Fair Online Advertising*

L. Elisa Celis[1], Anay Mehrotra[2], and Nisheeth K. Vishnoi[3]

[1,3]Yale University
[2]Indian Institute of Technology, Kanpur

January 30, 2019

Abstract

Online advertising platforms are thriving due to the customizable audiences they offer advertisers. However, recent studies show that the audience an ad gets shown to can be discriminatory with respect to sensitive attributes such as gender or ethnicity, inadvertently crossing ethical and/or legal boundaries. To prevent this, we propose a constrained optimization framework that allows the platform to control of the fraction of each sensitive type an advertiser's ad gets shown to while maximizing its ad revenues. Building upon Myerson's classic work, we first present an optimal auction mechanism for a large class of fairness constraints. Finding the parameters of this optimal auction, however, turns out to be a non-convex problem. We show how this non-convex problem can be reformulated as a more structured non-convex problem with no saddle points or local-maxima; allowing us to develop a gradient-descent-based algorithm to solve it. Our empirical results on the A1 Yahoo! dataset demonstrate that our algorithm can obtain uniform coverage across different user attributes for each advertiser at a minor loss to the revenue of the platform, and a small change in the total number of advertisements each advertiser shows on the platform.

*The code for the simulations is available at `https://github.com/AnayMehrotra/Fair-Online-Advertising`

Contents

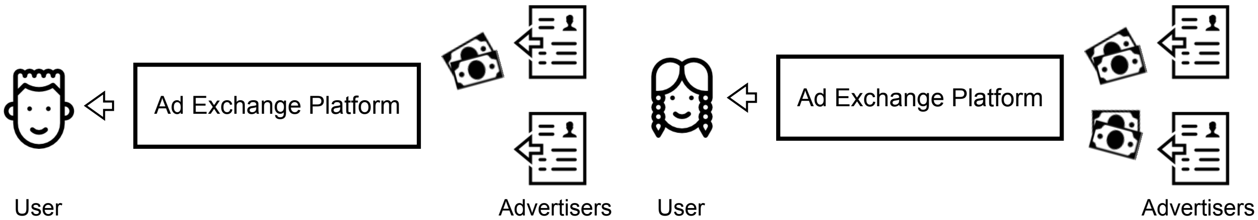
1 Introduction

Online advertisements are the main source of revenue for social-networking sites and search engines such as Google [1]. Ad exchange platforms allow advertisers to select the target audience for their ad by specifying desired user demographics, interests and browsing histories [5]. Every time a user loads a webpage or enters a search term, bids are collected from relevant advertisers [11], and an auction is conducted to determine which ad is shown, and how much the advertiser is charged [18, 31]. As it is not practical for advertisers to place individual bids for every user, the advertiser instead gives some high-level preferences about their budget and target audience, and the platform places bids on their behalf [12].

More formally, let there be n advertisers, and m types of users. Each advertiser i specifies their target demographic, average bid, and budget to the platform, which then decides a distribution, $\mathcal{P}_{ij}$, of bids of advertiser $i \in [n]$ for user type $j \in [m]$. These distributions represent the value of the user to the advertiser, and ensure that the advertiser only bids for users in their target demographic, with the expected bid not exceeding the amount specified by the advertiser [7]. At each time step, a user visits a web page (e.g., Facebook or Twitter), the user's type $j \in [m]$ is observed, and a bid v_i is drawn from $\mathcal{P}_{ij}$, for each advertiser $i \in [n]$. Receiving these bids as input, the mechanism $\mathcal{M}$ decides an allocation $x(v)$ and price $p(v)$ for the advertisement slot. Overall, such targeted advertising leads to higher utilities for the advertisers who show content to relevant audiences, for the users who view related advertisements, and for the platform which can benefit from selling targeted advertisements [8–10, 30].

