilities $\bar{\boldsymbol{\xi}}_t[0]$ and $\bar{\boldsymbol{\xi}}_t[1]$ without computing the multiplicative weights $w_t^{\mathbf{v}}$ for each $\mathbf{v} \in \mathcal{V}$ explicitly, and therefore Algorithm 9 can be run with polynomial per-round runtime. Intuitively, rather than sum over every vector in $\mathcal{V} = \times_{i=1}^n \mathcal{V}_i$ as in the definitions of $\bar{\boldsymbol{\xi}}_t[0]$ and $\bar{\boldsymbol{\xi}}_t[1]$, we show how to sum over individual elements in $\cup_{i=1}^n \mathcal{V}_i$, of which there are $\tau + n = \tilde{\mathcal{O}}(\sqrt{Tn} + n)$. We provide the complete proof in Appendix 5.8.

Theorem 5.4.4. *The purchase probabilities $\bar{\boldsymbol{\xi}}_t[0]$ and $\bar{\boldsymbol{\xi}}_t[1]$ in Algorithm 9 can be computed in $\mathcal{O}(n+\tau) = \mathcal{O}(n + \sqrt{Tn\log(T/n) + \log(1/\delta)})$ time.*

Proof sketch. Our proof begins with the observation that for each index $i \in [n]$, the thresholds $0 \leqslant p_{i_1} \leqslant \cdots \leqslant p_{i_{m_i}}$ in $\mathcal{V}_i$ divide the price range $[0, H]$ into $m_i + 1$ non-overlapping "buckets": $[0, p_{i_1}), [p_{i_2}, p_{i_3}), \ldots, [p_{i_{m_i}}, H]$. Using the notation $m = \max_i m_i$, the total number of buckets is $\mathcal{O}(mn)$. Each context $(h(\mathbf{x}_t), p_t)$ corresponds to exactly one bucket: the bucket $[p_{i_j}, p_{i_{j+1}})$ containing p_t where $h(\mathbf{x}_t) = i$. Moreover, the decision of each policy in each bucket is constant since the policies all use the same thresholds, namely, the boundaries of these buckets. Given a vector $\boldsymbol{v} \in \mathcal{V}$ and a bucket k, let $a_k^{\mathbf{v}} \in \{0, 1\}$ be the policy's recommendation of "buy" or "do not buy" for any item that falls in that bucket. This allows us to rewrite the policy decision $\boldsymbol{\xi}_t^{\mathbf{v}}$ as an alternative sum: $\boldsymbol{\xi}_t^{\mathbf{v}} = \left(\sum_{k:a_k^{\mathbf{v}}=0} \mathbb{1}(\text{item in } k), \sum_{k:a_k^{\mathbf{v}}=1} \mathbb{1}(\text{item in } k)\right)^\top$. Intuitively, $\mathbb{1}(\text{item in } k)$ is only nonzero for the bucket that this item belongs to, and the policy's decision for the item is the same as the policy's decision for that bucket. Using this fine-grained argument, we then show that all the purchase probabilities $\bar{\boldsymbol{\xi}}_t[0]$ and $\bar{\boldsymbol{\xi}}_t[1]$ can be computed in polynomial time without explicitly computing the exponentially many weights $w_t^{\boldsymbol{v}}$. ■

Remark 5.4.5. *Under* adversarial *prices, we can run n independent copies of an algorithm for Lipschitz contextual bandits (for example, the algorithm from Section 8.3 of the textbook by Slivkins et al. [198]) to obtain an expected regret bound of $\tilde{O}(T^{2/3}n^{1/3})$.*

[3] By running n copies of `Exp4.VC` in parallel for each context, we would obtain a regret bound of $\mathcal{O}(\sqrt{Tn\log(Tn/\delta)})$, but by using a more careful analysis in this section, we improve the dependence on n.

5.5 Conclusion and future directions

We presented learning algorithms for buyers who participate in auctions with limited item information. This model captures a broad set of practical applications, including advertising auctions. Our algorithms are no-regret with respect to an oracle buyer who has perfect knowledge of the distribution over items and the masking function that the seller uses to obfuscate the item information. We proposed no-regret learning algorithms in a variety of settings, including when the distribution over items is either known or unknown to the buyer, and when the prices are either stochastic or adversarial.

To the best of our knowledge, this is the first result on no-regret learning strategies for buyers with partial item information. Many interesting questions remain open for future research. First, we have assumed that the valuation is bounded, and it would be an interesting direction for future research to extend our results to unbounded distributions. Second, we focused on posted-prices in this work. It would be interesting to consider extensions to second-price auctions with partial item information. With stochastic prices, the problem seems potentially feasible, but adversarial prices might pose challenges because the adversary could block the learner from estimating the second-highest bid. Moreover, can the platform release partial information in a way that optimally trades off between revenue and privacy? When the distribution over items is unknown, our algorithm works with general masking functions. Can better purchasing strategies be developed by exploiting the properties of a specific set of masking functions?

5.6 Appendix: Proofs for Section 5.2

Proposition 5.2.3. *The strategy s^* that maximizes $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[(v^*(\mathbf{x})-p(\mathbf{x}))s^*(h(\mathbf{x}),p(\mathbf{x}),h,\mathcal{P}) \mid h(\mathbf{x})]$ is*

$$s^*(h(\mathbf{x}),p(\mathbf{x}),h,\mathcal{P}) = \mathbb{1}\left(\mathop{\mathbb{E}}_{x\sim\mathcal{P}}[v^*(\mathbf{x}) \mid h(\mathbf{x})] > p(\mathbf{x})\right).$$

Proof. For a decision $b \in \{0,1\}$, let R_b denote the expected utility of buying or not buying an item given the published information and price:

$$\begin{aligned}
R_b &\stackrel{\text{def}}{=} \mathbb{E}[(v^*(\mathbf{x}) - p(\mathbf{x})) \cdot b \mid h(\mathbf{x}), p(\mathbf{x})] \\
&= \mathbb{E}[(v^*(\mathbf{x}) - p(\mathbf{x})) \cdot b \mid v^*(\mathbf{x}) \geqslant p(\mathbf{x}), h(\mathbf{x}), p(\mathbf{x})] \cdot \mathbb{P}[v^*(\mathbf{x}) \geqslant p(\mathbf{x}) \mid h(\mathbf{x}), p(\mathbf{x})] \\
&\quad - \mathbb{E}[(p(\mathbf{x}) - v^*(\mathbf{x})) \cdot b \mid v^*(\mathbf{x}) < p(\mathbf{x}), h(\mathbf{x}), p(\mathbf{x})] \cdot \mathbb{P}[v^*(\mathbf{x}) < p(\mathbf{x}) \mid h(\mathbf{x}), p(\mathbf{x})].
\end{aligned}$$

Then, by definition, the optimal strategy $s^*(h(\mathbf{x}),p(\mathbf{x}),p,\mathcal{P}) = \mathbb{1}(R_1 > R_0) = \mathbb{1}(R_1 > 0)$.

Further, letting $x^+ = \max\{0, x\}$, by the law of total expectation,

$$\begin{aligned}
&\mathbb{E}[(v^*(\mathbf{x}) - p(\mathbf{x}))^+ \mid h(\mathbf{x}), p(\mathbf{x})] \\
&= \mathbb{E}[(v^*(\mathbf{x}) - p(\mathbf{x})) \mid v^*(\mathbf{x}) \geqslant p(\mathbf{x}), h(\mathbf{x}), p(\mathbf{x})] \cdot \mathbb{P}[v^*(\mathbf{x}) \geqslant p(\mathbf{x}) \mid h(\mathbf{x}), p(\mathbf{x})] \\
&\quad + \mathbb{E}[0 \mid v^*(\mathbf{x}) < p(\mathbf{x}), h(\mathbf{x}), p(\mathbf{x})] \cdot \mathbb{P}(v^*(\mathbf{x}) < p(\mathbf{x}) \mid h(\mathbf{x}), p(\mathbf{x}))
\end{aligned}$$

$$= \mathbb{E}[(v^*(\mathbf{x}) - p(\mathbf{x})) \mid v^*(\mathbf{x}) \geqslant p(\mathbf{x}), h(\mathbf{x}), p(\mathbf{x})] \cdot \mathbb{P}[v^*(\mathbf{x}) \geqslant p(\mathbf{x}) \mid h(\mathbf{x}), p(\mathbf{x})].$$

Similarly, letting $x^- = -\min\{0, x\}$,

$$\begin{aligned}&\mathbb{E}[(v^*(\mathbf{x}) - p(\mathbf{x}))^- \mid h(\mathbf{x}), p(\mathbf{x})] \\ &= \mathbb{E}[(p(\mathbf{x}) - v^*(\mathbf{x})) \mid v^*(\mathbf{x}) < p(\mathbf{x}), h(\mathbf{x}), p(\mathbf{x})] \cdot \mathbb{P}(v^*(\mathbf{x}) < p(\mathbf{x}) \mid h(\mathbf{x}), p(\mathbf{x})).\end{aligned}$$

Therefore, the optimal strategy is:

$$\begin{aligned}s^*(h(\mathbf{x}), p(\mathbf{x}), h, p, \mathcal{P}) &= \mathbb{1}\left(\mathbb{E}[(v^*(x) - p(\mathbf{x}))^+ \mid h(\mathbf{x}), p(\mathbf{x})] - \mathbb{E}[(v^*(x) - p(\mathbf{x}))^- \mid h(\mathbf{x}), p(\mathbf{x})] > 0\right) \\ &= \mathbb{1}\left(\mathbb{E}[(v^*(x) - p(\mathbf{x})) \mid h(\mathbf{x}), p(\mathbf{x})] > 0\right) \\ &= \mathbb{1}\left(\mathbb{E}[v^*(x) \mid h(\mathbf{x}), p(\mathbf{x})] > p(\mathbf{x})\right).\end{aligned}$$

The lemma statement then follows from Equation (5.1). ■

5.7 Appendix: Proofs for Section 5.3

Definition 5.7.1 (Lovász and Vempala [157])**.** Define the centroid of f_π as $c_f = \int \mathbf{x} df_\pi$. Define the variance of f_π as $\text{var}(f_\pi) = \int \|\mathbf{x} - c_f\|^2 df_\pi$. Also denote $\mathcal{L}(\theta)$ as the level set $\{\mathbf{x} : v^*(\mathbf{x})\pi(\mathbf{x}) \geqslant \theta\}$.

Theorem 5.3.5. *With probability at least $1-\delta$, the regret of Algorithm 8 is $R_T = \mathcal{O}(\sqrt{Td\ell \log(T\ell/\delta)})$.*

Proof. First, in the exploration phase, the total regret is upper bounded by Ht'.

Now consider the regret in the exploitation phase. Notice that, we would have obtained $\{\mathbf{x}_i, h_{\mathbf{w}}(\mathbf{x}_i)\}_{i=1}^{t'}$ number of i.i.d. samples from the exploration phase.

Let $\epsilon = \frac{\ell}{t'}\left(d \ln \frac{2et'}{d} + \ln \frac{\ell \cdot 2^{\ell+2}}{\delta'}\right)$. By Lemma 5.3.4 and a union bound over all rounds, with probability at least $1 - \delta' T$, the total expected regret incurred by Algorithm 8 is $R_T \leqslant Ht' + 2\epsilon H(T - t')$.

Let $t' = \sqrt{Td\ell \log\left(\frac{2\ell}{\delta'}\right)}$, we have

$$R_T \leqslant H\sqrt{Td\ell \log\left(\frac{2\ell}{\delta'}\right)} + 2H\ell d\sqrt{\frac{T}{d\ell \log\left(\frac{2\ell}{\delta'}\right)}} \log\left(\frac{2\ell}{d}\sqrt{Td\ell \log\left(\frac{2\ell}{\delta'}\right)}\right) + 2H\ell\sqrt{\frac{T}{d\ell \log\left(\frac{2\ell}{\delta'}\right)}} \log\left(\frac{\ell 2^{\ell+2}}{\delta'}\right).$$

Note that $\log\left(\frac{\ell 2^{\ell+2}}{\delta'}\right) = \log\left(\frac{\ell}{\delta'}\right) + \ell \log(4) \leqslant \log\left(\frac{\ell}{\delta'}\right) + d \log(4)$. Setting $\delta' = \delta/T$, we have that with probability at least $1 - \delta$,

$$R_T = \tilde{\mathcal{O}}\left(\sqrt{Td\ell \log\left(\frac{2\ell T}{\delta}\right)}\right). \tag{5.3}$$

This completes the proof. ■

Algorithm for an unknown horizon T

In Theorem 5.3.5, we assumed that we knew the time horizon T, which allowed us to set the correct length for the exploration phase. This assumption can be lifted by using the doubling trick which runs the algorithm in independent intervals that are doubling in length, as summarized by Algorithm 10. The regret bound remains the same up to constant factors.

Algorithm 10 Explore-then-Commit with Unknown Horizon

1: **Input:** starting epoch length T_0.
2: **for** $i = 1, 2, \dots$ **do**
3: $\quad T_i \leftarrow 2^i T_0$.
4: $\quad$ Run Algorithm 8 with $T = T_i$.
5: **end for**

Corollary 5.7.2. *Suppose that $v^*(\mathbf{x})\pi(\mathbf{x})$ is logconcave and well-rounded. Then with probability at least $1-\delta$, the regret of Algorithm 10 is $R_T = \tilde{\mathcal{O}}(\sqrt{Td\ell \log(T\ell/\delta)})$.*

Proof. Denote the total expected accumulated regret as R_T, and the expected accumulated regret in each interval with length T_i as R_{T_i}. Denote the number of intervals as $m \leqslant \log_2 \frac{2T}{T_0}$. Then, by Theorem 5.3.5 we have that for some universal constant C and with probability at least $1 - \delta' m$:

$$\begin{aligned}
R_T &\leqslant \sum_{i=1}^{m} R_{T_i} \\
&\leqslant \sum_{i=1}^{m} C\sqrt{T_i d\ell \log\left(\frac{2\ell T_i}{\delta'}\right)} \\
&\leqslant C\sqrt{d\ell} \sum_{i=1}^{m} \sqrt{T_i \left(\log\left(\frac{2\ell}{\delta'}\right) + \log T_i\right)} \\
&\leqslant C\sqrt{d\ell} \left(\sqrt{\log\left(\frac{2\ell}{\delta'}\right)} \cdot \sum_{i=1}^{m} \sqrt{2^i T_0} + \sum_{i=1}^{m} \sqrt{2^i T_0 \log(2^i T_0)}\right) \\
&= \mathcal{O}\left(\sqrt{d\ell T \log\left(\frac{2\ell}{\delta'}\right)}\right).
\end{aligned}$$

Let $\delta' = \delta/m$ and note that $m \leqslant \log_2 \frac{2T}{T_0}$, applying a union bound over all intervals, we have that with probability at least $1-\delta$,

$$R_T = \tilde{\mathcal{O}}\left(\sqrt{d\ell T \log\left(\frac{2\ell}{\delta}\right)}\right).$$

This completes the proof. ■

5.8 Appendix: Proofs for Section 5.4

Lemma 5.4.1. *The VC dimension of* Π *is* n.

Proof. We argue that any function contained in Π cannot be used to label $n+1$ input points in all possible ways. For simplicity denote $h(\mathbf{x}) = y \in [n]$. Consider a set of input points $\{y_i, p_i\}_{i=1}^{n+1}$. By the pigeonhole principle there must exist at least two elements $(y_i, p_i), (y_j, p_j)$ such that $y_i = y_j$ and $p_i \neq p_j$. Therefore by definition, we have that for any $\mathbf{v} \in \mathbb{R}^n$,

$$\pi_{\mathbf{v}}(y_i, p_i) = \mathbb{1}(\mathbf{v}[y_i] > p_i) = \mathbb{1}(\mathbf{v}[y_j] > p_i),$$
$$\pi_{\mathbf{v}}(y_j, p_j) = \mathbb{1}(\mathbf{v}[y_j] > p_j) = \mathbb{1}(\mathbf{v}[y_i] > p_j).$$

Without loss of generality, assume that $p_i < p_j$. Then the pair of labels $\pi_{\mathbf{v}}(y_i, p_i) = 1$ and $\pi_{\mathbf{v}}(y_j, p_j) = 0$ can never be achieved for any $\mathbf{v} \in \mathbb{R}^n$. Thus $VCdim(\Pi) < n + 1$.

Next, consider a set of n input points $\{(i, 0.5)\}_{i=1}^{n}$. Each point in this set is labeled using a different index i, so all possible combinations of labels can be achieved using vectors $\boldsymbol{v} \in \{0, 1\}^n$. Thus we conclude that $VCdim(\Pi) = n$.

■

Lemma 5.4.2. *Let* $\{(h(\mathbf{x}_1), p_1), \ldots, (h(\mathbf{x}_\tau), p_\tau)\}$ *be a subset of* $[n] \times [0, H]$. *For each* $i \in [n]$, *let* $i_1, i_2, \ldots, i_{m_i} \in [\tau]$ *be the set of indices where* $h(\mathbf{x}_{i_j}) = i$ *for all* $j \in \{1, 2, \ldots, m_i\}$. *Define* $\mathcal{V}_i = \{0, p_{i_1}, p_{i_2}, \ldots, p_{i_{m_i}}\}$ *and* $\mathcal{V} = \times_{i=1}^{n} \mathcal{V}_i$. *Then*

$$\left\{ \begin{pmatrix} \pi_{\boldsymbol{v}}(h(\mathbf{x}_1), p_1) \\ \vdots \\ \pi_{\boldsymbol{v}}(h(\mathbf{x}_\tau), p_\tau) \end{pmatrix} \right\}_{\boldsymbol{v} \in [0,H]^n} = \left\{ \begin{pmatrix} \pi_{\boldsymbol{v}}(h(\mathbf{x}_1), p_1) \\ \vdots \\ \pi_{\boldsymbol{v}}(h(\mathbf{x}_\tau), p_\tau) \end{pmatrix} \right\}_{\boldsymbol{v} \in \mathcal{V}}.$$

Proof. We will show that for every $\boldsymbol{v} \in [0, H]^n$, there exists a vector $\boldsymbol{v}_0 \in \times_{i=1}^{n} \mathcal{V}_i$ such that $\pi_{\boldsymbol{v}}(h(\mathbf{x}_j), p_j) = \pi_{\boldsymbol{v}_0}(h(\mathbf{x}_j), p_j)$ for every $j \in [\tau]$. To this end, fix an index $i \in [n]$ and without loss of generality, let $i_1, i_2, \ldots, i_{m_i}$ be sorted such that $0 := p_{i_0} < p_{i_1} < p_{i_2} < \cdots < p_{i_{m_i}}$. Let $i' \in \{0, 1, 2, \ldots, i_{m_i}\}$ be the largest index such that $\boldsymbol{v}[i] \geqslant p_{i'}$. Define $\boldsymbol{v}_0[i] = p_{i'}$. For every index $i_j \leqslant i'$, we know that $\boldsymbol{v}[i] \geqslant p_{i'} = \boldsymbol{v}_0[i] \geqslant p_{i_j}$, so $\pi_{\boldsymbol{v}}(h(\mathbf{x}_{i_j}), p_{i_j}) = \mathbb{1}(\boldsymbol{v}[i] \geqslant p_{i_j}) = 1 = \mathbb{1}(\boldsymbol{v}_0[i] \geqslant p_{i_j}) = \pi_{\boldsymbol{v}_0}(h(\mathbf{x}_{i_j}), p_{i_j})$. Meanwhile, for every index $i_j > i'$, we know that $p_{i'} = \boldsymbol{v}_0[i] \leqslant \boldsymbol{v}[i] < p_{i_j}$. Therefore, $\pi_{\boldsymbol{v}}(h(\mathbf{x}_{i_j}), p_{i_j}) = 0 = \pi_{\boldsymbol{v}_0}(h(\mathbf{x}_{i_j}), p_{i_j})$. In either case, we have that $\pi_{\boldsymbol{v}}(h(\mathbf{x}_{i_j}), p_{i_j}) = \pi_{\boldsymbol{v}_0}(h(\mathbf{x}_{i_j}), p_{i_j})$. Since this is true for every index $i \in [n]$, the lemma statement holds. ■

Theorem 5.4.4. *The purchase probabilities* $\bar{\boldsymbol{\xi}}_t[0]$ *and* $\bar{\boldsymbol{\xi}}_t[1]$ *in Algorithm 9 can be computed in* $\mathcal{O}(n + \tau) = \mathcal{O}(n + \sqrt{Tn \log(T/n) + \log(1/\delta)})$ *time.*

Proof. Label the elements of $\mathcal{V}_i$ as $0 := p_{i,0} < p_{i,1} < \cdots < p_{i,m_i}$. Let $m = \max m_i$. For the ease of notation, define the variables $p_{i,m_i+1} = p_{i,m_i+2} = \cdots p_{i,m} = H$. For each $i \in [n]$, $j \in \{1, \ldots, m\}$, and $t \in \{\tau + 1, \tau_2, \ldots, T\}$, we define the variable

$$d_{i,j}(t) = \mathbb{1}(h(\mathbf{x}_t) = i \text{ and } p_t \in (p_{i,j-1}, p_{i,j}]).$$

We also set $d_{i,0}(t) = \mathbb{1}(h(\mathbf{x}_t) = i \text{ and } p_t = 0)$. Next, for each $\boldsymbol{v} \in \times_{i=1}^n \mathcal{V}_i$, $i \in [n]$, and $j \in \{0, \dots, m\}$, we define the variable $a_{i,j}^{\boldsymbol{v}} = \mathbb{1}(\boldsymbol{v}[i] \geqslant p_{i,j})$.

Claim 5.8.1. *For $b \in \{0, 1\}$,*

$$\boldsymbol{\xi}_t^{\boldsymbol{v}}[b] = \sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=b} d_{i,j}(t). \tag{5.4}$$

Proof of Claim 5.8.1. First, let $i_t = h(\mathbf{x}_t)$. If $p_t = 0$, define $j_t = 0$ and otherwise, define j_t such that $p_t \in (p_{i_t,j_t-1}, p_{i_t,j_t}]$. Therefore, $d_{i,j}(t) = 1$ if $(i, j) = (i_t, j_t)$ and $d_{i,j}(t) = 0$ otherwise. This means that

$$\sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=b} d_{i,j}(t) = \mathbb{1}(a_{i_t,j_t}^{\boldsymbol{v}} = b). \tag{5.5}$$

We split the proof into two cases: $b = 0$ and $b = 1$.

Case 1: $b = 0$**.** We know that $\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] = 1 - \pi_{\boldsymbol{v}}(p_t, h(\mathbf{x}_t)) = \mathbb{1}(\boldsymbol{v}[h(\mathbf{x}_t)] < p_t) = \mathbb{1}(\boldsymbol{v}[i_t] < p_t)$.

If $\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] = 0$, then $\boldsymbol{v}[i_t] \geqslant p_t$. Since $\boldsymbol{v} \in \mathcal{V}$, we know that $\boldsymbol{v}[i_t] = p_{i_t,j}$ for some $j \in [m]$. Moreover, since $p_t \in (p_{i_t,j_t-1}, p_{i_t,j_t}]$, the fact that $\boldsymbol{v}[i_t] \geqslant p_t$ means that $\boldsymbol{v}[i_t] \geqslant p_{i_t,j_t}$. Therefore, $a_{i_t,j_t}^{\boldsymbol{v}} = 1$. By Equation (5.5), this means that $\sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=0} d_{i,j}(t) = 0$, so Equation (5.4) holds.

Meanwhile, if $\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] = 1$, then $\boldsymbol{v}[i_t] < p_t \leqslant p_{i_t,j_t}$. Therefore, $a_{i_t,j_t}^{\boldsymbol{v}} = 0$. By Equation (5.5), this means that $\sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=0} d_{i,j}(t) = 1$, so Equation (5.4) holds.

Case 2: $b = 1$**.** We know that $\boldsymbol{\xi}_t^{\boldsymbol{v}}[1] = \mathbb{1}(\boldsymbol{v}[i_t] \geqslant p_t)$.

If $\boldsymbol{\xi}_t^{\boldsymbol{v}}[1] = 0$, then $\boldsymbol{v}[i_t] < p_t$. By the same logic as the previous case, this means that $a_{i_t,j_t}^{\boldsymbol{v}} = 0$. By Equation (5.5), this means that $\sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=1} d_{i,j}(t) = 0$, so Equation (5.4) holds.

Meanwhile, if $\boldsymbol{\xi}_t^{\boldsymbol{v}}[1] = 1$, then $\boldsymbol{v}[i_t] \geqslant p_t$. By the same logic as the previous case, $a_{i_t,j_t}^{\boldsymbol{v}} = 1$. By Equation (5.5), this means that $\sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=1} d_{i,j}(t) = 1$, so Equation (5.4) holds. ■

This means that

$$\begin{aligned}
\boldsymbol{\xi}_t^{\boldsymbol{v}} \cdot \hat{\mathbf{r}}_t &= \sum_{b=0}^{1} \sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=b} d_{i,j}(t) \hat{\mathbf{r}}_t[b] \\
&= \sum_{i=1}^{n} \sum_{j=0}^{m} d_{i,j}(t) \left(\hat{\mathbf{r}}_t[0]\, \mathbb{1}(a_{i,j}^{\boldsymbol{v}} = 0) + \hat{\mathbf{r}}_t[1]\, \mathbb{1}(a_{i,j}^{\boldsymbol{v}} = 1) \right) \\
&= \sum_{i=1}^{n} \sum_{j=0}^{m} d_{i,j}(t) \hat{\mathbf{r}}_t[a_{i,j}^{\boldsymbol{v}}].
\end{aligned}$$

Similarly,

$$\sum_{b=0}^{1} \frac{\boldsymbol{\xi}_t^{\boldsymbol{v}}[b]}{\bar{\boldsymbol{\xi}}_t^{\boldsymbol{v}}[b]} = \sum_{b=0}^{1} \sum_{(i,j):a_{i,j}^{\boldsymbol{v}}=b} \frac{d_{i,j}(t)}{\bar{\boldsymbol{\xi}}_t^{\boldsymbol{v}}[b]} = \sum_{i=1}^{n} \sum_{j=0}^{m} \frac{d_{i,j}(t)}{\bar{\boldsymbol{\xi}}_t^{\boldsymbol{v}}[a_{i,j}^{\boldsymbol{v}}]}.$$

Therefore,

$$w_{t+1}^{\boldsymbol{v}} = w_t^{\boldsymbol{v}} \exp\left(\frac{\gamma}{2}\left(\boldsymbol{\xi}_t^{\boldsymbol{v}} \cdot \hat{\mathbf{r}}_t + \sum_{b=0}^{1} \frac{\boldsymbol{\xi}_t^{\boldsymbol{v}}[b]}{\bar{\boldsymbol{\xi}}_t^{\boldsymbol{v}}[b]} \sqrt{\frac{\log |\mathcal{V}|/\delta}{2(T-\tau)}}\right)\right) = w_t^{\boldsymbol{v}} \exp\left(\sum_{i=1}^{n}\sum_{j=0}^{m} d_{i,j}(t) f_{a_{i,j}^{\boldsymbol{v}}}(t)\right),$$

where

$$f_{a_{i,j}^{\boldsymbol{v}}}(t) = \frac{\gamma}{2}\left(\hat{\mathbf{r}}_t[a_{i,j}^{\boldsymbol{v}}] + \frac{1}{\bar{\boldsymbol{\xi}}_t^{\boldsymbol{v}}[a_{i,j}^{\boldsymbol{v}}]} \sqrt{\frac{\log |\mathcal{V}|/\delta}{2(T-\tau)}}\right).$$

We can therefore write

$$\begin{aligned}
w_{t+1}^{\boldsymbol{v}} &= \prod_{\tau=1}^{t} \exp\left(\sum_{i=1}^{n}\sum_{j=0}^{m} d_{i,j}(\tau) f_{a_{i,j}^{\boldsymbol{v}}}(\tau)\right) \\
&= \exp\left(\sum_{\tau=1}^{t}\sum_{i=1}^{n}\sum_{j=0}^{m} d_{i,j}(\tau) f_{a_{i,j}^{\boldsymbol{v}}}(\tau)\right) \\
&= \exp\left(\sum_{i=1}^{n}\sum_{j=0}^{m}\sum_{\tau=1}^{t} d_{i,j}(\tau) f_{a_{i,j}^{\boldsymbol{v}}}(\tau)\right) \\
&= \prod_{i=1}^{n}\prod_{j=0}^{m} \exp\left(\sum_{\tau=1}^{t} d_{i,j}(\tau) f_{a_{i,j}^{\boldsymbol{v}}}(\tau)\right).
\end{aligned}$$

Letting $g_{i,j}(t,b) = \exp\left(\sum_{\tau=1}^{t} d_{i,j}(\tau) f_b(\tau)\right)$, we have that

$$w_{t+1}^{\boldsymbol{v}} = \prod_{i=1}^{n}\prod_{j=0}^{m} g_{i,j}(t, a_{i,j}^{\boldsymbol{v}}).$$

Let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ be the set of m increasing bit vectors $\boldsymbol{b}_1 = (1,0,0,\ldots,0)$, $\boldsymbol{b}_2 = (1,1,0,\ldots,0)$, $\boldsymbol{b}_3 = (1,1,1,\ldots,0)$, $\ldots$, $\boldsymbol{b}_m = (1,1,1,\ldots,1)$. For each $i \in [n]$ and $\boldsymbol{b}_j$, define

$$g_i(t, \boldsymbol{b}_j) = g_{i,1}(t, b_j[1]) g_{i,2}(t, b_j[2]) \cdots g_{i,m}(t, b_j[m]).$$

We now prove the following claim.

Claim 5.8.2. *The normalizing constant W_t can be computed in polynomial time as*

$$W_t = \prod_{i=1}^{n}\sum_{j=1}^{m_i} g_i(t, \boldsymbol{b}_j).$$

Proof of Claim 5.8.2. We first write

$$W_t = \sum_{\boldsymbol{v}\in\mathcal{V}} w_t^{\boldsymbol{v}}$$

$$
\begin{aligned}
&= \sum_{\boldsymbol{v}\in\mathcal{V}} \prod_{i=1}^{n}\prod_{j=0}^{m} g_{i,j}(t, a_{i,j}^{\boldsymbol{v}}) \\
&= \sum_{\boldsymbol{v}\in\mathcal{V}} \prod_{i=1}^{n}\prod_{j=0}^{m} g_{i,j}(t, \mathbb{1}(\boldsymbol{v}[i] \geqslant p_{i,j})) \\
&= \sum_{j_1=0}^{m_1} \cdots \sum_{j_n=0}^{m_n} \prod_{i=1}^{n}\prod_{j=0}^{m} g_{i,j}(t, \mathbb{1}(p_{i,j_i} \geqslant p_{i,j})).
\end{aligned}
$$

This last equality holds because $\mathcal{V} = \times_{i=1}^{n}\mathcal{V}_i$ and $\mathcal{V}_i = \{p_{i,0}, \ldots, p_{i,m_i}\}$. Moreover,

$$
W_t = \sum_{j_1=0}^{m_1} \cdots \sum_{j_n=0}^{m_n} \prod_{i=1}^{n}\prod_{j=0}^{j_i} g_{i,j}(t,1) \prod_{j=j_i+1}^{m} g_{i,j}(t,0)
$$

because $p_{i,j_i} \geqslant p_{i,j}$ for all $j \leqslant j_i$ and $p_{i,j_i} < p_{i,j}$ for all $j > j_i$.

By definition of the vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$,

$$
W_t = \sum_{j_1=0}^{m_1} \cdots \sum_{j_n=0}^{m_n} \prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i}) = \prod_{i=1}^{n}\sum_{j=1}^{m_i} g_i(t, \boldsymbol{b}_j),
$$

as claimed. ■

We next prove that $\bar{\boldsymbol{\xi}}_t$ can be computed in polynomial time. To do so, let $h(\mathbf{x}_t) = i_t$. If $p_t = 0$, define $j_t = 0$ and otherwise, define j_t such that $p_t \in (p_{i_t,j_t-1}, p_{i_t,j_t}]$. Next, define $\bar{m}_i$ as follows:

$$
\bar{m}_i = \begin{cases} j_t - 1 & \text{if } i = i_t \\ m_i & \text{otherwise.} \end{cases}
$$

Similarly, define $\underline{m}_i$ as follows:

$$
\underline{m}_i = \begin{cases} j_t & \text{if } i = i_t \\ 0 & \text{otherwise.} \end{cases}
$$

Then $\bar{\boldsymbol{\xi}}_t$ has the following form:

Claim 5.8.3. *The probabilities $\bar{\boldsymbol{\xi}}_t$ can be computed in polynomial time as:*

$$
\bar{\boldsymbol{\xi}}_t[0] = \frac{1}{W_t} \prod_{i=1}^{n}\sum_{j_i=0}^{\bar{m}_i} g_i(t, \boldsymbol{b}_{j_i})
$$

and

$$
\bar{\boldsymbol{\xi}}_t[1] = \frac{1}{W_t} \prod_{i=1}^{n}\sum_{j_i=\underline{m}_i}^{m_i} g_i(t, \boldsymbol{b}_{j_i}).
$$

Proof of Claim 5.8.3. Recall that $\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] = \mathbb{1}(\boldsymbol{v}[i_t] < p_t)$. Therefore,

$$\begin{aligned}\sum_{\boldsymbol{v}\in\mathcal{V}} w_t^{\boldsymbol{v}}\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] &= \sum_{\boldsymbol{v}\in\mathcal{V}}\prod_{i=1}^{n}\prod_{j=0}^{m} g_{i,j}(t, a_{i,j}^{\boldsymbol{v}})\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] \\ &= \sum_{\boldsymbol{v}\in\mathcal{V}}\prod_{i=1}^{n}\prod_{j=0}^{m} g_{i,j}(t, a_{i,j}^{\boldsymbol{v}})\,\mathbb{1}(\boldsymbol{v}[i_t] < p_t) \\ &= \sum_{j_1=0}^{m_1}\cdots\sum_{j_n=0}^{m_n}\left(\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})\right)\mathbb{1}(p_{i_t,j_{i_t}} < p_t).\end{aligned}$$

This last equality holds because $\mathcal{V} = \times_{i=1}^{n}\mathcal{V}_i$ and $\mathcal{V}_i = \{p_{i,0}, \ldots, p_{i,m_i}\}$. Without loss of generality, suppose that $i_t = 1$, so $\mathbb{1}(p_{i_t,j_{i_t}} < p_t) = \mathbb{1}(p_{1,j_1} < p_t)$. Then

$$\begin{aligned}\sum_{\boldsymbol{v}\in\mathcal{V}} w_t^{\boldsymbol{v}}\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] &= \sum_{j_1=0}^{m_1}\cdots\sum_{j_n=0}^{m_n}\left(\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})\right)\mathbb{1}(p_{1,j_1} < p_t) \\ &= \sum_{j_1=0}^{m_1}\mathbb{1}(p_{1,j_1} < p_t)\left(\sum_{j_2=0}^{m_2}\cdots\sum_{j_n=0}^{m_n}\left(\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})\right)\right).\end{aligned}$$

Since $p_t = 0$ if $j_t = 0$ and otherwise $p_t \in (p_{1,j_t-1}, p_{1,j_t}]$ we have that $\mathbb{1}(p_{1,j_1} < p_t) = 1$ if and only if $j_1 \leqslant j_t - 1$. Therefore,

$$\sum_{\boldsymbol{v}\in\mathcal{V}} w_t^{\boldsymbol{v}}\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] = \sum_{j_1=0}^{j_t-1}\left(\sum_{j_2=0}^{m_2}\cdots\sum_{j_n=0}^{m_n}\left(\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})\right)\right) = \sum_{j_1=0}^{\bar{m}_1}\cdots\sum_{j_n=0}^{\bar{m}_n}\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i}) = \prod_{i=1}^{n}\sum_{j_i=0}^{\bar{m}_i} g_i(t, \boldsymbol{b}_{j_i}).$$

Similarly, since $\boldsymbol{\xi}_t^{\boldsymbol{v}}[1] = \mathbb{1}(\boldsymbol{v}[i_t] \geqslant p_t)$, we have

$$\sum_{\boldsymbol{v}\in\mathcal{V}} w_t^{\boldsymbol{v}}\boldsymbol{\xi}_t^{\boldsymbol{v}}[1] = \sum_{j_1=0}^{m_1}\cdots\sum_{j_n=0}^{m_n}\left(\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})\right)\mathbb{1}(p_{i_t,j_{i_t}} \geqslant p_t).$$

Without loss of generality, suppose that $i_t = 1$, so $\mathbb{1}(p_{i_t,j_{i_t}} \geqslant p_t) = \mathbb{1}(p_{1,j_1} \geqslant p_t)$. Then

$$\sum_{\boldsymbol{v}\in\mathcal{V}} w_t^{\boldsymbol{v}}\boldsymbol{\xi}_t^{\boldsymbol{v}}[1] = \sum_{j_1=0}^{m_1}\mathbb{1}(p_{1,j_1} \geqslant p_t)\left(\sum_{j_2=0}^{m_2}\cdots\sum_{j_n=0}^{m_n}\left(\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})\right)\right).$$

Since $p_t = 0$ if $j_t = 0$ and otherwise $p_t \in (p_{1,j_t-1}, p_{1,j_t}]$ we have that $\mathbb{1}(p_{1,j_1} \geqslant p_t) = 1$ if and only if $j_1 \geqslant j_t$. Therefore,

$$\sum_{\boldsymbol{v}\in\mathcal{V}} w_t^{\boldsymbol{v}}\boldsymbol{\xi}_t^{\boldsymbol{v}}[0] = \sum_{j_1=j_t}^{m_1}\left(\sum_{j_2=0}^{m_2}\cdots\sum_{j_n=0}^{m_n}\left(\prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})\right)\right)$$

$$= \sum_{j_1=\underline{m}_1}^{m_1} \cdots \sum_{j_n=\underline{m}_n}^{m_n} \prod_{i=1}^{n} g_i(t, \boldsymbol{b}_{j_i})$$
$$= \prod_{i=1}^{n} \sum_{j_i=\underline{m}_i}^{m_i} g_i(t, \boldsymbol{b}_{j_i}).$$

■

Lastly, note that in the initialization phase of Algorithm 9, every decision b_t is computed with $\mathcal{O}(1)$ time. In the `Exp4` subroutine, by Claim 5.8.2 and Claim 5.8.3, every iteration is also computed in polynomial time. Therefore the theorem statement holds. ■

Chapter 6

Off-Policy Evaluation with Policy-Dependent Optimization Response

6.1 Introduction

The interface of causal inference and machine learning offers to "deliver the right intervention, at the right time, to the right person". An extensive line of research studies off-policy evaluation (OPE) and learning—evaluating the average causal outcomes under alternative personalized treatment assignment policies that differ from the treatment assignment which generated the data (and may have introduced confounding), so that one may optimize over the best such treatment rule [13, 73, 126, 135, 159, 202, 226]. Most of this work is based on the assumption that the appropriate decision criterion is an *average* of individuals across a population. But various operational restrictions or settings imply that a decision-maker's utility is often not realized as an *average* but rather as an *output* of a downstream planning or decision-making problem.