However, targeted advertising can also lead to discriminatory practices. For instance, searches with "black-sounding" names were much more likely to be shown ads suggestive of an arrest record [24]. Another study found that women were shown fewer advertisements for high paying jobs than men with similar profiles [4]. In fact, a recent experiments demonstrated that ads can be *inadvertently* discriminatory [17]; they found that STEM job ads, specifically designed to be unbiased by the advertisers, were shown to more men than women across all major platforms (Facebook Ads, Google Ads, Instagram and Twitter). On Facebook, a platform with 52% women [27] the advertisement was shown to 20% more men than women. They suggest that this is a result of *competitive spillovers* among advertisers, and is neither a pure reflection of pre-existing cultural bias, nor a result of human input to the algorithm. Such (inadvertent) discrimination has led to two recent cases filed against Facebook, which will potentially lead to civil lawsuits alleging employment and housing discrimination [14, 21, 25]

To gain intuition, consider the setting in which there are two advertisers with similar bids/budgets, but one advertiser specifically targets women (which is allowed for certain types of ads, e.g., related to clothing), while the second advertiser is does not target based on gender (e.g., because they are advertising a job). The first advertiser creates an imbalance on the platform by taking up ad slots for women and, as a consequence, the second advertiser ends up advertising to disproportionately fewer women and is inadvertently discriminatory. Currently, online advertising platforms have no mechanism to check this type of discrimination. In fact, the only way around this would be for the advertiser to set up separate campaigns for different user types and ensure that each one reached similar number of the sub-target audience, however doing so would violate discrimination rules as, in itself, each sub-advertisement would be discriminatorily selecting for a specific demographic [6].



(a) Only the first advertiser bids for this user. (b) Both advertisers place bids for this user.

Figure 1: The platform accepts bids from *relevant* advertisers and by conducting an auction decides the ad to be shown to the user. Different advertisers have different target audiences and only bid for users in their target audience.

Our Contributions

Our main contribution is an optimization-based framework which maximizes the revenue of the platform subject to satisfying constraints that prevent the emergence of inadvertent discrimination as described above. The constraints can be formulated as any one of a wide class of "fairness" constraints as presented in [2]. The framework allows for intersectionality, allowing constraints across multiple sensitive attributes (e.g., gender, race, geography and economic class) and allows for restricting different advertisers to different constraints.

Formally, building on Myerson's seminal work [20], we characterize the truthful revenue-optimal mechanism which satisfies the given constraints (Theorem 3.1). The user types, as defined by their sensitive attributes, are taken as input along with the type-specific bid distributions for each advertiser, and we assume that bids are drawn from these distributions independently. Our mechanism is parameterized by constant "shifts" which it applies to bids for each advertiser-type pair. Finding the parameters of this optimal mechanism, however, is a non-convex optimization problem, both the objective and the constraints. Towards solving this, we first propose a novel reformulation of the objective as a composition of a convex function constrained on a polytope, and an unconstrained non-convex function (Theorem 3.2). Interestingly, the non-convex function is reasonably well behaved, with no saddle-points or local-maxima. This allows us to develop a gradient descent based scheme (Algorithm 1) to solve the reformulated program, which under mild assumptions has a fast convergence rate of $\widetilde{O}(1/\varepsilon^2)$ (Theorem 3.3).

We evaluate our approach empirically by studying the effect of the constraints on the revenue of the platform and the advertisers using the *Yahoo! Search Marketing Advertising Bidding Data* [29]. We find that our mechanism can obtain uniform coverage across different user types for each advertiser while loosing less than 5% of the revenue (Figure 3(b)). Further, we observe that the total-variation distance between the fair and unconstrained distributions of total advertisements an advertiser shows on the platform is less than 0.05 (Figure 4).

To the best of our knowledge, we are the first to give a framework to prevent inadvertent discrimination in online auctions.

2 Our model

We refer the reader to the excellent treatise [15] on Mechanism design, for a detailed discussion of the preliminaries.
A mechanism $\mathcal{M}$ is defined by its allocation rule $x\colon \mathbb{R}^n \to [0,1]^n$, and its payment rule $p\colon \mathbb{R}^n \to \mathbb{R}^n_{\geq 0}$. Among these truthful mechanisms are those in which revealing the true valuation is optimal for

all bidders. Any allocation rule $x(v_1, v_2, \ldots, v_n)$, of a truthful mechanism is monotone in v_i for all $i \in [n]$. Further, it can be shown that for any mechanism there is a truthful mechanism which offers the same revenue to the seller, and the same utility to each bidder [19]. As such, we restrict ourselves to truthful mechanisms. It is a well known fact [22] that for any truthful mechanism the payment rule p, is uniquely defined by its allocation rule x. Hence, for any truthful mechanism our only concern is the allocation rule, x.