For example, in studying the effects of price incentives in a matching market (e.g., on a ride-share platform), a firm's revenue is not realized until it matches riders to drivers under certain constraints [158, 162]. While the marketplace may offer incentives to drive or accept rides and induce causal effects on individuals, the final utility is determined by the *new* matches, taking into account operational constraints and structure.

As another example, although job training (and personalized provision thereof) is commonly touted in causal inference and machine learning papers as a promising example for personalized treatment policy assignment [13, 135, 136], labor economists voice a general concern that "the possible existence of equilibrium effects on the efficiency of the programs seems quite real" [56, p.541]. The equilibrium concern is that personalized provision of job training may not lead to actual beneficial gains at *the population* level due to externalities (substitution effects/congestion in matching) of the labor search process in a finite market.

While impressive cluster-randomized trials have been deployed to assess these effects [57], it would be useful if there exists a framework that can model the equilibrium effect and evaluate treatment policies directly based on available data of individual-level causal effects. In some settings, population-level impacts may be well-modeled as a downstream optimization response. The development of such a framework is our focus in the current paper.

We study a new framework for policy evaluation and optimization where there is a *personalized treatment policy* on individual-level outcomes, and a *policy-dependent* optimization response. The key difference between this model and previous work on off-policy evaluation and optimization is that: although treatments realize causal effects on *individuals*, a treatment policy's value depends on a further downstream *policy-dependent* optimization. We study how to evaluate different policies without bias (off-policy evaluation) and how to optimize for the optimal policy under this framework (policy optimization).

Our contributions are as follows: we first introduce the model of policy-dependent optimization response[1], which we formulate as a nonconvex stochastic optimization problem. For off-policy evaluation, we develop a framework of *policy-dependent optimization response*, decompose the bias that arises in this framework ("optimization bias") and show how to control it via the design of estimators for the policy-dependent estimand. Finally, we provide a general algorithm for optimizing causal interventions. We corroborate the theoretical results with experimental comparisons.

Related work

We highlight the most relevant work from causal inference, off-policy evaluation, and optimization under uncertainty in the main text. We include additional or tangential discussion in Section 6.7.

There is an extensive literature on off-policy evaluation and optimization [see, e.g., 73, 159, 200, 226]. Relative to this line of work, we focus on the introduction of a downstream decision response, arising for example from operational constraints.

The case of constrained policies has been considered in the OPE literature. Our setting is conceptually different but overlaps in some application contexts. Specifically, we decouple the downstream *policy-dependent response*, i.e. over a similar constraint space, from treatment decisions that have causal effects. For example, Bhattacharya [37] studies the setting of "roommate assignment" with discrete types; i.e., perfect bipartite matching. A crucial difference is that in their setting, the causal treatment of interest *is* the assignment decision to the other individual type; while in our setting the causal treatment only affects certain *parameters* of the assignment decision, such as edge costs. We instantiate an analogous example in our framework to highlight our decoupled causal intervention and prediction decisions. Consider a setting with a causal intervention, such as a diversity information intervention affecting a student's probability of getting along with various types. Here the policy

[1]For terminology, we use policy-dependent (optimization) response or downstream policy-dependent response to refer to the same concept.

performs treatments on individual's diversity information, and the final assignment decision is policy-dependent response. Other work considers resource-budgeted allocation [139], which is structured because of reformulation of thresholds [156]. Sun [199] studies sharp asymptotics for the additional challenge of stochasticity in the budget.

Some works that illustrate the embedding of causal effect estimates in optimization-based decision problems include [184], although their formulation is ultimately a mixed-integer optimization.

In contrast to an extensive line of work on heterogeneous causal effect estimation [140, 193, 212], often crucially leveraging simpler structure of the treatment contrast rather than the conditional outcomes, in this work we require estimation of the latter due to the downstream optimization and distributional convergence for the perturbation method. In turn, combining causal outcome estimation with adjustments for optimization bias requires different properties of the estimation strategy, namely plug-in estimation of a modified regression model; we focus on estimators that modify the first-order conditions of a regression model to algebraically achieve an AIPW-type adjustment as discussed in Bang and Robins [25]. See also Chernozhukov et al. [50], Scharfstein et al. [192], Shi et al. [195], Tran et al. [205].

This work focuses on the challenge of *optimization bias* for policy evaluation introduced in our setting, for *generic* linear optimization problems. This well-known challenge of in-sample optimization bias ("sample average approximation bias") fundamentally demarcates the statistical regime of optimization under uncertainty from sample mean estimation [27, 194]. Recent work develops bagging, jackknife, perturbation and variance-corrected perturbation approaches for bias adjustment [106, 121, 131, 144]. We extend a perturbation method of Ito et al. [121] to the setting of nonlinear predictions.

6.2 Preliminaries

We first define the setting for off-policy evaluation with policy-dependent responses. We distinguish between the *causal decision policy* π and the *downstream optimization response* x. The causal decision policy π intervenes on individual units, while the policy-dependent responses are solutions to a downstream optimization problem on the causal responses of all the units.

1. Off-policy evaluation. We first describe the single time step off-policy policy evaluation and optimization problem [see 74, 114, for further context]. Let covariates be $W \in \mathcal{W} \subseteq \mathbb{R}^d$, binary treatment be $T \in \{0,1\}^2$, and potential outcomes be $c(T)$. Denote the covariates' distribution as $\mathcal{P}$. Without loss of generality we consider lower is better for c; e.g. we minimize costs. We consider a setting of learning causal responses from a dataset of tuples $\mathcal{D}_1 = \{(W_i, T_i, c_i)\}_{i=1}^n$ where treatment is assigned randomly or in an observational setting; henceforth we call this the *observational / experimental dataset*.

[2]The extension to non-binary treatments is immediate.

We let $\pi_t \colon \mathcal{W} \mapsto [0,1]$ denote a personalized policy mapping from covariates to a (probability of) treatment t. Later we will focus on parameterized policies, such as $\pi_t(w) = \mathit{sigmoid}(\varphi^\top w)$ or policies that admit global enumeration. The goal of off-policy learning is to optimize the causal interventions (aka policies) by estimating average outcomes induced by any given policy. Throughout we will follow the convention that, a random variable $c(\pi_t)$ denotes $c(\pi_t) = c(t)\,\mathbb{1}(Z_t = t)$, where $Z_\pi \in \{0,1\}$ is a Bernoulli random variable of policy assignment: $Z_\pi \sim \mathrm{Bern}(\pi_1)$. Then, the (random) outcome for a given covariate with policy π is:

$$c(\pi) = \textstyle\sum_t c(\pi_t) = \sum_t c(t)\,\mathbb{1}(Z_\pi = t).$$

The average treatment effect (ATE) of a policy $\pi(\cdot)$ is then $\mathbb{E}[c(\pi)]$, where the expectation is taken over the randomness of the covariates $W \sim \mathcal{P}$, assignments induced by π, and c conditional on realized treatment t and covariates.

2. Policy-dependent responses. *Policy-dependent optimization* solves a downstream stochastic linear optimization problem over a decision problem $x \in \mathcal{X} \subseteq \mathbb{R}^m$ on the m units given a causal intervention policy. In particular, m represents the dimension of the downstream decision problem. Relative to the downstream decision problem, causal outcomes may enter *either* as uncertain objective coefficients (in c) *or* constraint capacities (in b).[3]

Dimensionality of the responses. We consider two different asymptotic regimes: an *out-of-sample, fixed-dimension, fixed-m* regime and an *in-sample, growing-dimension, growing-n* regime. We formalize the former regime, the main focus of the paper, in the following assumption.

Assumption 6.2.1 (Out-of-sample, fixed-dimension regime)**.** As $n \to \infty$, the dimension of the optimization problem m, given by a new draw of contexts $\mathcal{D}_2 = \{W_i\}_{1:m}$ remains finite. The decision-dependent response on m units is measurable with respect to $\mathcal{D}_2$. Let $c_i(\pi) \stackrel{\text{def}}{=} \mathbb{E}[c(\pi(w)|w = W_i]$, we have that the policy value v^*_π is:

$$v^*_\pi = \mathbb{E}[\min_x \{\textstyle\sum_{i=1}^m c_i(\pi) x_i \colon Ax \leqslant b\}]. \tag{6.1}$$

Assumption 6.2.1 defines our *policy-dependent estimand* in this regime. The expectation is taken over the randomness of the policy π and the randomness of the finite samples $\{w_i\}_{i=1}^m$. The main text focuses on statements in the regime of Assumption 6.2.1. Evaluating regret with respect to a fixed dimension is standard or implicit in the predictive optimization literature.[4]

[3]Throughout the text we focus on uncertainty in c for notational clarity; strong duality implies the same results hold for uncertainty in the constraint right-hand-side, b. The decision is made conditionally on context information W but prior to realizations of potential outcomes, aka a policy-dependent response.

[4]The predictive optimization literature instead views each dimension of the decision variable as a multivariate outcome; relative to that, our regime can be interpreted as the setting of a scalar-valued contextual response.

Assumption 6.2.2 (In-sample, growing-dimension regime)**.** As $n \to \infty$, the limit of the objective function is an expectation over contexts.[5] The estimand is:

$$v_\pi^* = \mathbb{E}[\min_x \{\mathbb{E}[c(\pi)x] \colon Ax \leqslant b\}]. \tag{6.2}$$

Recall that a policy maps from covariates to a (probability of) treatment. Assumption 6.2.2 precisely takes an expectation over the two sources of randomness: the outer expectation is taken over the randomness of the policy, and the inner expectation is taken over the randomness of the covariates w.

The limiting object in the growing-dimension regime is a "fluid limit" or asymptotic regime: informally we assume a meaningfully constrained optimization in the limit. We instantiate our framework in the following example.

Example 6.2.3 (Min-cost bipartite matching)**.** Our framework is precisely motivated by the practical challenges in causal inference tasks, where the problem of "policy dependent" optimizations pops up repeatedly. For instance, for price incentives in a matching market (such as a rideshare platform), the revenue/welfare outcome is not realized until the riders and drivers are matched under constraints. As another example, consider a manager wants to assign agents to different jobs, and assigning an agent to a job is associated with some cost. Our goal is to assign each agent to at most one job such that the overall cost is minimized. To incentivize the workers to complete the jobs, the company might want to provide some bonus to the agents. However, the overall efficiency and total payments are not realized until all the assignments are determined.

The above type of application can be modeled as a min-cost bipartite matching problem, which is well known to have a totally unimodular linear relaxation. Clearly, the agents (or riders) and the jobs (or passenger requests) form the two sides of nodes for the matching. The edge costs in the matching stand for the cost or payment for an agent to complete that job. A treatment ($T = 1$) serves as intervention on the edge costs for that agent, and the covariates W could be any observable features of the agents, such as preferences, demographic information, etc. Given any allocation rule of the bonuses, the manager faces a downstream min-cost bipartite matching:

$$\min_{x \in \{0,1\}^{|\mathcal{E}|}} \left\{ \sum_{e \in \mathcal{E}} c_e(\pi) x_e \,:\, \sum_{e \in \mathcal{N}(i)} x_e = 1, \forall i \in \mathcal{V} \right\}. \tag{6.3}$$

Here $\mathcal{N}(i)$ is the set of all edges contains node i, the c_e are the edge costs, and $x = \{x_e\}_{e \in \mathcal{E}}$ represents the matching where $x_e = 1$ means that edge e is selected[6].

In Section 6.11 we include an additional example of predictive risk optimization, beyond linear optimization, which requires a different estimation strategy.

[5] Assume the constraint b scales with n in a meaningful problem-dependent way so that constraints are neither all slack nor infeasible in the limit.

[6] In the later analysis we use the linear relaxation with $x_e \in [0, 1]$ (continuous interval). For bipartite matching because of *total unimodularity* the linear relaxation is tight and equivalent to integral formulation.

3. Policy optimization with policy-dependent responses. Putting together the pieces of the previous subsections, the off-policy optimization over candidate policies $\pi \in \Pi$ is:

$$\min_{\pi\in\Pi} \min_{x\in\mathcal{X}} \left\{\sum_{i=1}^{m} c_i(\pi) x_i \colon Ax \leqslant b\right\}, \tag{6.4}$$

where m represents the dimension of the decision problem (e.g., the number of edges in Example 6.2.3), and x denotes the whole response vector $\{x_i\}_{i\in[m]}$.

We illustrate this framework by revisiting our examples.

Example 6.2.4 (Policy optimization for Example 6.2.3, min-cost matching)**.** In the min-cost bipartite matching example, the optimal assignments with a given policy π can be solved via the linear program in Equation (6.3). Suppose that we want to find the best intervention policy which gives the lowest matching cost. Then, the policy optimization problem is:

$$\min_{\pi\in\Pi} \min_{x\in\mathcal{X}} \left\{ \sum_{e\in\mathcal{E}} c_e(\pi) x_e : \sum_{e\in\mathcal{N}(i)} x_e = 1, \forall i; x_e \geqslant 0, \forall e \right\},$$

where Π denotes the set of all policies that are of interest.

6.3 Problem description: optimization bias

We focus on off-policy evaluation in view of the downstream optimization over the decision variables $x = \{x_i\}_{i\in[m]}$. We first discuss *plug-in* estimation approaches without causal adjustment to introduce the challenge of optimization bias in this regime. We then discuss causal estimation in Section 6.4.

From estimation bias to optimization bias. Denote $\mu_t(w) = \mathbb{E}[c(t) \mid W = w]$ as the conditional outcome mean of the population with treatment t and covariates w. We consider "predict-then-optimize" approaches which learn some $\hat{\mu}_t(w) = \mathbb{E}[c \mid W = w, T = t]$ and optimize with respect to it, so that our estimator is:

$$\hat{v}_\pi = \min_{x\in\mathcal{X}} \left\{\sum_{i=1}^{m} \sum_{t\in\{0,1\}} \pi_t(w_i)\hat{\mu}_t(w_i) x_i \colon Ax \leqslant b\right\}.$$

Note that due to the estimation and minimization step, $\hat{v}_\pi$ is not an unbiased estimator for v_π^*. Define the overall error of $\hat{v}_\pi$ with respect to the target estimand of Equation (6.1) as: $\text{err} = v_\pi^* - \mathbb{E}[\hat{v}_\pi]$. We decompose the overall error into two parts: the estimation bias of the plug-in estimator, and the optimization bias. Denote $\tilde{v}_\pi$, the best-in-class feasible estimate using the true conditional expectations μ_t^*:

$$\tilde{v}_\pi = \min_{x\in\mathcal{X}} \left\{\sum_{i=1}^{m} \sum_{t\in\{0,1\}} \pi_t(w_i)\mu_t(w_i) x_i \colon Ax \leqslant b\right\}.$$

Then, the estimation and optimization biases are:

(by triangle inequality, $|\text{err}| \leqslant |\text{bias}_{\text{est}}| + |\text{bias}_{\text{opt}}|$)

$$\text{bias}_{\text{est}} = \mathbb{E}[\hat{v}_\pi] - \mathbb{E}[\tilde{v}_\pi], \qquad \text{bias}_{\text{opt}} = v_\pi^* - \mathbb{E}[\tilde{v}_\pi].$$

<table>
<tr><th></th><th colspan="2">Out of sample, fixed m (Assumption 6.2.1)</th><th colspan="2">In-sample, growing n (Assumption 6.2.2, Section 6.10)</th></tr>
<tr><th></th><th>Evaluation</th><th>Policy optimization</th><th>Evaluation</th><th>Policy optimization</th></tr>
<tr><td>AIPW</td><td colspan="2">N/A</td><td>Sample splitting (finite VC-dim x)</td><td rowspan="3">Uniform generalization requires problem-dependent structure (finite VC-dim x)</td></tr>
<tr><td>WDM</td><td>Perturbation method</td><td rowspan="2">Uniform generalization from out-of-sample risk bounds</td><td>Perturbation</td></tr>
<tr><td>GRDR</td><td>Perturbation method</td><td>Perturbation
Doubly-robust estimation</td></tr>
</table>

Table 6.1: Summary of regimes and estimation properties. The main text provides methods for Assumption 6.2.1. Additional structural restrictions permit extensions for Assumption 6.2.2.

In-sample estimation bias due to optimization. It is well known that in-sample estimation of the value of optimization problems is biased; e.g., $\hat{v}$ is a biased estimate for the true objective value v_π^* due to optimization. Ito et al. [121] studies a bias correction for affine linear objectives with an unbiased estimate of a parameter θ. To understand the source of the bias due to optimization, observe that clearly $\sum_{i=1}^m \mu_t(w_i)x_i \geqslant \min_x \sum_{i=1}^m \mu_t(w_i)x_i$. The inequality remains valid when evaluating expectations over training datasets so that $\mathbb{E}[\sum_{i=1}^m \mu_t(w_i)x_i] \geqslant \mathbb{E}[\min_x \sum_{i=1}^m \mu_t(w_i)x_i]$. Noting that the RHS is the true objective v_π^*, we obtain in general the well-known optimistic bias, that $\mathbb{E}[\tilde{v}_\pi] \geqslant v_\pi^*$. In the policy evaluation setting, our estimates converge to the LHS, $\tilde{v}_\pi$, so that our estimator $\hat{v}_\pi$ is in general a *biased* estimate of the decision-dependent policy value even if we obtain *unbiased* estimates of the cost coefficient.

6.4 Causal estimation with policy-dependent responses

In this section we present an estimation approach building upon a perturbation method that adjusts for the aforementioned optimization bias. We summarize tradeoffs among estimation strategies in different regimes in Table 6.1 and possible extensions and additional structure in Section 6.10.

Estimating causal effects: estimation bias

Assumption 6.4.1 (Ignorability, overlap, SUTVA)**.** For all t, $c(t) \perp\!\!\!\perp T \mid W$. The evaluation policy is absolutely continuous with respect to treatment probabilities in the training dataset. Assume the stable unit treatment value assumption.

Confounding-adjusted plug-in estimators. In general, plug-in estimation of $\hat{\mu}_t(W)$ does *not* admit unbiased predictions because of selection bias and model misspecification. Existing importance-sampling based estimators, e.g. the inverse propensity weighting (IPW)

estimator and the doubly-robust augmented inverse probability weighting (AIPW) uses the propensity score to adjust confounding, under Assumption 6.4.1. Note importance sampling cannot *directly* be applied in our main regime of interest with out-of-sample evaluation as in Assumption 6.2.1, see Section 6.8 for a detailed overview.

We depart from previous work in off-policy evaluation, in view of the optimization bias adjustment (detailed in the next section), and study estimation methods that are *plug-in estimates* for OPE: $\mathbb{E}[c(\pi)] = \sum_t \mathbb{E}[\pi_t(W)\hat{\mu}_t(W)]$, for some outcome model $\hat{\mu}_t$ that is confounding-adjusted.

Note that IPW/AIPW-type estimators cannot be applied in the out-of-sample regime of Assumption 6.2.1. However, we may obtain out-of-sample risk bounds on the decision regret in this regime by virtue of out-of-sample generalization risk bounds on the generated regressors. We include more detailed discussion in Appendix 6.10.

Weighted direct method (WDM). Outcome regression, learning $\hat{\mu}_t(W) = \mathbb{E}[c \mid T = t, W]$ directly from $\mathcal{D}_1$, is sometimes called the *direct method*. However, when $\hat{\mu}$ is a misspecified regression model such a method incurs bias. Nonetheless, re-weighting the estimation $\hat{\mu}$ (maximum likelihood, empirical risk minimization) by the inverse probability weights $1/e$ is known to adjust for the covariate shift; by a similar argument as that of [45, 196, 215]. We call this approach *weighted direct method* (WDM), which solves:

$$\hat{\mu}_t^{\mathrm{WDM}} \in \arg\min_{\mu} \mathbb{E}\left[\tfrac{\mathbb{I}(T=t)}{e_t(W)}(c - \mu_t(W))^2\right]. \tag{6.5}$$

Doubly-robust direct method (GRDR). We also consider an approach that achieves doubly-robust estimation of the treatment-effect due to Bang and Robins [25]. [See also 192, 205]. This approach has been used for CATE estimation [50, 195]. The inverse propensity score reweighted treatment indicator is added as a covariate in the model, inducing coefficients ϵ_0, ϵ_1. Define

$$\hat{\mu}^{\mathrm{GRDR}} = \mu(W) + \epsilon_1(T/e_1(W)) + \epsilon_0((1-T)/e_0(W)).$$

Optimizing over $\hat{\mu}$ by (nonlinear) least-squares yields the following first-order optimality conditions for $\theta^{\mathrm{GRDR}} = [\bar{\theta}, \epsilon_1, \epsilon_0]$:

$$\mathbb{E}[(c-\hat{\mu})\nabla_\theta\hat{\mu}] = 0, \mathbb{E}[(c-\hat{\mu})(T/e_1(W))] = 0, \mathbb{E}[(c-\hat{\mu})((1-T)/e_0(W))] = 0. \tag{6.6}$$

Bang and Robins [25] show that the first-order optimality conditions ensure that plug-in estimation of an average treatment effect with the model is equivalent to AIPW, hence doubly-robust. Because it is designed primarily for estimation of the ATE, its use as an outcome predictor is more speculative. Although one can verify that its output is covariate-conditionally equivalent to CATE in expectation, and one can use this fact to again regress upon the pseudooutcomes, this final procedure would require re-verifying asymptotic convergence; we don't outline those arguments here. We include further discussion on the different estimation interpretations of GRDR in the two regimes in Appendix 6.10.

Estimating the decision-dependent estimand

Our procedure is adapted from the perturbation method of Ito et al. [121] which we describe here for completeness; we extend it from linear to nonlinear predictors. The method of Ito et al. [121] focuses on one parameter that we denote $\xi = [\theta, \gamma]$, where we assume as outlined in Assumption 6.4.4 that it encompasses parameters of the outcome and propensity model (respectively). Define the policy-induced outcome model, $\mu_\pi(w) = \sum_t \pi_t(w)\mu_t(w)$, the estimation error $\delta = \hat{\xi} - \xi^*$, and the (parametrized) optimal solution at a given predictive model $x(\xi)$. The perturbation method is motivated by a finite-difference approximation to the optimization bias induced by estimation error δ. Define the auxiliary functions given a scalar ϵ parametrizing the direction of δ:

$$\eta(\epsilon) = \mathbb{E}_\delta\left[\sum_{i=1}^m x(\xi^* + \epsilon\delta)\pi(W; \xi^*)\right], \quad \phi(\epsilon) = \mathbb{E}_\delta\left[\sum_{i=1}^m x(\xi^* + \epsilon\delta)\pi(W; \xi^* + \epsilon\delta)\right].$$

We require regularity conditions for derivatives of these functions to exist:

Assumption 6.4.2 (Perturbation method assumptions)**.** (i) The optimal solution $x(\xi)$ is unique. (ii) $\hat{\xi}$ is an unbiased estimator of ξ^*.

We generalize Prop. 3 of Ito et al. [121] for nonlinear models.

Proposition 6.4.3. *We have* $\eta(\epsilon) = \phi(\epsilon) - \epsilon\phi'(\epsilon) + O(\epsilon^2)$.

The plug-in estimated optimal value $\hat{v}_\pi$ unbiasedly estimates $\phi(1)$. Note $\phi'(1)$ is equivalent to the value of the bias. The perturbation method estimates $\phi'(1)$ by $(\phi(1+h)-\phi(1))/h$ for some small h.

It remains to estimate $\phi(1+h)$. First we obtain s samples of the perturbed parameter $\hat{\xi}_h = \xi^* + (1+h)\delta$, denoted as $\{\hat{\xi}_h^{(j)}\}_{j=1}^s$. Each replicate of $\hat{\xi}^{(j)}$ leads to an optimization estimate $\hat{v}^{(j)} = \min_x \sum_{i=1}^m x_i \hat{\mu}_\pi(w_i, \hat{\xi}_h^{(j)})$. The debiased estimator is:

$$\rho_h = \hat{v}^{(0)} - \tfrac{1}{h}(\hat{v}^{(0)} - \tfrac{1}{s}\sum_{j=1}^s \hat{v}^{(j)})$$

Our Proposition 6.4.3 then implies asymptotic unbiasedness (cf. Prop. 4 of Ito et al. [121]) so that $\lim_{h\to 0}\mathbb{E}[\rho_h] = \mathbb{E}[\min_x \sum_{i=1}^m x_i\hat{\mu}_\pi(w_i, \xi^*)]$. We summarize the method in Algorithm 11.

Asymptotic variance of estimation methods. We discuss the asymptotic variance of the *weighted direct method* and GRDR via classical asymptotic analysis of *generated regressors* (specifically, stacked estimation equations of GMM) [177]. We summarize this framework in Section 6.8 for completeness and include the main result here that we invoke.[7]

[7]Asymptotic normality of these approaches is taken as given in Bang and Robins [25], Cao et al. [45] and so we include these statements for completeness. For exposition and context of Donsker-type conditions in semiparametric inference, see Kennedy [134] or other references.

Algorithm 11 Perturbation method, Alg. 2 of [121])

Input: Estimation strategy $\diamond \in \{\text{WDM}, \text{GRDR}\}$; h: finite different parameter; π: policy.
Estimate $\hat{\xi}_\diamond = [\hat{\theta}_\diamond, \hat{\gamma}_\diamond]$ for $\hat{\mu}^\diamond$ from $\mathcal{D}_1$
$\hat{v}^{(0)} \leftarrow \min_{x \in \mathcal{X}} \sum_{i=1}^m x_i \sum_{t \in \{0,1\}} \pi_t(w_i) \hat{\mu}_t^\diamond(w_i; \hat{\xi}_\diamond)$
Generate $\{\xi_\diamond^{(j)}\}_{j=1}^s$: if by parametric bootstrap, learn $\hat{\xi}_\diamond^{(j)}$ from $\frac{N}{(1+h)^2}$ samples randomly chosen from $\mathcal{D}_1$ with replacement.
Otherwise if using $\hat{\Sigma}$, estimator of asymptotic variance of ξ, approximate the distribution of $\xi^* + (1+h)\delta$. Add $\hat{\xi}$ to $\hat{\theta}$ where $\hat{\delta} \sim N(0, \frac{(1+h)^2-1}{N}\hat{\Sigma})$. Then set $\hat{\xi}_\diamond^{(j)} = \hat{\xi} + \hat{\delta}_j$.
for $j = 1, \ldots, S$: **do**
    $\hat{v}^{(j)} \leftarrow \min_{x \in \mathcal{X}} \sum_{i=1}^m x_i \sum_{t \in \{0,1\}} \pi_t(w_i) \hat{\mu}_t^\diamond(w_i; \hat{\xi}_\diamond^{(j)})$.
end for
Output $\rho_h = \hat{v}_0 - \frac{1}{h}(\hat{v}^{(0)} - \frac{1}{s}\sum_{j=1}^s \hat{v}^{(j)})$.

Assumption 6.4.4 (Estimators via GMM with generated regressors)**.** Suppose the propensity score e and outcome model μ are indexed by true parameters γ^*, θ^* that solve the respective estimating equations $\mathbb{E}[h(W, \gamma^*)] = 0, \quad \mathbb{E}[g(W, \theta^*, \gamma^*)] = 0$. The functions $e_t(w), \mu_t(w)$ are in a Donsker class.

Remark 6.4.5 (Strength of assumptions)**.** *Algorithm 11 requires both unbiased and asymptotically normal predictions—stronger conditions than merely inference on the ATE. The Donsker assumption preserves asymptotic normality with generated regressors. The framework allows for nonparametric estimation via linear sieves (but not some high-dimensional regimes; see Ackerberg et al. [3]).*

Theorem 6.4.6 (Thm. 6.1, eq. 6.12 of Newey and McFadden [177])**.** *Suppose Assumption 6.4.4 holds. Let $\hat{G}_\alpha, \hat{G}_\theta, \hat{H}$ denote the Jacobian matrices of partial derivatives of the moment conditions g, h with respect to the respective parameters, i.e. $\hat{G}_\gamma = n^{-1}\sum_{i=1}^n \nabla_\gamma g(w_i, \hat{\theta}, \hat{\gamma})$. Let $\hat{V}_\gamma = (\hat{H}^{-1}\hat{h}_i)(\hat{H}^{-1}\hat{h}_i)^\top$. Then an estimator of the asymptotic variance is:*
$\hat{V}_\theta = \hat{G}_\theta^{-1}\left(n^{-1}\sum_{i=1}^n \hat{g}_i\hat{g}_i^\top\right)(\hat{G}_\theta^{-1})^\top + \hat{G}_\theta^{-1}\hat{G}_\gamma\hat{V}_\gamma\hat{G}_\gamma^\top(\hat{G}_\theta^{-1})^\top$.

Since $\hat{V}_\gamma$ depends only on the specification of the propensity score, to completely specify the asymptotic variance for the above formula we state the mixed terms $\hat{G}_\gamma, \hat{G}_\theta$.

Proposition 6.4.7 (Asymptotic normality of WDM)**.** *Let $e_t(w), \mu_t(w)$ satisfy Assumption 6.4.4 with the moment condition $g_t(W, \theta, \gamma) = e_t(W; \gamma)^{-1}(c - \mu_t(W; \theta))^2$ and $g = [g_0, g_1]$. Then*

$$\hat{G}_\gamma = \begin{bmatrix} \mathbb{E}_n[2T(c - \mu(W;\theta))\frac{\partial}{\partial\theta}(e_1^{-1}(W,\gamma))\frac{\partial\mu}{\partial\theta}] \\ \mathbb{E}_n[2(1-T)(c - \mu(W;\theta))\frac{\partial}{\partial\theta}(e_0^{-1}(W,\gamma))\frac{\partial\mu}{\partial\theta}] \end{bmatrix}.$$

Algorithm 12 Subgradient method for policy optimization

1: **Input:** step size η, linear objective function f.
2: **for** $j = 1, 2, \cdots$ **do**
3: At φ^k, obtain a subgradient in subdifferential $\mathcal{S}^*(\pi_\varphi^k) = \{x^* : f(x^*; \pi_\varphi^k) = \min_x f(x; \pi_\varphi^k)\}$
4: Compute subgradient $\nabla_\varphi(\min_x f(x; \pi_\varphi^k)) \leftarrow \nabla_\varphi f(x^*; \pi_\varphi)$
5: Update subgradient step: $\varphi^{k+1} \leftarrow \varphi^k - \eta \nabla_\varphi \left(\min_x f(x; \pi_\varphi^k)\right)$
6: **end for**

These formulas are generally computable from standard output of optimization solvers for nonlinear least squares: gradients and Hessians. In practice, using the parametric bootstrap may be simpler at a higher computational cost.

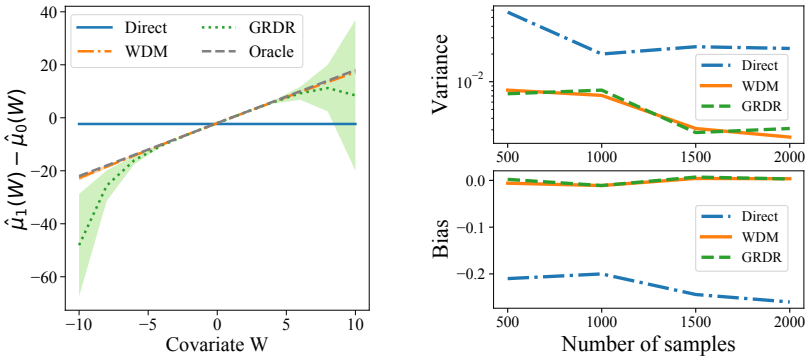
Optimizing Causal Interventions

Algorithm 11 provides estimation for a fixed policy. We now discuss how to optimize over policies; e.g., implementing the outer optimization over policies $\min_{\pi \in \Pi}$ in Equation (6.4). We focus on the case where the policy $\pi_t(w)$ is parametrized by and differentiable in a parameter $\varphi \in \Psi$. For example, for the logistic policy parameterization, $\pi_t(w) = sigmoid(\varphi_t^\top w)$. We consider a robust subgradient method, based on Danskin's theorem, detailed in Algorithm 12. Such an approach is a common heuristic used in adversarial machine learning.

We solve the inner optimization problem to full optimality in line 3 and take (sub)gradient steps for the outer optimization. We evaluate (sub)gradients of the inner optimization solution in line 3 by evaluating the gradient of the objective with respect to φ, fixing the inner optimization variable x^*. Danskin's theorem implies that ∇_φ is a subgradient [60]. The inner minimization can be solved via a linear optimization oracle for any fixed choice of policy. This use of the linear optimization oracle can be beneficial when special problem structures, such as matching and network flows, may also admit readily-available algorithmic solutions to full optimization.

The perturbation method is compatible with our optimization procedure because the bias-adjusted perturbation estimated from Algorithm 11 is affine in the optimization problems corresponding to each parameter replicate. Hence, run Algorithm 11 with an expanded linear objective over the s-product space $x' \in \mathcal{X}^s$ where $f(\tilde{x}, \pi) = \hat{v}_\pi^{(0)}(\tilde{x}_0) - \frac{1}{h}(\hat{v}_\pi^{(0)}(\tilde{x}_0) - \frac{1}{s}\sum_{j=1}^s \hat{v}_\pi^{(j)}(\tilde{x}_j))$.

So, re-optimize $\tilde{x}_j^* \in \arg\min_{x \in \mathcal{X}} \sum_{i=1}^m x_i \hat{\mu}_\pi^\diamond(w_i; \hat{\xi}_\diamond^{(j)})$ and apply Danskin's theorem to each optimization problem in the sum over $\hat{v}_\pi^{(j)}$ comprising $f(x', \pi)$. In fact, though adversarial machine learning focuses on min-max rather than our min-min optimization problem, this particular approach is simply subgradient descent on a nonconvex function (the solution to the inner optimization).



(a) Effect of W (b) Effect of dataset size

Figure 6.1: *(In-sample estimation of $\hat{\mu}_1(W) - \hat{\mu}_0(W)$, with model mis-specification).* Comparison of direct / WDM / GRDR to the oracle. (a) Conditional estimation error averaged over ten random train sets; shaded area indicates std. error. (b) Bias / variance comparison with varying training data size.

6.5 Experimental evaluations

Since real data suitable for both policy evaluation and downstream optimization is unavailable, we focus on synthetic data and downstream bipartite matching. We first illustrate estimation properties of the different approaches before showing the improvement obtained via policy optimization. Though we are not aware of prior approaches that are directly comparable for optimizing causal policy with a downstream optimization-dependent response, we include more comparisons to nonparametric estimators (e.g. causal forests [212]), and full implementation details.

1. Causal effect estimation. First, we investigate and illustrate the properties of different estimators. We generated dataset $\mathcal{D}_1 = \{(W, T, c)\}$ with covariate $W \sim \mathcal{N}(0, 1)$, confounded treatment T, and outcome c. Treatment is drawn with probability

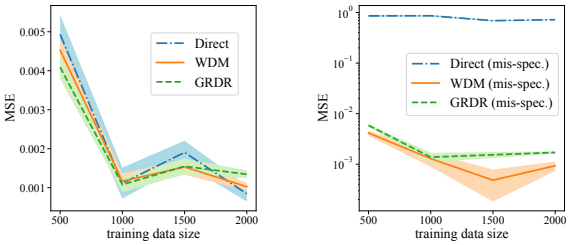
$$\pi_t^b(W) = (1 + e^{-\varphi_1 W + \varphi_2})^{-1}$$

, $\varphi_1 = \varphi_2 = 0.5$. The true outcome model is given by a degree-2 polynomial,[8]

$$c_t(w) = poly_\theta(t, w) + \epsilon$$

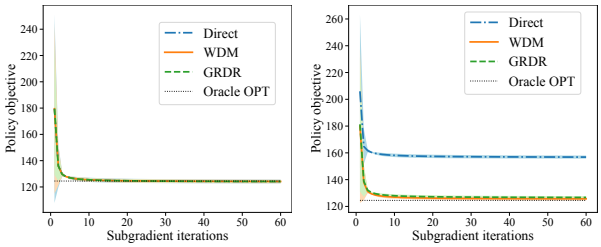
, where $\epsilon \sim \mathcal{N}(0, 1)$. . In Figure 6.1a and 6.1b, we illustrate the (covariate-conditional) estimation error of the three estimators. In the mis-specified setting that induces confounding, the outcome model is a vanilla linear regression over W without the polynomial expansion. The direct method results in more bias under mis-specification, while WDM and GRDR are robust as expected.

[8] If not stated otherwise we spread the coefficients as $poly_\theta(t, w) = (1, w, t, w^2, wt, t^2) \cdot ([5, 1, -1, 2, 2, -1])^\top$. Additional supporting experiments under other nonlinear data-generating processes are in Appendix 6.12.



(a) well-specified models (b) with mis-specification

Figure 6.2: *(Policy evaluation via perturbation method (Algorithm 11).* Comparison of direct / WDM / GRDR estimators over increasing size of training data (averaged over ten runs).



(a) well-specified models (b) with mis-specification

Figure 6.3: *(Policy optimization).* Subgradient policy optimization with direct / WDM / GRDR estimation methods and a fixed test set. Averaged over ten random training datasets of size=1000.

2. Policy evaluation. We compare the perturbation method (Algorithm 11) with three different estimators (direct, WDM, and GRDR). In both the well-specified / mis-specified model setting, we evaluate the mean-squared-error (MSE) of the estimated policy value with the three estimators, where the MSE is computed with regard to the ground-truth outcome model. Training data size n increases from 500 to 2000 samples. We scaled the MSE down by the number of edges (a constant) and computed the MSE in terms of the averaged cost per edge in the matching.

For the policy-dependent optimization, we evaluate a min-cost bipartite matching problem, where the causal policy intervene on the edge costs (as detailed in Example 6.2.3). Specifically, the bipartite graph contains $m = 500$ left side nodes $W_1, \cdots, W_m$, and $m' = 300$ right side nodes. The policy π_t applies treatments to the left side nodes and the outcome is the edge cost of edges with that node. While we grow the training data size, we fixed m, m' (with $m > m'$) and evaluate over ten random draw of train/test data for each value of n. Figure 6.2 plots the results. When there is mis-specification, even a large training dataset cannot bring bias correction for the direct method, where both WDM and GRDR enjoy smaller and decreasing MSE.

We also conduct an ablation study for the corresponding performance in the mis-specified setting (i.e., no bootstrapping in Alg. 11). Results indicate that the perturbation method is helpful for MSE reduction for both WDM and GRDR. We further conduct evaluations with different bootstrap replicates' sizes, and the above conclusions remain robust for different replicate sizes (additional results in Appendix 6.12).