Let $\mathcal{P}$ be the distribution of valuation of a bidder, $\mathrm{pdf}\colon \mathbb{R} \to \mathbb{R}_{>0}$ be its probability density function, and $\mathrm{cdf}\colon \mathbb{R} \to [0,1]$ be its cumulative density function, then we define the virtual valuation $\phi\colon \mathrm{supp}(\mathcal{P}) \to \mathbb{R}$, as $\phi(v) \coloneqq v - (1-\mathrm{cdf}(v))(\mathrm{pdf}(v))^{-1}$. We say $\mathcal{P}$ is regular iff $\phi(v)$ is non-decreasing in v. Likewise, we say $\mathcal{P}$ is strictly regular iff $\phi(v)$ is strictly increasing in v.

Myerson's Optimal Mechanism. Myerson's mechanism is defined as the VCG mechanism [3, 13, 28] where the virtual valuation ϕ_i, is submitted as the bid v_i, for each bidder i. If the valuations v_i, and therefore the virtual valuations ϕ_i are *independent*, then for any truthful mechanism the virtual surplus, $\sum_{i\in[n]} \phi_i x_i(\phi_i)$, is equal to the revenue, $\sum_{i\in[n]} p_i$, in expectation. Since VCG is surplus maximizing, if Myerson's mechanism is truthful then it maximizes the revenue.

Notation. We represent the distribution of valuations of advertiser $i \in [n]$ for user type $j \in [m]$ by $\mathcal{P}_{ij}$. Let $\phi_{ij} \in \mathbb{R}$ be the virtual valuation of advertiser $i \in [n]$ for user type $j \in [m]$, $f_{ij}\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ be its probability density function, and $F_{ij}\colon \mathbb{R} \to [0,1]$ be its cummulative density function. We denote the joint virtual valuation of all advertisers for user type j, by $\phi_j \in \mathbb{R}^n$, and its joint probability density function by $f_j\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$. The user types $j \in [m]$, are distributed according to the known distribution $\mathcal{U}$, and the mechanism's allocation rule for user type j, is denoted by $x_j\colon \mathbb{R}^n \to [0,1]^n$.

2.1 Fairness Constraints

We would like to guarantee that advertisers have a fair coverage across user types. We do so by placing constraints on the coverage of an advertiser. Formally, we define an advertiser i's coverage of user type j as the probability that advertiser i wins the auction conditioned on the user being of type j.

$$q_{ij}(x_j) \coloneqq \Pr_{\mathcal{U}}[j] \int\limits_{\mathrm{supp}(\phi_j)} x_{ij}(\phi_j) df_j(\phi_j), \qquad \text{(Coverage, 1)}$$

where $x_{ij}(\phi_j)$ is the i-th component of $x_j(\phi_j)$.

Towards ensuring that an advertiser has a fair coverage of different user types, we consider the proportional coverage of the advertiser on each user type. Given vectors ℓ_j, $u_j \in [0,1]^n \ \forall\ j \in [m]$, we define (ℓ, u)-fairness constraints for each advertiser i and user type j, as a lower bound ℓ_{ij}, and an upper bound u_{ij}, on the proportion of users of type j the advertiser shows ads to, i.e., we impose the following constraints,

$$\ell_{ij} \leq \frac{q_{ij}}{\sum_{t\in[m]} q_{it}} \leq u_{ij} \quad \forall\ i \in [n] \text{ and } j \in [m]. \qquad \text{((}\ell, u\text{)-fairness constraints, 2)}$$

2.2 Discussion of Fairness Constraints

Returning to the example presented in the introduction, we can ensure that the advertiser shows $x\%$ of total ads to women, by choosing a lower bound of x for this advertiser on women.

More generally, for m user types, moderately placed lower bounds and upper bounds ($\ell_{ij} \sim 1/m$ and $u_{ij} \sim 1/m$), for some subset of advertisers, ensure this subset has a uniform coverage across all user types, while allowing other advertisers to target a specific user types.