3. Policy optimization. Lastly, we integrate the policy evaluation and the sub-gradient method (Alg 12) to conduct policy optimization. At each iteration of Alg 12, the perturbation algorithm (Alg 11) and one of the three different estimation methods are applied to evaluate

the policy objective. We consider a logistic policy $\pi_t(W) = sigmoid(\varphi_1 \cdot W + \varphi_2)$. To study the convergence and the effectiveness of the subgradient algorithm for minimization, we fix a test set and perform subgradient descent over 60 iterations for each run. We average the policy values at each iteration over ten runs, where in each run we generate a random set of training data and a random initialization of the starting policy.

We compare to the oracle estimator using the ground truth outcome model (Oracle OPT). Results are presented in Figure 6.3. Again, WDM and GRDR quickly converge to the oracle estimation, while the large bias of the direct method leads to poor policy optimization. We further evaluate the impact of average random selected initial policies to the performance, and compared Figure 6.3 with the results using a fixed initial policy. We observe that in this relatively low-dimensional example, the policy value converges to estimation-oracle-optimal after a few iterations (additional results and full training details in Appendix 6.12).

6.6 Conclusion and future directions

We studied a new framework for causal policy optimization with a *policy-dependent* optimization response. We proposed evaluation algorithms and analysis to address the fundamental challenge of an additional optimization bias. Simulations for both the policy evaluation and optimization algorithms demonstrate the effectiveness of this approach. Interesting further directions include studying individual fairness of optimal allocations in applications such as school assignments or job matching, and/or computational improvements to the policy optimization algorithms.

6.7 Appendix: Further related works and comparisons

Other variants of off-policy evaluation and optimization with global dependence on the population. There is also a line of work on off-policy evaluation, for example evaluating policies by non-average functionals over populations (median [148], quantile [46, 182]); but estimation crucially depends on specific reformulations based on problem structure. These functionals are typically risk deviations and reformulated in relation to estimation of quantiles rather than generic optimization formulations. In contrast, our debiasing is more general and independent of the functional form of the risk deviation beyond the second-stage problem being a stochastic linear optimization problem. However, risk deviations typically lead to *min-max* problems, while our formulation has aligned objective functions of treatment and response and is inherently a partial optimization of a nonconvex problem, hence a *min-min* problem.

Such min-max formulations also appear in sensitivity analysis and approaches to unobserved confounding via min-max robustness [125, 127] and similar algorithms for policy optimization appear there. However, the formulation of this work focuses specifically on

generic optimization problems in a different conceptual setting. Estimation is quite different due to the requirement of compatibility of estimation and a perturbation approach.

Our focus on the downstream global system response bears a distant conceptual resemblance to interference [117], but is fundamentally very different: we require the *causal* response satisfies the stable unit treatment value assumption (SUTVA), while the source of interaction across units (analogous to the exposure mapping) *is completely known to* and *under the control of* the decision-maker.

Lastly, recent work develops specialized estimation more closely using the structure of economic systems, such as marketplace interference [151] or mean-field equilibrium [169, 213]. This work is often closely coupled with the application structure. Our use of policy-dependent structure is coarse, considering only a generic linear optimization response and in turn adjusting for the introduced optimization bias.

Decision-dependent classifier shift. Our last example of estimating decision-dependent predictive loss is broadly motivated by decision-dependent distribution shifts but focuses on settings with distinct treatments, rather than other frameworks where a classifier is a treatment studying model-based or utility-model based approaches [109, 180].

Structured prediction. Finally, although there is extensive work on structured prediction in machine learning, our framework is very different: while structured prediction maps contextual inputs to the space of complex outputs (the space of network flows, matchings; the optimization decision vector), our data environment consists of contexts and separable, unit-dependent outcomes.

6.8 Appendix: Additional details for estimation

Preliminaries

Adjusting for confounding: IPW and AIPW. In general, plug-in estimation of $\hat{\mu}_t(W)$ does *not* admit unbiased predictions because of selection bias and model misspecification. A key object that adjusts for selection bias is the

$$\textit{propensity score: } e_t(W) = \mathbb{P}(T = t \mid W).$$

Although importance sampling cannot *directly* be applied in our main regime of interest with out-of-sample evaluation as in Assumption 6.2.1, we introduce key properties which can be used to debias outcome models. (See Section 6.10 for discussion of an alternative in-sample OPE regime).

Inverse propensity weighting (IPW) transforms treatment-conditional expectations to the population expectation—by iterated expectations and Assumption 6.4.1 (ignorability), we have:

$$\sum_t \mathbb{E}\left[\mathbb{E}\left[c\frac{\mathbb{I}[Z_\pi = t]}{e_t(W)} \mid W\right]\right] = \sum_t \mathbb{E}\left[c\frac{\mathbb{I}[Z_\pi = t]}{e_t(W)}\right] = \sum_t \mathbb{E}[c(\pi_t)] = \mathbb{E}[c(\pi)]. \tag{6.7}$$

In general, *doubly-robust augmented inverse probability weighting (AIPW)* estimation improves the variance when both models are well-specified and achieves overall unbiasedness under unbiasedness of either outcome or propensity:

$$\sum_t \mathbb{E}\left[\pi_t(W)\mathbb{I}[Z_\pi = t](\tfrac{\mathbb{I}[T=t]}{e_t(W)}(c - \mu_t(W)) + \mu_t(W))\right] = \mathbb{E}[c(\pi)].$$

6.9 Appendix: Proofs

Asymptotic variance of two-step estimation via GMM asymptotic variance We first recall a general framework for deriving asymptotic variance with generated regressors as discussed in [177]. This section is summarized from the presentation in there to keep derivations self-contained.

We focus on the approach based on the asymptotic variance for GMM, viewing the nuisance estimations as "stacked" moment equations for $\hat{\theta}$ (second-stage estimate) and $\hat{\gamma}$ (the first-stage estimation). Then, applying blockwise inversion to the GMM asymptotic variance obtains the asymptotic variance of the (parameter) vector, $\begin{bmatrix}\theta\\ \gamma\end{bmatrix}$ in terms of the first component (top left block).

Stack the moment equations for $\hat{\theta}, \hat{\gamma}$ to get:

$$\begin{aligned}\mathbb{E}[g(W, \theta_0, \gamma_0)] &= 0,\\ \mathbb{E}[h(W, \gamma_0] &= 0.\end{aligned}$$

Note adaptivity occurs iff $G_\gamma = \nabla_\gamma \mathbb{E}[g(W, \theta_0, \gamma_0)] = 0$

Define the following sample Jacobians of moment conditions with respect to the parameters:

$$\hat{G}_\theta = n^{-1}\sum_{i=1}^{n} \nabla_\theta g(w_i, \hat{\theta}, \hat{\gamma}), \qquad \hat{G}_\gamma = n^{-1}\sum_{i=1}^{n} \nabla_\gamma g(w_i, \hat{\theta}, \hat{\gamma}), \qquad \hat{H} = n^{-1}\sum_{i=1}^{n} \nabla_\gamma h(w_i, \hat{\gamma}).$$

For short introduce notation for the empirical moments $\hat{g}_i, \hat{h}_i$:

$$\hat{g}_i = g(w_i, \hat{\theta}, \hat{\gamma}), \qquad \hat{h}_i = h(w_i, \hat{\gamma}),$$

so that we can define the sample second moment matrix $\hat{\Omega}$ as follows:

$$\hat{\Omega} = n^{-1}\sum_{i=1}^{n} (\hat{g}_i, \hat{h}_i)^\top (\hat{g}_i, \hat{h}_i).$$

The estimator for asymptotic variance is given by:

$$\hat{V} = \begin{bmatrix}\hat{G}_\theta & \hat{G}_\gamma\\ 0 & \hat{H}\end{bmatrix}^{-1} \Omega^{-1} \begin{bmatrix}\hat{G}_\theta & \hat{G}_\gamma\\ 0 & \hat{H}\end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \hat{G}_\theta^{-1} & -\widehat{G}_\theta^{-1}\hat{G}_\gamma\hat{H}^{-1} \\ 0 & \hat{H}^{-1} \end{bmatrix} \widehat{\Omega} \begin{bmatrix} \hat{G}_\theta^{-1} & -\widehat{G}_\theta^{-1}\widehat{G}_\gamma\hat{H}^{-1} \\ 0 & \hat{H}^{-1} \end{bmatrix}.$$

If the moment functions are uncorrelated then the first-step estimation increases the second-step variance. We may obtain a simplification as follows. Let $\hat{\phi}_i = -\hat{H}^{-1}\hat{h}_i$. Then the upper left block is

$$\hat{V}_\theta = \widehat{G}_\theta^{-1}\left[n^{-1}\sum_{i=1}^{n}\left\{\hat{g}_i + \hat{G}_\gamma\hat{\psi}_i\right\}\left\{\hat{g}_i + \widehat{G}_\gamma\hat{\psi}_i\right\}'\right]\widehat{G}_\theta^{-1}.$$

For $\hat{V}_\gamma = n^{-1}\sum_{i=1}^{n}\hat{\phi}_i\hat{\phi}_i'$, an asymptotic variance estimator for $\hat{\theta}$ is

$$\hat{V}_\theta = \hat{G}_\theta^{-1}\left(n^{-1}\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'\right)(\hat{G}_\theta^{-1})' + \hat{G}_\theta^{-1}\hat{G}_\gamma\hat{V}_\gamma(\hat{G}_\gamma)'(\hat{G}_\theta^{-1})'.$$

Outcome regressions with generated regressors

In the appendix we also discuss an approach based on GRDR .

Proposition 6.9.1 (Asymptotic normality of GRDR)**.** *Let $e_t(w), \mu(w)$ satisfy Assumption 6.4.4 with the moment condition for μ given by Equation* (6.6)*. Then*

$$\hat{G}_\gamma = \begin{bmatrix} \mathbb{E}_n[2(\epsilon_1(\frac{\partial}{\partial\gamma}e_1^{-1}(W;\gamma))) + \epsilon_0(\frac{\partial}{\partial\gamma}e_0^{-1}(W;\gamma)))\frac{\partial\mu}{\partial\tilde{\theta}}] \\ -\mathbb{E}_n[2T(c-\mu(W;\theta))(\frac{\partial}{\partial\gamma}e_1^{-1}(W;\gamma))] \\ -\mathbb{E}_n[2(1-T)(c-\mu(W;\theta))(\frac{\partial}{\partial\gamma}e_0^{-1}(W;\gamma))] \end{bmatrix}.$$

Proof of Proposition 6.4.7.

$$\mathbb{E}_n[\nabla_\gamma g(W;\theta,\beta)] = \mathbb{E}_n\left[T(\frac{\partial e}{\partial\gamma}e^{-1})\cdot 2(c-\mu(W;\beta))\frac{\partial\mu}{\partial\tilde{\theta}}\right].$$

■

Proof of Proposition 6.9.1, GRDR*, [25].* The stacked estimation equations are as follows, for the parameters $[\gamma, \theta, \epsilon_1, \epsilon_0]$:

$$\begin{aligned} &-\mathbb{E}\left[2(T-e_T(W;\gamma))\frac{\partial e}{\partial\gamma}\right] = 0, \\ &-\mathbb{E}\left[2(c-\tilde{\mu}(W;\theta)\frac{\partial\mu}{\partial\theta}\right] = 0, \\ &-\mathbb{E}[2T(c-\tilde{\mu}_1(W;\theta)e_1^{-1}(W;\gamma)] = 0, \\ &-\mathbb{E}[2(1-T)(c-\tilde{\mu}_0(W;\theta)e_0^{-1}(W;\gamma)] = 0 \end{aligned}$$

and the Jacobian of partial derivatives is:

$$G_\gamma = \begin{bmatrix} \mathbb{E}[2(\epsilon_1(\frac{\partial}{\partial\gamma}e_1^{-1}(W;\gamma))) + \epsilon_0(\frac{\partial}{\partial\gamma}e_0^{-1}(W;\gamma)))\frac{\partial\mu}{\partial\theta}] \\ -\mathbb{E}[2(c-\mu(W;\theta))T(\frac{\partial}{\partial\gamma}e_1^{-1}(W;\gamma))] \\ -\mathbb{E}[2(c-\mu(W;\theta))(1-T)(\frac{\partial}{\partial\gamma}e_0^{-1}(W;\gamma)] \end{bmatrix}.$$

■

Nonlinear Generalization of Perturbation Method

Preliminaries. We include the proof for completeness. It is the same argument of Ito et al. [121] with the addition of linear expansions of the nonlinear model μ around the parameter ξ.

Let μ denote a generic prediction model which may depend nonlinearly upon its parameter ξ. Define for our context the true-optimal-decision $x(\xi^*)$, and sample-optimal-decision $\hat{x}(\hat{\xi})$:

$$x^* \in \arg\max \sum_{i=1}^m \mu^*(W_i, \xi^*)x_i,$$
$$\hat{x} \in \arg\max \sum_{i=1}^m \mu(W_i; \hat{\xi})x_i.$$

Define the auxiliary functions η, ϕ evaluated along paths indexed by ϵ:

$$\eta(\epsilon) = \mathbb{E}_\delta\left[\sum_{i=1}^m x(\xi^* + \epsilon\delta)\mu(W_i;\xi^*)\right],$$
$$\phi(\epsilon) = \mathbb{E}_\delta\left[\sum_{i=1}^m x(\xi^* + \epsilon\delta)\mu(W_i;\xi^* + \epsilon\delta)\right].$$

We focus on the case exclusively where the function f of interest is affine, i.e. so that $f(z^*, \xi^*) = \sum_{i=1}^m z_i^* g(\xi^*; X_i)$.

Proof of Proposition 6.4.3. Let $\xi_\epsilon^* = \xi^* + \epsilon\delta$. First, we will show

$$\eta(\epsilon) - \phi(\epsilon) = \epsilon\,\mathbb{E}_\delta\left[\sum_{i=1}^m \hat{x}_i(\nabla_\xi\mu\Big|_{\xi_\epsilon}\delta)\right] + O(\epsilon^2). \tag{6.8}$$

and then that $\epsilon\phi'(\epsilon)$ equals the right-hand-side of the above.

Step 1 (Showing Equation (6.8)):

Expand the definition of η, ϕ and apply a Taylor expansion of $\mu(W_i;\xi^*+\epsilon\delta)$ from $\mu(W_i;\xi^*)$.

$$\eta(\epsilon) - \phi(\epsilon) = \mathbb{E}_\delta\left[\sum_{i=1}^m \left(\hat{x}_i\mu(W_i;\xi^*) - \hat{x}_i\mu(W_i;\hat{\xi})\right)\right] = \mathbb{E}_\delta\left[\sum_{i=1}^m \hat{x}_i\left(\mu(W_i;\xi^*) - \mu(W_i;\hat{\xi})\right)\right]$$

$$= \mathbb{E}_\delta \left[\sum_{i=1}^{m} \hat{x}_i \left(\epsilon (\nabla_\xi \mu \Big|_{\xi_\epsilon} \delta) + O(\|\epsilon\delta\|_2^2) \right) \right].$$

Step 2: $\epsilon\phi'(\epsilon) =$ RHS of Equation (6.8).
Let $\xi^*_{\epsilon+h} = \xi^* + (\epsilon + h)\delta$.
By definition,

$$\begin{aligned}
\phi'(\epsilon) &= \lim_{h \to 0} \frac{\phi(\epsilon + h) - \phi(\epsilon)}{h} \\
&= \lim_{h \to 0} \frac{1}{h} \left(\mathbb{E}_\delta \left[\sum_{i=1}^{m} x_i(\xi^*_{\epsilon+h}) \mu(W_i; \xi^*_{\epsilon+h}) - \sum_{i=1}^{m} x_i(\xi^*_\epsilon) \mu(W_i; \xi^*_\epsilon) \right] \right).
\end{aligned}$$

Add / subtract $\hat{x}(\xi_\epsilon)\mu(W_i; \xi_\epsilon)$:

$$\begin{aligned}
\phi'(\epsilon) &= \lim_{h \to 0} \left\{ \frac{1}{h} \left(\mathbb{E}_\delta \left[\sum_{i=1}^{m} \left(x_i(\xi^*_{\epsilon+h}) \mu(W_i; \xi^*_\epsilon) + x_i(\xi^*_{\epsilon+h})(\mu(W_i; \xi^*_{\epsilon+h}) - \mu(W_i; \xi^*_\epsilon)) \right) \right] \right) - \frac{1}{h} \left(\mathbb{E}_\delta \left[\sum_{i=1}^{m} x_i(\xi^*_\epsilon) \mu(\right. \right. \right. \\
&= \lim_{h \to 0} \left\{ \frac{1}{h} \left(\mathbb{E}_\delta \left[\sum_{i=1}^{m} (x_i(\xi^*_{\epsilon+h}) - x_i(\xi^*_\epsilon)) \mu(W_i; \xi^*_\epsilon) + \sum_{i=1}^{m} (x_i(\xi^*_{\epsilon+h})(\mu(W_i; \xi^*_{\epsilon+h}) - \mu(W_i; \xi^*_\epsilon)) \right] \right) \right\} \\
&= \lim_{h \to 0} \left\{ \frac{1}{h} \left(\mathbb{E}_\delta \left[\sum_{i=1}^{m} (x_i(\xi^*_{\epsilon+h}) - x_i(\xi^*_\epsilon)) \mu(W_i; \xi^*_\epsilon) + \sum_{i=1}^{m} (x_i(\xi^*_{\epsilon+h})(\mu(W_i; \xi^*_{\epsilon+h}) - \mu(W_i; \xi^*_\epsilon)) \right] \right) \right\}.
\end{aligned}$$

The last line follows by a Taylor expansion of μ from ξ^*_ϵ to $\xi^*_{\epsilon+h}$ and noting that the first term converges as x does, $\lim_{h \to 0} \frac{1}{h} (\sum_{i=1}^{m} (x_i(\xi^*_{\epsilon+h}) - x_i(\xi^*_\epsilon)) \mu(W_i; \xi^*_\epsilon)) = 0$. under regularity conditions common in perturbation analysis of stochastic programs, such as uniqueness of the solution.

Therefore, interchanging limits and the expectation:

$$\begin{aligned}
\phi'(\epsilon) &= \mathbb{E}_\delta \left[\lim_{h \to 0} \left\{ \frac{1}{h} \left(\sum_{i=1}^{m} x_i(\xi^*_{\epsilon+h}) (\nabla_\xi \mu \Big|_{\xi_\epsilon} (h\delta) + O(\|h\epsilon\|_2^2)) \right) \right\} \right] \\
&= \mathbb{E}_\delta \left[\lim_{h \to 0} \left\{ \sum_{i=1}^{m} x_i(\xi^*_{\epsilon+h}) (\nabla_\xi \mu \Big|_{\xi_\epsilon} \delta + O(h) \right\} \right] \\
&= \mathbb{E}_\delta \left[\sum_{i=1}^{m} x_i(\xi^*_\epsilon) (\nabla_\xi \mu \Big|_{\xi_\epsilon} \delta) \right].
\end{aligned}$$

■

6.10 Appendix: Alternative asymptotic regime (Assumption 6.2.2)

In the main text we focused on a fixed-dimension regime. We describe some extensions that may be possible to handle an in-sample, growing dimension, growing-n regime described in Assumption 6.2.2. We do generally require additional structural information to apply more familiar OPE estimators such as IPW/AIPW and other adaptations of bias adjustment methods.

The strongest such additional structural knowledge is that the optimization is highly structured so as to admit a finite VC dimension; to circumvent issues related to the growing dimension.

Assumption 6.10.1. $x(\pi, W)$ has finite VC dimension.

Such a structural characterization is established in special cases such as multi-knapsack linear programs [206], or large-market limits of stable matching markets [16]; but need not hold in general.

Preliminaries

For completeness, we describe the analogous estimands/estimators for the *in-sample, growing-n* regime as described in the main text for the out-of-sample, fixed-m regime.

Plug-in estimation is evaluated on the training dataset as follows:

$$\hat{v}_\pi = \min_{x \in \mathcal{X}} \left\{ n^{-1} \sum_{i=1}^{n} \sum_{t \in \{0,1\}} \pi_t \hat{\mu}_t(w_i) x_i \colon Ax \leqslant b \right\}. \tag{6.9}$$

We describe extensions of approaches to handle in-sample bias in this growing-dimension setting, although we generally require more structure on the problem. As described in Assumption 6.2.2 we typically require a problem-dependent asymptotic scaling; for example that we jointly scale up the problem size as well as the constraints. We provide a concrete example for Example 6.2.3.

Example 6.10.2 (Fluid limit for Example 6.2.3)**.** The number of workers is αn and the number of jobs is βn.

Sample splitting

We first discuss an analogous sample splitting extension of Ito et al. [121] which combines their sample splitting procedure with standard cross-fitting for doubly robust estimators [49]. However, a naive extension requires four folds and is therefore expected to perform poorly in finite samples.

Let K denote the number of folds, we will exposit the case of two folds and the K-fold generalization is standard. Denote two main folds of the data, $\mathcal{I}_{k_1}, \mathcal{I}_{k_2}$ denoting the index sets for the data, and $\mathcal{I}_{k_{1e}}, \mathcal{I}_{k_{1\mu}}$ be subfolds of $\mathcal{I}_{k_1}$ (respectively subfolds for $\mathcal{I}_{k_2}$). As suggested by the notation, we use distinct subfolds $\mathcal{I}_{k_{1e}}, \mathcal{I}_{k_{1\mu}}$ to learn the nuisance estimates e, μ (from the respective subfold).

The main difference from standard cross-fitting is that in Assumption 6.10.1 we assume the optimization problem is well-parametrized in covariates: the optimization solution is well-described as a function of $x(W)$. We also require that there is a sensible way of sampling datapoints and projecting the feasible set $\mathcal{X}$ onto each subsampled index set. E.g. when subsampling in a matching example, after subsampling nodes the new feasible set in each index set $\mathcal{X}_1, \mathcal{X}_2$ preserves all edges between nodes in the original feasible set $\mathcal{X}$.

Let $\Gamma_t(O_i; e, \mu)$ denote the score associated with observation $O_i = (W_i, T_i, c_i)$ under either IPW or AIPW, with input nuisance functions e, μ. For example, as in Section 6.4, $\Gamma_t^{\text{IPW}}(O; e, \mu) = \frac{\mathbb{I}[T=t]c}{e_t(w)}$ and $\Gamma_t^{\text{AIPW}}(O; e, \mu) = \frac{\mathbb{I}[T=t](c-\mu_t(w)}{e_t(w)} + \mu_t(w)$.

As is standard in cross-fitting we use distinct main folds in order to estimate nuisances (indexed by parameters ξ) for input into $\hat{x}$: that is,

$$\hat{x}_1(\hat{\xi}_2; W) \in \arg\min_{x(W)\in\mathcal{X}_2} \frac{1}{|\mathcal{I}_2|} \sum_{i\in\mathcal{I}_2} \sum_{t\in\{0,1\}} \pi_t(W_i)\Gamma_t(O_i; e^{-k(i)}, \mu^{-k(i)})x_i,$$

and analogously for $\hat{x}_2$.

Then evaluation estimates the value within each fold using the optimal solution from the other fold:

$$\hat{v}_1 = \frac{1}{|\mathcal{I}_2|} \sum_{i\in\mathcal{I}_2} \sum_{t\in\{0,1\}} \pi_t(W_i)\Gamma(O_i; e^{-k(i)}, \mu^{-k(i)})\hat{x}_1(\hat{\xi}_1, W_i),$$

and we return the average over folds, $\frac{1}{2}(\hat{v}_1 + \hat{v}_2)$ or more generally the average of $\{\hat{v}_k\}_{k\in K}$.

When $K > 2$, for each fold k, we will use two subfolds $I_{-k,e}$ and $I_{-k,\mu}$ to estimate e, μ, and then obtain $\hat{x}_{-k}$ from I_{-k}. We evaluate the estimated objective with $\hat{x}_{-k}$, $\hat{e}_k$ and $\hat{\mu}_k$ and average over all folds.

Proposition 6.10.3 (Unbiased estimation by sample splitting.)**.** $\frac{1}{K}\sum_i^K \hat{v}_k = \mathbb{E}[\tilde{v}]$

Proof. Immediate from standard analysis of AIPW and sample-splitting of [121]. ■

Comparison of estimation properties in the two regimes (Table 6.1)

Out-of-sample, fixed-dimension. IPW/AIPW-type estimators cannot be applied in the out-of-sample regime of Assumption 6.2.1, by definition of the regime. However, we may obtain out-of-sample risk bounds on the decision regret in this regime, simply by virtue of out-of-sample generalization risk bounds on the generated regressors. For example, we effectively assume near-parametric regimes for the propensity score so that the conditions of Theorem 1

of Bertail et al. [34], providing a generalization risk bound for two-stage reweighted empirical risk minimization with estimated weights (as in our Section 6.4), are met. Under assumption of uniformly bounded decision variables, applying the Cauchy-Schwarz inequality directly implies that statistical estimation consistency of our estimation approaches imply decision regret consistency, so that the estimation bias vanishes at a $O_p(n^{-\frac{1}{2}})$ rate. (However the statistical rate of optimization bias adjustment remains unclear).

In-sample, growing-dimension, growing-n. An analogous extension to sample splitting as in Ito et al. [121] is possible in highly structured situations satisfying Assumption 6.10.1. For a *fixed* optimization solution x, uniform generalization over $\pi \in \Pi$ is a consequence of uniform generalization with a stochastic (bounded) envelope function. However, in this regime, uniform generalization over both $\pi \in \Pi$ and $x \in \mathcal{X}$ is difficult because in the regime of Assumption 6.2.2, the dimension of the optimization grows as $n \to \infty$. Typical approaches to uniform convergence would require $x^*(\pi, W)$ (the optimal optimization solution at a fixed π) to converge uniformly over the space of policies and W.

Different estimation interpretations of GRDR **in the two regimes.** Note that benefits of GRDR in terms of doubly-robust estimation of the ATE (mixed-bias, rate double-robustness) are only relevant in the *in-sample regime* of Assumption 6.2.2. Recent work does show this specification obtains empirical benefits for confounded outcome estimation, in appeal to the sufficient balancing properties of the propensity score, that may also apply to the regime of Assumption 6.2.1.

6.11 Appendix: Beyond linearity: decision-dependent classifier risk

The downstream optimization can also in turn be a prediction risk problem: treatments shift distributions upon which predictive risk models are trained [179]. For example, the medical system simultaneously treats individuals but is also interested in large-scale predictive models from passively collected electronic health records, trained upon the realizations of health outcomes of the entire population, and so may generate distribution shifts in these predictive risk models [6, 85]. Therefore, the post-treatment predictive risk model introduces a downstream causal-policy dependent optimization response.

In the previous sections, we focused on linear optimization because plug-in estimation is consistent when the random variable enters linearly into the optimization problem. The challenge with nonlinearity is that such plug-in-approaches are no longer consistent and can introduce policy-dependent nuisance estimation functions.

Nonetheless, special structure of the problem can admit alternative estimation strategies.

Problem setup

Example 6.11.1 (Decision-dependent classifier drift.)**.** We shift to notation more typical in statistics/machine learning to emphasize the setting. We model decision-dependent shift of predictive risk models in a repeated measurement setting.[9] Our observation trajectories[10] each comprise of $(L_0, Y_0, T_0, L_1(T_0), Y_1(T_0))$: baseline covariates (L_0, Y_0), time-0 treatment $T_0 \in \{0, 1\}$, and post-treatment covariates and outcome $(L_1(T_0), Y_1(T_0))$.

For example, L could measure patient state and Y a cardiac event within a given time period. Upon observation of (L_0, Y_0) a patient is treated with T_0; for example with more aggressive or wait-and-see treatment depending on $Y_0(T_0)$. While optimal treatment regimes focus on averages of individual-level outcomes $Y_1(T_0)$; in our policy-dependent response setting we model the problem of, for example, continuously monitoring "feedback loops" that may surface in predictive risk models that may generally be trained using large electronic health record databases. Said differently, we could have modeled this abstractly as a policy evaluation with the augmented set of "covariates", jointly (L_0, Y_0), and "outcomes" $(L_1(T_0), Y_1(T_0))$. However, we focus on the ultimate downstream predictive risk model which depends (nonlinearly) on all the outcomes of the population's units. The policy evaluation problem could evaluate the predictive loss of the downstream predictive model $f(L_1(\pi), \beta)$, where there is downstream optimization of the squared loss over β.

$$\min_{\beta} \mathbb{E}[(Y_1(\pi) - f(L_1(\pi), \beta))^2]. \tag{6.10}$$

The causal graph of Section 6.11 describes the two-stage observation of individuals[11], comprising observation trajectories $(L_0, Y_0, T_0, L_1(T_0), Y_1(T_0))$. We consider in the general case a two-stage setting with a treatment affecting covariates and outcomes; upon which a predictive risk model is trained.

Example 6.11.2 (Policy optimization for Example 6.11.1, policy-dependent prediction.)**.** Consider the case of two different treatments with similar (conditional) average treatment effects, but one induces higher variability in outcomes which increases the fundamental noise level in the regression: harming the population prediction model and incurring higher loss. Optimizing between these two treatments, scalarizing population outcomes by the global term would result in choosing the less variable treatment.

Therefore, in this framework we may be interested in the following scalarized policy optimization problem:

$$\min_{\pi} \min_{\beta} \lambda \, \mathbb{E}[Y_1(\pi)] + (1 - \lambda) \, \mathbb{E}[(Y_1(\pi) - f(L_1(\pi), \beta))^2].$$

[9] That is, observing covariates and outcomes from the same unit, measured at different time periods.

[10] We are therefore modeling "feedback loops" between outcomes Y_0 and the treatments administered to manage them via temporally distinct repeated measurements.

[11] We do not consider for now the causal effects of the prediction model, although this could be a direction for future work.

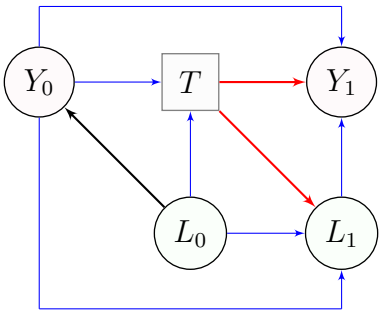


Figure 6.4: Causal diagram for decision-dependent classifier drift.

Estimation

Plug-in estimators. Our goal is off-policy evaluation of predictive risk model (parameter) solving the downstream prediction risk optimization after treatment under treatment regime π:

$$\beta^* \in \arg\min_{\beta} \mathbb{E}[(Y_1(\pi) - f(L_1(\pi); \beta)^2].$$

However for the special case that we consider of linear regression response (squared loss, linear parametrization), we may orthogonalize the *first-order optimality* conditions of the policy-dependent optimization, e.g. recognizing that, β^* solves the first order conditions

$$\mathbb{E}[L_1^\top (Y_1 - L_1\beta)] = 0. \tag{6.11}$$

Hence for the case of least squares and linear regression, we may focus on estimation refinements for β: observe that the estimation requires estimation of certain transformations of (X_1, Y_1) unit's downstream outcome, $x_1 y_1$ and $x_1^\top x_1$, and we may estimate the following matrix-regression or vector-valued regression nuisance estimates:

$$\mathbb{E}[L_1 Y_1 \mid T = t, X_0, Y_0], \qquad \mathbb{E}[L_1 L_1^\top \mid T = t, L_0, Y_0]$$

Hence standard AIPW-type approaches can be applied with the above nuisances and the censored observations $l_1 y_1(t), l_0 y_0(t)$.

This suggests that when β is our parameter of interest (or functions thereof), we can leverage double robustness. And, if we have a small space of policies we can optimize by enumeration. However the same challenges regarding nonlinearity remain if we want to estimate the final squared loss of θ.

Remark 6.11.3 (Restriction to linear models and the challenge for generalization to nonlinear models)**.** *Note that the challenge with generalizing to non-least-squares losses or nonlinear predictors is that due to nonlinearity, doubly-robust estimation of outcome X_1 need not provide the same benefits of bias reduction. Although an alternative approach is to instead estimate the squared loss as the composite outcome $\mathbb{E}[(Y_1(t) - \theta^\top L_1(t))^2 \mid t, L_0, Y_0]$ because of our policy-dependent response optimizing over θ, we would have policy-dependent nuisance functions so this becomes intractable.*

6.12 Appendix: Additional experiment details and results

In this section, we provide more details on the experimental setup as well as further results.

Causal effect estimation setup

Fr the causal effect estimation we generated the training dataset $\mathcal{D}_1 = \{(W, T, c)\}$ with covariate $W \sim \mathcal{N}(0, 1)$, confounded treatment T, and outcome c. Treatment is drawn with probability $\pi_t^b(W) = \frac{1}{1+e^{-\varphi_1 W+\varphi_2}}$, $\varphi_1 = \varphi_2 = 0.5$. The true outcome model is given by a degree-2 polynomial:

$$poly_\theta(t, w) = (1, w, t, w^2, wt, t^2) \cdot ([5, 1, -1, 2, 2, -1])^\top.$$

We generate the outcome samples as $c_t(w) = poly_\theta(t, w) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$. All random samples are generated using `numpy.random` package. In the mis-specified setting that induces confounding, the outcome model is a vanilla linear regression over W without the polynomial expansion.

In Fig. 6.1a and Fig. 6.1b in the main text, we illustrate the (covariate-conditional) estimation over the covariates' landscape for the direct method, the weighted direct method (WDM), and the doubly robust method (GRDR) when there is a model mis-specification. We provide the estimation results without model mis-specification in Fig. 6.5.

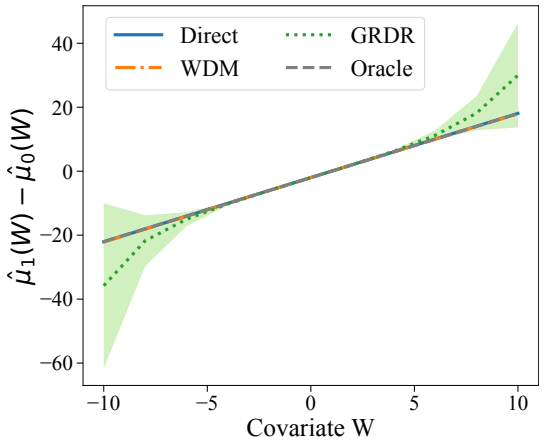


Figure 6.5: *(In-sample estimation of $\hat{\mu}_1(W) - \hat{\mu}_0(W)$, no model mis-specification).* Comparison of direct / weighted direct (WDM) / doubly robust method (GRDR) to the oracle estimator for estimation of conditional ATE over different covariate values. Results are averaged over ten random training datasets; shading area indicates the standard error.

When there is no model mis-specification that that induces confounding, we observe that both the three estimation methods perform well against the oracle estimation.

Policy evaluation

For policy evaluation, we compare the perturbation method (Algorithm 11) when being applied with three different estimators (direct, WDM, and GRDR). For consistency, throughout the evaluations, we follow the same true outcome model and covariate distribution as in the previous subsection for causal effect estimation. In both the well-specified model setting and the mis-specified model setting, the mean-squared-error (MSE) of the estimated policy value with the three estimators is computed with regard to the ground truth outcome model (aka oracle).

When evaluating Algorithm 11, we generated $S = 20$ bootstrap replicates. The downstream matching problem is evaluated with $m = 500$ left-hand-side nodes, and $m' = 300$ right-hand-side nodes. The min-cost matching requires each node to be matched to no more than one node on the other side, and was computed by the `linear_sum_assignment` function of the `scipy.optimize` package in Python 3. We evaluated a fixed logistic policy $\pi_t(W) = sigmoid(\phi \cdot W + b)$ with $\phi = 1, b = 0.5$.

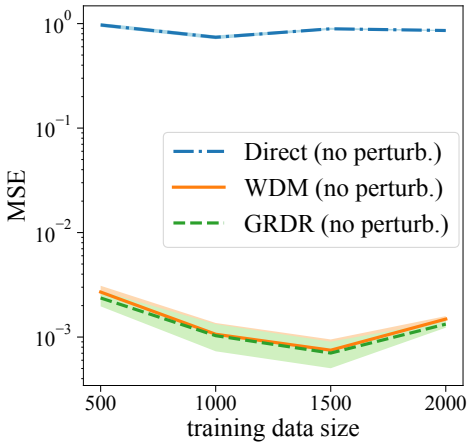


Figure 6.6: *(In-sample estimation of $\hat{\mu}_1(W) - \hat{\mu}_0(W)$ with model mis-specification, no perturbation applied).* Comparison of direct / weighted direct (WDM) / doubly robust method (GRDR) over increasing size of training data. Results are averaged over ten random training datasets; shading area indicates the standard error.

Figure 6.2 shows that when there is mis-specification, even a large training dataset cannot bring bias correction for the direct method, where both WDM and GRDR enjoy smaller and decreasing MSE. As an ablation study, we also compare to the corresponding performance in the mis-specified setting when we do not perform the perturbations (i.e. no bootstrapping in Alg. 11). In detail, we directly return $\hat{v}^{(0)}$ without doing the later bootstrap procedure. Figure 6.6 indicates that the perturbation method is helpful for MSE reduction for both WDM and GRDR.

We further conduct evaluations with different bootstrap replicates' sizes (controlled by variable h in Algorithm 11). We include these results in Table 6.2. Results in Table 6.2 show

Table 6.2: *(Perturbation method, varying replicate size.)* Performance for different estimator/model combinations. Mean-squared-errors (MSE) are computed with regard to the oracle outcome model.

	Estimation	$h = 1$	$h = 2$	$h = 3$	$h = 4$
Mis-specified model	Direct	0.59±0.05	0.70±0.05	0.76±0.06	0.70±0.06
	WDM	0.00031±0.0001	0.00042±0.0002	0.00041±0.0002	0.00048 ±0.0003
	GRDR	0.00040±0.0002	0.00046±0.0002	0.00031±0.0001	0.00035±0.0001
Well-specified model	Direct	0.00079±0.0004	0.00067±0.0004	0.00062±0.0003	0.00024±0.0002
	WDM	0.00076±0.0004	0.00067±0.0003	0.00080±0.0002	0.00031±0.0001
	GRDR	0.00082±0.0002	0.00067±0.0002	0.00080±0.0002	0.00031±0.0001

that WDM and GRDR remain more superior and that is robust with different replicate sizes. For the evaluations over different h values, we used training data with 3000 samples. In each iteration, the number of bootstrap replicates is 20.

Policy optimization

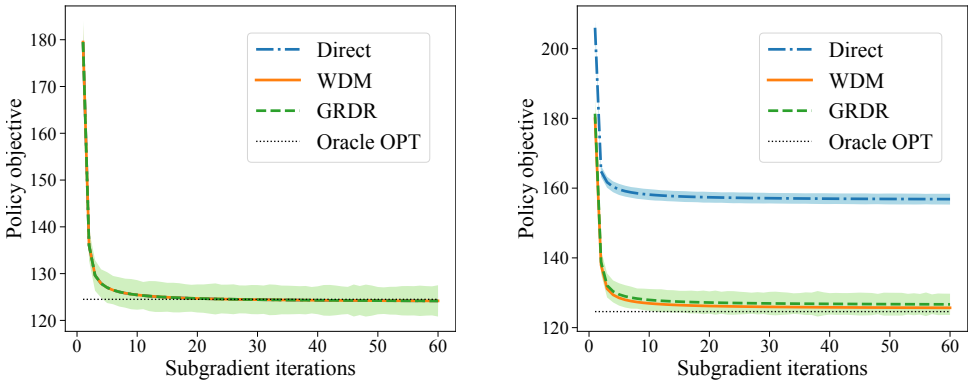
For policy optimization, we implemented the subgradient method as in Algorithm 12, and obtained causal effect estimators from Algorithm 11.