Importantly, while ensuring fairness across multiple user types our constraints allow for targeting within any single user type. This is vital as the advertiser may not derive the same utility from each user, and could be willing to pay a higher amount for more relevant users in the same user type. For example, if the advertiser is displaying job ads, then a user already looking for job opportunities may be of a higher value to the advertiser than one who is not.

For a detailed discussion on how such constraints can encapsulate other popular metrics, such as risk-difference, we refer the reader to [2].

2.3 Optimization Problem

We would like to develop a mechanism which maximizes the revenue while satisfying the upper and lower bound constraints in Equation (2). Towards formally stating our problem, we define the revenue of mechanism $\mathcal{M}$, with an allocation rule $x_j \colon \mathbb{R}^n \to [0,1]^n$, for user type j as

$$\mathrm{rev}_{\mathcal{M}} \coloneqq \sum_{i\in[n],j\in[m]} \Pr_{\mathcal{U}}[j] \int_{\mathrm{supp}(\phi_j)} \phi_{ij} x_{ij}(\phi_j) d\phi_j, \tag{Revenue, 3}$$

where $x_{ij}(\phi_j)$, and ϕ_{ij} are the i-th component of $x_j(\phi_j)$, and ϕ_j respectively. Thus, we can express our optimization problem with respect to *functions* $x(\cdot)$, or as an infinite dimensional optimization problem as follows.

(**Infinite-dimensional fair advertising problem).** For all user types $j \in [m]$, find the optimal allocation rule $x_j(\cdot) \colon \mathbb{R}^n \to [0,1]^n$ for,

$$\max_{x_{ij}(\cdot)\geq 0} \mathrm{rev}_{\mathcal{M}}(x_1, x_2, \ldots, x_m) \tag{4}$$

$$\text{s.t.} \quad q_{ij}(x_j) \geq \ell_{ij} \sum_{t\in[m]} q_{it}(x_t) \ \forall \ j \in [m], i \in [n] \tag{5}$$

$$q_{ij}(x_j) \leq u_{ij} \sum_{t\in[m]} q_{it}(x_t) \ \forall \ j \in [m], i \in [n] \tag{6}$$

$$\sum_{i\in[n]} x_{ij}(\phi_j) \leq \ 1 \ \forall \ j \in [m], \phi_j, \tag{7}$$

where (5) encodes the lower bound constraints, (6) encodes the upper bound constraints, and (7) ensures that only one ad is allocated.

In the above problem, we are looking for a collection of optimal continuous function $x^\star$. To be able to solve this problem, we need – in the least – a finite dimensional formulation of the fair online advertisement problem.

3 Theoretical Results

Our first result is structural, and gives a characterization of the optimal solution $x^\star$, to the infinite-dimensional fair advertising problem, in terms of a matrix $\alpha \in \mathbb{R}^{n\times m}$, making it a finite-dimensional optimization problem with respect to α.

Theorem 3.1. (Characterization of an optimal allocation rule). There exists an $\alpha = \{\alpha_j\}_{j\in[m]} \in \mathbb{R}^{n\times m}$, such that, if for all $j \in [m]$ $\mathcal{P}_j$ is strictly regular and independent, then , the set of allocation rules $x_j(\cdot, \alpha_j)\colon \mathbb{R}^n \to [0,1]^n$ $\forall$ $j \in [m]$, defined below, is optimal for the infinite-dimensional fair advertising problem.

$$x_j(v_j, \alpha_j) \coloneqq \operatorname*{argmax}_{i\in[n]}(\phi_{ij} + \alpha_{ij}) \qquad (\alpha\text{-shifted mechanism, 8})$$

where we randomly breaks if any (this is equivalent to the allocation rule of VCG mechanism.)

We present the proof of Theorem 3.1 in Section 6.1. In the proof, we analyze the dual problem of the infinite-dimensional fair advertising problem. We reduce the dual problem to one lagrangian variable, by fixing the lagrangian variables corresponding lower bound (5) and upper bound (6) constraints to their optimal values.

The resulting problem turns out to be the dual of the unconstrained revenue maximizing problem, for which Myerson's mechanism is the optimal solution. We interpret the fixed lagrangian variables as shifting the original virtual valuations, ϕ_{ij}. It then follows that for some $\alpha \in \mathbb{R}^{n\times m}$, the α-shifted mechanism (8) is the optimal solution to the infinite-dimensional fair advertising problem.

Now, our task is reduced from finding an optimal allocation rule, to finding an α characterizing the optimal allocation rule. Towards this, let us define the revenue, $\mathrm{rev_{shift}}\colon \mathbb{R}^{n\times m} \to \mathbb{R}$, and coverage $q_{ij}\colon \mathbb{R}^{n\times m} \to [0,1]$, as functions of α.

$$\mathrm{rev_{shift}}(\alpha) \coloneqq \sum_{i\in[n]j\in[m]} \Pr_{\mathcal{U}}[j] \int_{\mathrm{supp}(f_{ij})} y f_{ij}(y) \prod_{k\in[n]\backslash\{i\}} F_{kj}(y + \alpha_{ij} - \alpha_{kj})dy \qquad (\text{Revenue } \alpha\text{-shifted mechanism, 9})$$

$$q_{ij}(\alpha) \coloneqq \Pr_{\mathcal{U}}[j] \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) \prod_{k\in[n]\backslash\{i\}} F_{kj}(y + \alpha_{ij} - \alpha_{kj})dy. \qquad (\text{Coverage } \alpha\text{-shifted mechanism, 10})$$

Depending on the nature of the distribution, the gradients ${\partial \mathrm{rev_{shift}}(\alpha)}/{\partial \alpha_i}$ and ${\partial q_{ij}(\alpha)}/{\partial \alpha_i}$ may not be a monotone function of the α (e.g., consider the exponential distribution). Therefore, in general neither is $\mathrm{rev_{shift}}(\cdot)$ a concave, nor is $q_{ij}(\cdot)$ a convex function of α (see Section B for a concrete example). Hence, this optimization problem is non-convex both in its objective and in its constraints. We require further insights to solve the problem efficiently.

Towards this we observe that revenue is a concave function of q. Consider two allocation rules obtaining coverages $q_1,\ q_2 \in [0,1]^{n\times m}$, and revenues $R_1,\ R_2 \in \mathbb{R}$ respectively. If we use the first with a probability $\gamma \in [0,1]$, then we achieve a coverage $\gamma q_1 + (1-\gamma)q_2$, and a revenue $\gamma R_1 + (1-\gamma)R_2$. Therefore, the optimal allocation rule achieving $\gamma q_1 + (1-\gamma)q_2$, has at least $\gamma R_1 + (1-\gamma)R_2$ revenue. Choosing the allocation rules which maximize the revenue for q_1 and q_2 respectively, this argument shows that revenue is a concave function of the coverage q.

Let $\mathrm{rev}\colon [0,1]^{(n-1)\times m} \to \mathbb{R}$, be the maximum revenue of the platform as a function of coverage q.[1] Consider the following two optimization problems.

[1] We drop q_{ij} for some $i \in [n]$ and each $j \in [m]$. This is crucial to calculate $\nabla \mathrm{rev}(\cdot)$, see Remark 4.1.

(Optimal coverage problem). Find the optimal $q \in [0,1]^{n\times m}$ for,

$$\max_{q\in[0,1]^n} \ \mathrm{rev}(q) \tag{11}$$

$$\text{s.t. } q_{ij} \geq \ell_{ij}\sum_{t\in[m]} q_{it} \qquad \forall\, j \in [m], i \in [n] \tag{12}$$

$$q_{ij} \leq u_{ij}\sum_{t\in[m]} q_{it} \qquad \forall\, j \in [m], i \in [n] \tag{13}$$

$$\sum_{i\in[n]} q_{ij} \leq \ \mathrm{Pr}_{\mathcal{U}}[j] \qquad \forall\, j \in [m] \tag{14}$$

(Optimal shift problem). Given the target coverage $\delta \in [0,1]^{n\times m}$, find the optimal $\alpha \in \mathbb{R}^{n\times m}$ for,

$$\min_{\alpha\in\mathbb{R}^{n\times m}} \mathcal{L}(\alpha) := \|\delta - q(\alpha)\|_F^2 \tag{15}$$

Our next result relates the solution of the above two problems with the infinite-dimensional fair advertising problem.