In detail, for a given training dataset, we first obtained $S + 1$ outcome estimators (i.e. $\{\hat{\mu}_t^{\diamond}(w_i; \hat{\xi}_{\diamond}), \hat{\mu}_t^{\diamond}(w_i; \hat{\xi}_{\diamond})^{(j)}, j = 1 \cdots S\}$ in Algorithm 11) via bootstrap. Then, at each iteration of running subgradient descent, we evaluate the current policy using the $S + 1$ outcome estimators respectively, and obtain $S + 1$ subgradients of it. We then aggregate these subgradients by the bootstrap aggregation (as in Step 6, Algorithm 11).

We evaluate subgradients of the inner optimization solution in Algorithm 12 (step 4) by evaluating the gradient of the objective with respect to φ, fixing the inner optimization variable x^*. The fact that ∇_{φ} is a subgradient is a consequence of Danskin's theorem [60]. The inner minimization (the matching problem) is again solved by the `linear_sum_assignment` function of the `scipy.optimize` package in Python 3.

To further study the impact of the random initial policies to begin with the subgradient descent algorithm, in Figure 6.7 we obtained the corresponding results of Figure 6.3, but with a fixed initial policy. We observe that again WDM and GRDR quickly converges to the oracle estimation, while the large bias of the direct method leads to poor policy optimization. Moreover, in this relatively low-dimension example, although random initialization of the policy leads to a larger variance in earlier iterations, the policy value convergences to oracle policy objective quickly after a few iterations.

For the evaluations of policy optimization, we used training datasets with size 1000, and a downstream min-cost matching with $m = 100, m' = 60$. The learning rate was tuned over $[0.01, 0.1, 1]$. All of our evaluations were run on a 2.3 GHz 8-Core Intel Core i9 CPU. All the differentiation operations were handled by the automatic differentiation library in `JAX`.



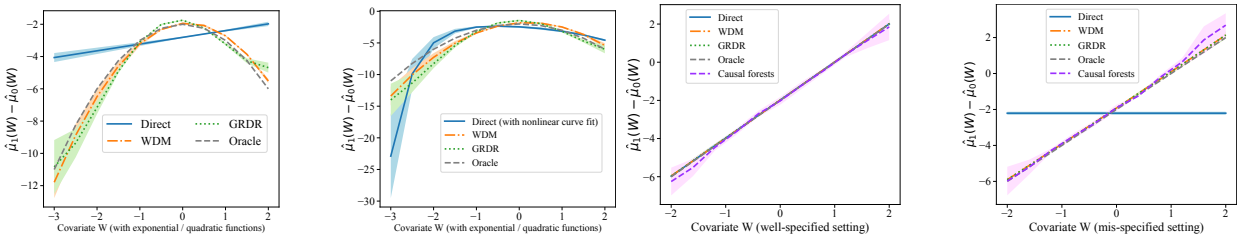
(a) Policy optimization with well-specified models

(b) Policy optimization with model mis-specification

Figure 6.7: *(Policy optimization (fixed test set)).* Results of subgradient policy optimization with direct / weighted direct (WDM) / doubly robust (GRDR) estimation methods and a fixed test set. Averaged over ten random training datasets of size 1000.

Additinal comparisons and evaluations

Additional evaluations with more complex non-linear outcome model. We conduct further robustness checks with nonlinear data-generating processes: exponential and quadratic. The outcome is $c_t(w) = a_1 + a_2 \exp(b_1 + b_2 w) + c_1 t + c_2 t w^2 + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$ is an external noise, and $[a_1, a_2, b_1, b_2, c_1, c_2] = [5, 0.05, 0.5, -2, -2, -1]$. We fit this function with nonlinear least squares (*scipy.optimize.curve_fit*). Indeed, Figure 6.8 shows that the direct method is sensitive to the model mis-specification bias without using a near-specified nonlinear curve fit. However the weighted direct method (WDM) and the doubly robust estimator (GRDR) remain robust; even if starting with misspecified parametric models.



(a) (b) (c) (d)

Figure 6.8: (In-sample estimation of $\hat{\mu}_1(W) - \hat{\mu}_0(W)$ for exponential function, (a) without / (b) with curve fit. Comparisons of the CATE estimates with nonparametric estimators. (c) without / (d) with model mis-specification.

Additional comparison to non-parametric estimators. We further compare our estimators against existing CATE estimators (although in general, these estimators tuned to estimate contrasts may improve upon differencing outcome models). We compared the DM/WDM/GRDR to the Causal Random Forests estimator proposed in (Wager and Athey, 2018) [12], following the setup in Section 5.1. Moreover, we also compared how the CATE estimators affect the policy evaluation task with the perturbation method (section 5.2). In Figure6.8(c,d) comparing CATE estimates, the non-parametric random forest estimator is indeed unbiased, while the naive direct method has a large bias under model mis-specification.

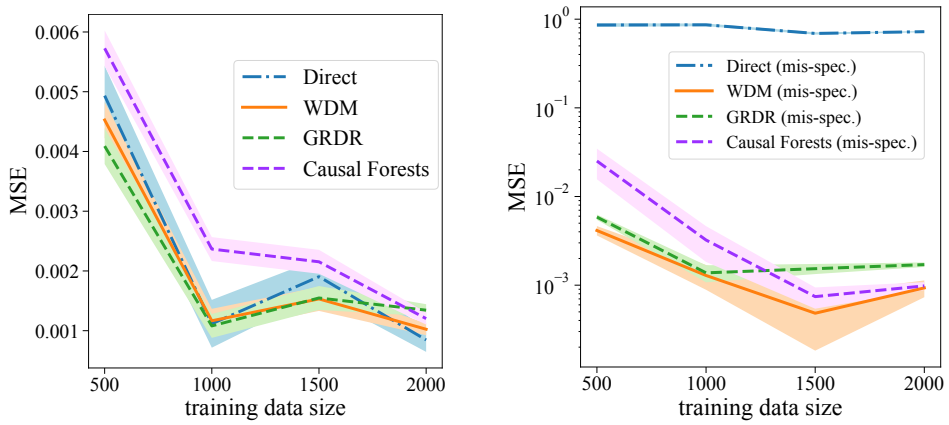


Figure 6.9: *(Policy evaluation via perturbation method (Algorithm 1).* Comparison of direct / WDM / GRDR / Causal Forests estimators over increasing size of training data.

[12]Based on implementation by Battocchi et al, *EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation.*

Bibliography

[1] R. Abebe, S. Barocas, J. Kleinberg, K. Levy, M. Raghavan, and D. G. Robinson. Roles for computing in social change. In *Conference on Fairness, Accountability, and Transparency (FAccT)*, 2020.

[2] M. Abramowitz, I. A. Stegun, and R. H. Romer. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 1988.

[3] D. Ackerberg, X. Chen, and J. Hahn. A practical asymptotic variance estimator for two-step semiparametric estimators. *Review of Economics and Statistics*, 94(2):481–498, 2012.

[4] A. Agarwal, A. Beygelzimer, M. Dudík, J. Langford, and H. Wallach. A reductions approach to fair classification. In *International Conference on Machine Learning (ICML)*, 2018.

[5] S. Aghaei, M. J. Azizi, and P. Vayanos. Learning optimal and fair decision trees for non-discriminative decision-making. In *AAAI Conference on Artificial Intelligence*, volume 33, pages 1418–1426, 2019.

[6] D. Agniel, I. S. Kohane, and G. M. Weber. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *Bmj*, 361, 2018.

[7] S. Agrawal and N. Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory*, pages 39–1, 2012.

[8] G. M. Amdahl. Computer architecture and amdahl's law. *Computer*, 46(12):38–46, 2013.

[9] M. Anthony and P. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 2009.

[10] A. Arefeva and D. Meng. Revealing information in auctions: The optimal auction versus the second-price auction. *Economics Letters*, 204:109895, 2021.

[11] N. Arnosti, M. Beck, and P. Milgrom. Adverse selection and auction design for internet display advertising. *American Economic Review*, 106(10):2852–66, 2016.

[12] S. Athey and I. Segal. An efficient dynamic mechanism. *Econometrica*, 81(6):2463–2485, 2013.

[13] S. Athey, S. Wager, et al. Efficient policy learning. Technical report, 2017.

[14] P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

[15] P. Awasthi, M. Kleindessner, and J. Morgenstern. Equalized odds postprocessing under imperfect group information. In *Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

[16] E. M. Azevedo and J. D. Leshno. A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268, 2016.

[17] M. Babaioff, R. Kleinberg, and A. Slivkins. Multi-parameter mechanisms with implicit payment computation. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, pages 35–52, 2013.

[18] M. Babaioff, N. Nisan, and I. Talgam-Cohen. Fair allocation through competitive equilibrium from generic incomes. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 180–180, 2019.

[19] M. Babaioff, N. Nisan, and I. Talgam-Cohen. Competitive equilibrium with indivisible goods and generic budgets. *Mathematics of Operations Research*, 46(1):382–403, 2021.

[20] M. Bagnoli and T. Bergstrom. Log-concave probability and its applications. In *Rationality and Equilibrium*, pages 217–241. Springer, 2006.

[21] M.-F. Balcan, T. Sandholm, and E. Vitercik. Sample complexity of automated mechanism design. *arXiv preprint arXiv:1606.04145*, 2016.

[22] M.-F. Balcan, T. Sandholm, and E. Vitercik. A general theory of sample complexity for multi-item profit maximization. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 173–174, 2018.

[23] S. R. Balseiro and Y. Gur. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

[24] D. L. Bandalos. *Measurement Theory and Applications for the Social Sciences*. Guilford Publications, 2017.

[25] H. Bang and J. M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

[26] S. Barocas, M. Hardt, and A. Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2019. `http://www.fairmlbook.org`.

[27] G. Bayraksan and D. P. Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108(2):495–514, 2006.

[28] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, October 2009.

[29] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[30] S. Bera, D. Chakrabarty, N. Flores, and M. Negahbani. Fair algorithms for clustering. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 4954–4965, 2019.

[31] D. Bergemann and M. Pesendorfer. Information structures in optimal auctions. *Journal of Economic Theory*, 137(1):580–609, 2007.

[32] D. Bergemann, T. Heumann, and S. Morris. Selling impressions: Efficiency vs. competition. 2021.

[33] R. Berk, H. Heidari, S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, S. Neel, and A. Roth. A convex framework for fair regression. *arXiv preprint arXiv:1706.02409*, 2017.

[34] P. Bertail, S. Clémençon, Y. Guyonvarch, and N. Noiry. Learning from biased data: A semi-parametric approach. In *International Conference on Machine Learning*, pages 803–812. PMLR, 2021.

[35] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[36] A. Beygelzimer, J. Langford, L. Li, L. Reyzin, and R. Schapire. Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 19–26. JMLR Workshop and Conference Proceedings, 2011.

[37] D. Bhattacharya. Inferring optimal peer assignment from experimental data. *Journal of the American Statistical Association*, 104(486):486–500, 2009.

[38] S. Brânzei, N. Devanur, and Y. Rabani. Proportional dynamics in exchange economies. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 180–201, 2021.

[39] J. Brustle, Y. Cai, and C. Daskalakis. Multi-item mechanisms without item-independence: Learnability via robustness. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 715–761, 2020.

[40] E. Budish, G. P. Cachon, J. B. Kessler, and A. Othman. Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation. *Operations Research*, 65(2):314–336, 2017.

[41] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Journal of Machine Learning Research (JMLR)*, 2018.

[42] B. Burns, B. Grant, D. Oppenheimer, E. Brewer, and J. Wilkes. Borg, omega, and kubernetes. *ACM Queue*, 14:70–93, 2016. URL `http://queue.acm.org/detail.cfm?id=2898444`.

[43] Y. Cai and C. Daskalakis. Extreme-value theorems for optimal multidimensional pricing. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 522–531. IEEE, 2011.

[44] Y. Cai and C. Daskalakis. Learning multi-item auctions with (or without) samples. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 516–527. IEEE, 2017.

[45] W. Cao, A. A. Tsiatis, and M. Davidian. Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96 (3):723–734, 2009.

[46] Y. Chandak, S. Niekum, B. C. da Silva, E. Learned-Miller, E. Brunskill, and P. S. Thomas. Universal off-policy evaluation. *arXiv preprint arXiv:2104.12820*, 2021.

[47] K. Chen, I. Hu, Z. Ying, et al. Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *Annals of Statistics*, 27(4):1155–1163, 1999.

[48] L. Chen, S. Liu, B. Li, and B. Li. Scheduling jobs across geo-distributed datacenters with max-min fairness. *IEEE Transactions on Network Science and Engineering*, 6(3): 488–500, 2018.

[49] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

[50] V. Chernozhukov, W. K. Newey, V. Quintas-Martinez, and V. Syrgkanis. Riesznet and forestriesz: Automatic debiased machine learning with neural nets and random forests. *arXiv preprint arXiv:2110.03031*, 2021.

[51] K. J. Cohen and R. M. Cyert. Theory of the firm; resource allocation in a market economy. Technical report, Prentice-Hall, 1965.

[52] R. Cole and T. Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, pages 243–252, 2014.

[53] A. Cotter, H. Jiang, and K. Sridharan. Two-player games for efficient non-convex constrained optimization. In *International Conference on Algorithmic Learning Theory (ALT)*, 2019.

[54] A. Cotter, H. Jiang, S. Wang, T. Narayan, S. You, K. Sridharan, and M. R. Gupta. Optimization with non-differentiable constraints with applications to fairness, recall, churn, and other goals. *Journal of Machine Learning Research (JMLR)*, 20(172):1–59, 2019.

[55] K. Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43:1241, 1990.

[56] B. Crépon and G. J. Van Den Berg. Active labor market policies. *Annual Review of Economics*, 8:521–546, 2016.

[57] B. Crépon, E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *The Quarterly Journal of Economics*, 128(2):531–580, 2013.

[58] S. Crockett, S. Spear, and S. Sunder. Learning competitive equilibrium. *Journal of Mathematical Economics*, 44(7-8):651–671, 2008.

[59] V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. 2008.

[60] J. M. Danskin. The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–664, 1966.

[61] M. A. Davenport, R. G. Baraniuk, and C. D. Scott. Tuning support vector machines for minimax and Neyman-Pearson classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[62] A. De Corniere and R. De Nijs. Online advertising and privacy. *The RAND Journal of Economics*, 47(1):48–72, 2016.

[63] G. Debreu. Existence of competitive equilibrium. *Handbook of Mathematical Economics*, 2:697–743, 1982.

[64] C. Delimitrou and C. Kozyrakis. Paragon: Qos-aware scheduling for heterogeneous datacenters. *ACM SIGPLAN Notices*, 48(4):77–88, 2013.

[65] N. R. Devanur, Z. Huang, and C.-A. Psomas. The sample complexity of auctions with side information. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, pages 426–439, 2016.

[66] N. Dikkala and É. Tardos. Can credit increase revenue? In *International Conference on Web and Internet Economics*, pages 121–133. Springer, 2013.

[67] I. Dissanayake, J. Zhang, and B. Gu. Task division for team success in crowdsourcing contests: Resource allocation and alignment effects. *Journal of Management Information Systems*, 32(2):8–39, 2015.

[68] D. Dolev, D. G. Feitelson, J. Y. Halpern, R. Kupferman, and N. Linial. No justified complaints: On fair sharing of multiple resources. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 68–75, 2012.

[69] M. Donini, L. Oneto, S. Ben-David, J. Shawe-Taylor, and M. Pontil. Empirical risk minimization under fairness constraints. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[70] D. Dua and C. Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`.

[71] J. Duchi and H. Namkoong. Learning models with uniform performance via distributionally robust optimization. *arXiv preprint arXiv:1810.08750*, 2018.

[72] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *International Conference on Machine Learning (ICML)*, 2008.

[73] M. Dudík, J. Langford, and L. Li. Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601*, 2011.

[74] M. Dudík, D. Erhan, J. Langford, and L. Li. Doubly robust policy evaluation and optimization. *Statistical Science*, 29(4):485–511, 2014.

[75] M. Dudík, N. Haghtalab, H. Luo, R. E. Schapire, V. Syrgkanis, and J. W. Vaughan. Oracle-efficient online learning and auction design. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science*, pages 528–539. IEEE, 2017.

[76] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.

[77] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In *Innovations in Theoretical Computer Science (ITCS)*, pages 214–226. ACM, 2012.

[78] E. Eban, M. Schain, A. Mackey, A. Gordon, R. A. Saurous, and G. Elidan. Scalable learning of non-decomposable objectives. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

[79] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American economic review*, 97(1):242–259, 2007.

[80] A. Epasto, A. Muñoz Medina, S. Avery, Y. Bai, R. Busa-Fekete, C. Carey, Y. Gao, D. Guthrie, S. Ghosh, J. Ioannidis, et al. Clustering for private interest-based advertising. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2802–2810, 2021.

[81] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171:115–166, 2018.

[82] P. Eső and B. Szentes. Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, 74(3):705–731, 2007.

[83] Z. Feng, C. Podimata, and V. Syrgkanis. Learning to bid without knowing your value. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 505–522, 2018.

[84] S. Filippi, O. Cappe, A. Garivier, and C. Szepesvári. Parametric bandits: The generalized linear case. In *Conference on Neural Information Processing Systems*, volume 23, pages 586–594, 2010.

[85] S. G. Finlayson, A. Subbaswamy, K. Singh, J. Bowers, A. Kupke, J. Zittrain, I. S. Kohane, and S. Saria. The clinician and dataset shift in artificial intelligence. *The New England Journal of Medicine*, 385(3):283–286, 2021.

[86] M. Fourcade and K. Healy. Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society*, 38(8):559–572, 2013.

[87] S. A. Friedler, C. Scheidegger, S. Venkatasubramanian, S. Choudhary, E. P. Hamilton, and D. Roth. A comparative study of fairness-enhancing interventions in machine learning. In *Conference on Fairness, Accountability, and Transparency (FAccT)*, 2019.

[88] J.-J. Ganuza. Ignorance promotes competition: an auction model with endogenous private valuations. *RAND Journal of Economics*, pages 583–598, 2004.

[89] L. Georgiadis, M. J. Neely, and L. Tassiulas. *Resource Allocation and Cross-layer Control in Wireless Networks*. Now Publishers Inc, 2006.

[90] A. Gershkov. Optimal auctions and information disclosure. *Review of Economic Design*, 13(4):335–344, 2009.

[91] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica. Dominant resource fairness: Fair allocation of multiple resource types. In *NSDI*, volume 11, pages 24–24, 2011.

[92] A. Ghodsi, M. Zaharia, S. Shenker, and I. Stoica. Choosy: Max-min fair sharing for datacenter jobs with constraints. In *Proceedings of the 8th ACM European Conference on Computer Systems*, pages 365–378, 2013.

[93] T. B. Gillis. False dreams of algorithmic fairness: The case of credit pricing. *Available at SSRN 3571266*, 2020.