Theorem 3.2. Given a solution $q^\star \in [0,1]^{n\times m}$ to the optimal coverage problem, the solution $\alpha^\star$, to the optimal shift problem with $\delta = q^\star$, defines an optimal α-shifted mechanism (8) for the infinite-dimensional fair advertising problem.

Proof. Adding the all 1 vector, 1_n, to α_j for any $j \in [m]$, does not change the allocation rule of the α-shifted mechanism. Thus, it suffices to show that for all $\delta \in [0,1]^{n\times m}$, there is a unique α, such that, $q(\alpha) = \delta$ and $\alpha_{1j} = 0$ for all $j \in [m]$.

We change show that for all $\delta \in [0,1]^{n\times m}$, there is at-least one $\alpha \in \mathbb{R}^{n\times m}$. In fact, the greedy algorithm which increases all α_{ij}, such that, $q_{ij}(\alpha) < \delta_{ij}$ and $i \neq 1$ finds the required α.

Consider distinct $\alpha, \beta \in \mathbb{R}^{n\times m}$, such that, $\alpha_{1j} = 0$ and $\beta_{ij} = 0$. We can show that $q(\alpha) \neq q(\beta)$, by Consider the advertiser and user type pair (i', j') whose shift changes by the largest magnitude. We can show that $q_{i'j'}(\alpha) \neq q_{i'j'}(\beta)$, thereby proving that $q(\alpha) \neq q(\beta)$. □

The above theorem allows us to find the optimal α, by solving the *optimal coverage problem* and *optimal shift problem*. Towards this, let us consider the *optimal coverage problem*. We already know that its objective is a concave function. We can further observe that its constraints are linear in q. In particular, the constraints define a polytope, $\mathcal{Q} \subseteq [0,1]^{n\times m}$, which we refer to as the constraint-polytope. Therefore, it is a convex program, and a possible direction to solve this program is to use a gradient based methods.

The trouble is that we do not have direct access to $\nabla\mathrm{rev}$. A key idea is that, if we let $\alpha = q^{-1}(\delta)$, then we can calculate $\nabla\mathrm{rev}(\delta)$ by solving the following linear-system,

$$(J_q(\alpha))^\top \nabla\mathrm{rev}(\delta) = \nabla\mathrm{rev}_{\mathrm{shift}}(\alpha).$$

Where, $J_q(\alpha)$ is the Jacobian of $\mathrm{vec}(q(\alpha)) \in \mathbb{R}^{(n-1)m}$, with respect $\mathrm{vec}(\alpha) \in \mathbb{R}^{(n-1)m}$.[2] It turns out that this Jacobian $J_q(\alpha)$ is invertible for all $\alpha \in \mathbb{R}^{n\times m}$, and therefore the above linear system has an exact solution (see Section 4.1 for the details).

Let us consider the *optimal shift problem*. The objective of the problem is non-convex (see Figure 8(b) and Section B for an example.) Interestingly, $\nabla\mathcal{L}(\alpha)$ is a linear combination of $\nabla q_{ij}(\alpha)$

[2] $\mathrm{vec}(\cdot)$ represents the vectorization operator.

for all $i \in [n]$ and $j \in [m]$. Since the rows of the Jacobian, $\{\nabla q_j(\alpha)\}_{j\in[m]}$ are linearly independent, the gradient is never zero unless we are at the global minimum where $\alpha = q^{-1}(\delta)$. This guarantees that the objective function does not have any saddle-points or local-maximum, and that any local-minimum is a global minimum. Using this we can develop efficient algorithms to solve the *optimal coverage problem* (Lemma 6.2).

This brings us to our main algorithmic result, which is an algorithm to find the optimal allocation rule for the *infinite-dimensional fair advertising problem.*

Theorem 3.3. (An algorithm to solve the infinite-dimensional fair advertising problem). There is an algorithm (Algorithm 1) which outputs $\alpha \in \mathbb{R}^{n\times m}$, such that, if assumptions (16), (17), (18), and (19) are satisfied, then the α-shifted mechanism (8) achieves a revenue ε-close to the optimal for the infinite-dimensional fair advertising problem in

$$\widetilde{O}\left(\frac{n^7 \log m}{\varepsilon^2} \frac{(\mu_{\max}\rho)^2}{(\mu_{\min}\eta)^4}(L + n^2\mu_{\max}^2)\right) \text{ steps.}$$

Where the arithmetic calculations in each step are bounded by calculating $\nabla \mathrm{rev}(\cdot)$ once, and $\widetilde{O}$ hides log factors in $n,\ \rho,\ \eta,\ \mu_{\max},\ 1/\varepsilon$ and $1/\mu_{\min}$.