[94] G. Goh, A. Cotter, M. Gupta, and M. Friedlander. Satisfying real-world goals with dataset constraints. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

[95] Y. A. Gonczarowski and N. Nisan. Efficient empirical revenue maximization in single-parameter auction environments. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 856–868, 2017.

[96] Y. A. Gonczarowski and S. M. Weinberg. The sample complexity of up-to-ε multi-dimensional revenue maximization. *Journal of the ACM*, 68(3):1–28, 2021.

[97] V. Grari, B. Ruf, S. Lamprier, and M. Detyniecki. Achieving fairness with decision trees: An adversarial approach. *Data Science and Engineering*, 5(2):99–110, 2020.

[98] T. G. Group. *Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys*. The Williams Institute, 2014.

[99] S. Guha, B. Cheng, and P. Francis. Privad: Practical privacy in online advertising. In *USENIX Conference on Networked Systems Design and Implementation*, pages 169–182, 2011.

[100] C. Guo, Z. Huang, and X. Zhang. Settling the sample complexity of single-parameter revenue maximization. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 662–673, 2019.

[101] W. Guo, M. Jordan, and E. Zampetakis. Robust learning of optimal auctions. *Advances in Neural Information Processing Systems*, 34:21273–21284, 2021.

[102] W. Guo, M. Jordan, and E. Vitercik. No-regret learning in partially-informed auctions. In *International Conference on Machine Learning*, pages 8039–8055. PMLR, 2022.

[103] W. Guo, M. I. Jordan, and A. Zhou. Off-policy evaluation with policy-dependent optimization response. *Advances in Neural Information Processing Systems*, 2022.

[104] W. Guo, K. Kandasamy, J. Gonzalez, M. Jordan, and I. Stoica. Learning competitive equilibria in exchange economies with bandit feedback. In *International Conference on Artificial Intelligence and Statistics*, pages 6200–6224. PMLR, 2022.

[105] M. Gupta, A. Cotter, M. M. Fard, and S. Wang. Proxy fairness. *arXiv preprint arXiv:1806.11212*, 2018.

[106] V. Gupta, M. Huang, and P. Rusmevichientong. Debiasing in-sample policy performance for small-data, large-scale optimization. *Large-Scale Optimization (June 2, 2021)*, 2021.

[107] A. Gutman and N. Nisan. Fair allocation without trade. 2012. URL `https://www.cs.huji.ac.il/~noam/notrade.pdf`.

[108] A. Hanna, E. Denton, A. Smart, and J. Smith-Loud. Towards a critical race methodology in algorithmic fairness. *Conference on Fairness, Accountability, and Transparency (FAccT)*, 2020.

[109] M. Hardt, N. Megiddo, C. Papadimitriou, and M. Wootters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 111–122, 2016.

[110] M. Hardt, E. Price, and N. Srebro. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

[111] M. Harris, C. H. Kriebel, and A. Raviv. Asymmetric information, incentives and intrafirm resource allocation. *Management Science*, 28(6):604–620, 1982.

[112] T. B. Hashimoto, M. Srivastava, H. Namkoong, and P. Liang. Fairness without demographics in repeated loss minimization. In *International Conference on Machine Learning (ICML)*, 2018.

[113] B. Hindman, A. Konwinski, M. Zaharia, A. Ghodsi, A. D. Joseph, R. H. Katz, S. Shenker, and I. Stoica. Mesos: A platform for fine-grained resource sharing in the data center. In *NSDI*, volume 11, pages 22–22, 2011.

[114] K. Hirano and J. R. Porter. Asymptotic analysis of statistical decision rules in econometrics. In *Handbook of Econometrics*, volume 7, pages 283–354. Elsevier, 2020.

[115] B. Hooks. *Yearning: Race, gender, and cultural politics*. 1992.

[116] Z. Huang, Y. Mansour, and T. Roughgarden. Making the most of your samples. *SIAM Journal on Computing*, 47(3):651–674, 2018.

[117] M. G. Hudgens and M. E. Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

[118] H. Hussain, S. U. R. Malik, A. Hameed, S. U. Khan, G. Bickler, N. Min-Allah, M. B. Qureshi, L. Zhang, W. Yongji, N. Ghani, et al. A survey on resource allocation in high performance distributed computing systems. *Parallel Computing*, 39(11):709–736, 2013.

[119] B. Hutchinson and M. Mitchell. 50 years of test (un)fairness: Lessons for machine learning. In *Conference on Fairness, Accountability, and Transparency (FAccT)*, 2019.

[120] M. Isard, V. Prabhakaran, J. Currey, U. Wieder, K. Talwar, and A. Goldberg. Quincy: fair scheduling for distributed computing clusters. In *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*, pages 261–276, 2009.

[121] S. Ito, A. Yabe, and R. Fujimaki. Unbiased objective estimation in predictive optimization. In *International Conference on Machine Learning*, pages 2176–2185. PMLR, 2018.

[122] A. Z. Jacobs and H. Wallach. Measurement and fairness. *arXiv preprint arXiv:1912.05511*, 2019.

[123] A. Juels. Targeted advertising... and privacy too. In *Cryptographers' Track at the RSA Conference*, pages 408–424. Springer, 2001.

[124] S. M. Kakade, I. Lobel, and H. Nazerzadeh. An optimal dynamic mechanism for multi-armed bandit processes. *arXiv preprint arXiv:1001.4598*, 2010.

[125] N. Kallus and A. Zhou. Confounding-robust policy improvement. *Advances in Neural Information Processing Systems*, 31, 2018.

[126] N. Kallus and A. Zhou. Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics*, pages 1243–1251. PMLR, 2018.

[127] N. Kallus and A. Zhou. Minimax-optimal policy learning under unobserved confounding. *Management Science*, 67(5):2870–2890, 2021.

[128] N. Kallus, X. Mao, and A. Zhou. Assessing algorithmic fairness with unobserved protected class using data combination. *Conference on Fairness, Accountability, and Transparency (FAccT)*, 2020.

[129] K. Kandasamy, J. E. Gonzalez, M. I. Jordan, and I. Stoica. Mechanism design with bandit feedback. *arXiv preprint arXiv:2004.08924*, 2020.

[130] K. Kandasamy, G.-E. Sela, J. E. Gonzalez, M. I. Jordan, and I. Stoica. Online learning demands in max-min fairness. *arXiv preprint arXiv:2012.08648*, 2020.

[131] R. Kannan, G. Bayraksan, and J. R. Luedtke. Data-driven sample average approximation with covariate information. *Optimization Online. URL: http://www. optimization-online. org/DB_HTML/2020/07/7932. html*, 2020.

[132] M. Kasy and R. Abebe. Fairness, equality, and power in algorithmic decision-making. *ICML Workshop on Participatory Approaches to Machine Learning*, 2020.

[133] E. Kaufmann, N. Korda, and R. Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *International Conference on Algorithmic Learning Theory*, pages 199–213. Springer, 2012.

[134] E. H. Kennedy. Semiparametric theory and empirical processes in causal inference. In *Statistical Causal Inferences and their Applications in Public Health Research*, pages 141–167. Springer, 2016.

[135] T. Kitagawa and A. Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

[136] M. C. Knaus, M. Lechner, and A. Strittmatter. Heterogeneous employment effects of job search programmes: A machine learning approach. *Journal of Human Resources*, pages 0718–9615R1, 2020.

[137] N. Kolodny. Why equality of treatment and opportunity might matter. *Philosophical Studies*, 176:3357–3366, 2019.

[138] I. Krumpal. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality and Quantity*, 47:2025–2047, 2011.

[139] A. Kube, S. Das, and P. J. Fowler. Allocating interventions based on predicted outcomes: A case study on homelessness services. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 622–629, 2019.

[140] S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.

[141] M. J. Kusner, J. R. Loftus, C. Russell, and R. Silva. Counterfactual fairness. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[142] P. Lahoti, A. Beutel, J. Chen, K. Lee, F. Prost, N. Thain, X. Wang, and E. H. Chi. Fairness without demographics through adversarially reweighted learning. *arXiv preprint arXiv:2006.13114*, 2020.

[143] K. Lai, L. Rasmusson, E. Adar, L. Zhang, and B. A. Huberman. Tycoon: An implementation of a distributed, market-based resource allocation system. *Multiagent and Grid Systems*, 1(3):169–182, 2005.

[144] H. Lam and H. Qian. Bounding optimality gap in stochastic optimization via bagging: Statistical efficiency and stability. *arXiv preprint arXiv:1810.02905*, 2018.

[145] A. Lamy, Z. Zhong, A. K. Menon, and N. Verma. Noise-tolerant fair classification. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[146] C. Lazar and S. Vijaykumar. A resolution in algorithmic fairness: Calibrated scores for fair classifications. *arXiv preprint arXiv:2002.07676*, 2020.

[147] T. N. Le, X. Sun, M. Chowdhury, and Z. Liu. Allox: compute allocation in hybrid clusters. In *Proceedings of the Fifteenth European Conference on Computer Systems*, pages 1–16, 2020.

[148] L. Leqi and E. H. Kennedy. Median optimal treatment regimes. *arXiv preprint arXiv:2103.01802*, 2021.

[149] D. Li and I. Jewitt. Cheap talk advertising in auctions: horizontally vs vertically differentiated products. 2017.

[150] H. Li and X. Shi. Discriminatory information disclosure. *American Economic Review*, 107(11):3363–85, 2017.

[151] H. Li, G. Zhao, R. Johari, and G. Y. Weintraub. Interference, bias, and variance in two-sided marketplace experimentation: Guidance for platforms. *arXiv preprint arXiv:2104.12222*, 2021.

[152] J. Li and J. Xue. Egalitarian division under leontief preferences. *Economic Theory*, 54 (3):597–622, 2013.

[153] J. Li, S. Huang, and A. M.-C. So. A first-order algorithmic framework for Wasserstein distributionally robust logistic regression. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[154] L. Li, Y. Lu, and D. Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pages 2071–2080. PMLR, 2017.

[155] L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt. Delayed impact of fair machine learning. In *International Conference on Machine Learning (ICML)*, 2018.

[156] R. Lopez, C. Li, X. Yan, J. Xiong, M. Jordan, Y. Qi, and L. Song. Cost-effective incentive allocation via structured counterfactual inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 4997–5004, 2020.

[157] L. Lovász and S. Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 57–68. IEEE, 2006.

[158] H. Ma, F. Fang, and D. C. Parkes. Spatio-temporal pricing for ridesharing platforms. *Operations Research*, 2021.

[159] C. F. Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246, 2004.

[160] A. Mas-Colell, M. D. Whinston, J. R. Green, et al. *Microeconomic Theory*, volume 1. Oxford university press New York, 1995.

[161] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, pages 1269–1283, 1990.

[162] J. Mejia and C. Parker. When transparency fails: Bias and financial incentives in ridesharing platforms. *Management Science*, 67(1):166–184, 2021.

[163] P. Milgrom. Simplified mechanisms with an application to sponsored-search auctions. *Games and Economic Behavior*, 70(1):62–70, 2010.

[164] P. Milgrom and P. R. Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, 2004.

[165] U. Misra, R. Liaw, L. Dunlap, R. Bhardwaj, K. Kandasamy, J. E. Gonzalez, I. Stoica, and A. Tumanov. Rubberband: cloud-based hyperparameter tuning. In *Proceedings of the Sixteenth European Conference on Computer Systems*, pages 327–342, 2021.

[166] J. Morgenstern and T. Roughgarden. The pseudo-dimension of near-optimal auctions. *arXiv preprint arXiv:1506.03684*, 2015.

[167] J. Morgenstern and T. Roughgarden. Learning simple auctions. In *Conference on Learning Theory*, pages 1298–1318. PMLR, 2016.

[168] H. Mozannar, M. I. Ohannessian, and N. Srebro. Fair learning with private demographic data. *arXiv preprint arXiv:2002.11651*, 2020.

[169] E. Munro, S. Wager, and K. Xu. Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*, 2021.

[170] M. Mutnỳ and A. Krause. Efficient high dimensional bayesian optimization with additivity and quadrature fourier features. *Conference on Neural Information Processing Systems 31*, pages 9005–9016, 2019.

[171] R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1): 58–73, 1981.

[172] H. Namkoong and J. Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

[173] H. Narasimhan, A. Cotter, and M. Gupta. Optimizing generalized rate metrics with three players. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[174] H. Narasimhan, A. Cotter, and M. R. Gupta. On making stochastic classifiers deterministic. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[175] H. Narasimhan, A. Cotter, Y. Zhou, S. Wang, and W. Guo. Approximate heavily-constrained learning with lagrange multiplier models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[176] A. Narayanan. Translation tutorial: 21 fairness definitions and their politics. In *Conference on Fairness, Accountability, and Transparency (FAccT)*, 2018.

[177] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245, 1994.

[178] D. C. Parkes, A. D. Procaccia, and N. Shah. Beyond dominant resource fairness: Extensions, limitations, and indivisibilities. *ACM Transactions on Economics and Computation*, 3(1):1–22, 2015.

[179] C. Paxton, A. Niculescu-Mizil, and S. Saria. Developing predictive models using electronic medical records: challenges and pitfalls. In *AMIA Annual Symposium Proceedings*, volume 2013, page 1109. American Medical Informatics Association, 2013.

[180] J. Perdomo, T. Zrnic, C. Mendler-Dünner, and M. Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.

[181] D. Post, S. Coppinger, and G. Sheble. Application of auctions as a pricing mechanism for the interchange of electric power. *IEEE Transactions on Power Systems*, 10(3): 1580–1584, 1995.

[182] Z. Qi, J.-S. Pang, and Y. Liu. Estimating individualized decision rules with tail controls. *arXiv preprint arXiv:1903.04367*, 2019.

[183] O. Rafieian and H. Yoganarasimhan. Targeting and privacy in mobile advertising. *Marketing Science*, 40(2):193–218, 2021.

[184] A. Rahmattalabi, P. Vayanos, K. Dullerud, and E. Rice. Learning resource allocation policies from observational data with an application to homeless services delivery. *arXiv preprint arXiv:2201.10053*, 2022.

[185] A.-K. Roesler and B. Szentes. Buyer-optimal learning and monopoly pricing. *American Economic Review*, 107(7):2072–80, 2017.

[186] T. Roughgarden and O. Schrijvers. Ironing in the dark. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 1–18, 2016.

[187] T. Roughgarden and J. R. Wang. Minimizing regret with multiple reserves. *ACM Transactions on Economics and Computation*, 7(3):1–18, 2019.

[188] P. Rusmevichientong and J. N. Tsitsiklis. Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411, 2010.

[189] C. Russell, M. J. Kusner, J. Loftus, and R. Silva. When worlds collide: Integrating different counterfactual assumptions in fairness. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[190] K. Rzadca, P. Findeisen, J. Swiderski, P. Zych, P. Broniek, J. Kusmierek, P. Nowak, B. Strack, P. Witusowski, S. Hand, et al. Autopilot: workload autoscaling at google. In *Proceedings of the Fifteenth European Conference on Computer Systems*, pages 1–16, 2020.

[191] A. Saumard and J. A. Wellner. Log-concavity and strong log-concavity: a review. *Statistics surveys*, 8:45, 2014.

[192] D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models (rejoinder). *Journal of the American Statistical Association*, 94(448):1135–1146, 1999.

[193] U. Shalit, F. D. Johansson, and D. Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*, pages 3076–3085. PMLR, 2017.

[194] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2021.

[195] C. Shi, D. M. Blei, and V. Veitch. Adapting neural networks for the estimation of treatment effects. *arXiv preprint arXiv:1906.02120*, 2019.

[196] H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000.

[197] W. Simonsen. *Citizen participation in resource allocation*. Routledge, 2018.

[198] A. Slivkins et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

[199] L. Sun. Empirical welfare maximization with constraints. *arXiv preprint arXiv:2103.15298*, 2021.

[200] A. Swaminathan and T. Joachims. Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research*, 16(1): 1731–1755, 2015.

[201] V. Syrgkanis. A sample complexity measure with applications to learning optimal auctions. *arXiv preprint arXiv:1704.02598*, 2017.

[202] P. Thomas, G. Theocharous, and M. Ghavamzadeh. High-confidence off-policy evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

[203] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[204] M. Tiwari, T. Groves, and P. C. Cosman. Competitive equilibrium bitrate allocation for multiple video streams. *IEEE Transactions on Image Processing*, 19(4):1009–1021, 2009.

[205] L. Tran, C. Yiannoutsos, K. Wools-Kaloustian, A. Siika, M. Van Der Laan, and M. Petersen. Double robust efficient estimators of longitudinal treatment effects: Comparative performance in simulations and a case study. *The international Journal of Biostatistics*, 15(2), 2019.

[206] S. van de Geer and L. Stougie. On rates of convergence and asymptotic normality in the multiknapsack problem. *Mathematical Programming*, 51(1):349–358, 1991.

[207] H. R. Varian. Equity, envy, and efficiency. 1973.

[208] H. R. Varian and H. R. Varian. *Microeconomic Analysis*, volume 3. Norton New York, 1992.

[209] V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth, et al. Apache hadoop yarn: Yet another resource negotiator. In *Proceedings of the 4th Annual Symposium on Cloud Computing*, page 5. ACM, 2013.

[210] S. Venkataraman, Z. Yang, M. Franklin, B. Recht, and I. Stoica. Ernest: Efficient performance prediction for large-scale advanced analytics. In *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*, pages 363–378, 2016.

[211] C. Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 2009.

[212] S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

[213] S. Wager and K. Xu. Experimenting in equilibrium. *Management Science*, 2021.

[214] M. J. Wainwright. *High-dimensional Statistics: a Non-asymptotic Viewpoint*, volume 48. Cambridge University Press, 2019.

[215] L. Wang, Y. Bai, A. Bhalla, and T. Joachims. Batch learning from bandit feedback through bias corrected reward imputation. In *ICML Workshop on Real-World Sequential Decision Making*, 2019.

[216] S. Wang and M. R. Gupta. Deontological ethics by monotonicity shape constraints. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

[217] S. Wang, W. Guo, H. Narasimhan, A. Cotter, M. Gupta, and M. Jordan. Robust optimization for fairness with noisy protected groups. *Advances in Neural Information Processing Systems*, 33:5190–5203, 2020.

[218] J. Weed, V. Perchet, and P. Rigollet. Online learning in repeated auctions. In *Conference on Learning Theory*, pages 1562–1583. PMLR, 2016.

[219] R. Wolski, J. S. Plank, J. Brevik, and T. Bryan. Analyzing market-based resource allocation strategies for the computational grid. *The International Journal of High Performance Computing Applications*, 15(3):258–281, 2001.

[220] B. Woodworth, S. Gunasekar, M. I. Ohannessian, and N. Srebro. Learning non-discriminatory predictors. In *Conference on Learning Theory (COLT)*, pages 1920–1953, 2017.

[221] I.-C. Yeh and C. hui Lien. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36:2473–2480, 2009.

[222] M. B. Zafar, I. Valera, M. G. Rodriguez, and K. P. Gummadi. Fairness constraints: Mechanisms for fair classification. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

[223] M. B. Zafar, I. Valera, M. Gomez-Rodriguez, and K. P. Gummadi. Fairness constraints: A flexible approach for fair classification. *Journal of Machine Learning Research (JMLR)*, 20:1–42, 2019.

[224] S. M. Zahedi, Q. Llull, and B. C. Lee. Amdahl's law in the datacenter era: a market for fair processor allocation. In *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pages 1–14. IEEE, 2018.

[225] L. Zhang. Proportional response dynamics in the fisher market. *Theoretical Computer Science*, 412(24):2691–2698, 2011.

[226] Y. Zhao, D. Zeng, A. J. Rush, and M. R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.