Roughly, the above algorithm has a convergence rate of $\widetilde{O}(1/\varepsilon^2)$, under the assumptions which we list below.

Assumptions

For all $i \in [n],\ j \in [m],$ and $y_1, y_2 \in \mathrm{supp}(f_{ij})$

1. $q_{ij} > \eta$ (η-coverage, 16)
2. $\mu_{\min} \le f_{ij}(y_1) \le \mu_{\max}$ (Distributed distribution, 17)
3. $|f_{ij}(y_1) - f_{ij}(y_2)| < L|y_1 - y_2|$ (Lipschitz distribution, 18)
4. $\big|\mathbb{E}[\phi_{ij}]\big| = \Big|\int_{\mathrm{supp}(f_{ij})} z f_{ij}(z)dz\Big| < \rho.$ (Bounded bid, 19)

Assumption (16) guarantees that all advertisers have at least an η probability of winning on every user type, assumption (17) places lower and upper bounded on the probability density functions of the ϕ_{ij}, assumption (18) guarantees that the probability density functions of the ϕ_{ij} is L-Lipschitz continuous, and assumption (19) assumes that the expected ϕ_{ij} is bounded.

We expect Assumptions (16) and (19) to hold in any real-world setting. We can drop the lower bound in Assumption (17) by introducing "jumps" in α, to avoid ranges where the measure of bids is small. Removing assumption (18) would be an interesting direction for future work.

Some remarks

We inherit the assumption of *independent* and *regular distributions* from Myerson. In addition, we require the the distributions of valuations are *strictly regular* to guarantee that ties between bidders happen with 0 probability.[3] We note that the above allocation rule is monotone and allocates the ad spot to the bidder with the highest shifted valuation $\phi_{ij} + \alpha_{ij}$ for a given user. Thus it defines a unique truthful mechanism and the corresponding payment rule. We refer to α as the shift of the mechanism.

[3]We can drop this assumption by incorporating specific randomized tie-breaking rules, that retain fairness.

Algorithm 1 Algorithm1$(\mathcal{Q}, G, L, \eta, \mu_{\max}, \mu_{\min}, \varepsilon)$

Input: Constraint polytope $\mathcal{Q} \subseteq [0,1]^{n\times m}$, Lipschitz constant $G > 0$ of $\mathrm{rev}(\cdot)$, Lipschitz constant $L > 0$ of $f_{ij}(\cdot)$, the minimum coverage $\eta > 0$, the lower and upper bounds, $\mu_{\min}$ and $\mu_{\max}$, of $f_{ij}(\cdot)$, and a constant $\varepsilon > 0$ controlling the accuracy.
Output: Shifts $\alpha \in \mathbb{R}^{n\times m}$ for the optimal mechanism.

```
Initialize γ:=ε/2G², ξ:=(Gγ)², T=(√2G/ε)²
Compute q_1 := proj_Q(q(0_{n×m}))
Compute α_1 := Algorithm2(q_t, α_t, ξ, L, η, μ_max, μ_min)
for t = 1,2,...,T do
    Compute J_q(α_t)
    Compute rev(q_t) from
        J_q(α_t)^⊤ ∇rev(q_t) := ∇rev_shift(α_t)
    Update q_{t+1} := proj_Q(q_t + γ∇rev(q_t))
    Update
        α_{t+1} := Algorithm2(q_t, α_t, ξ, L, η, μ_max, μ_min)
end for
return α
```

4 Our Algorithm

Algorithm 1 solves the *optimal coverage problem* by performing a projected gradient descent in the constraint polytope, $\mathcal{Q}$. It starts with an initial coverage, $q_1 \in \mathcal{Q}$, and the corresponding approximate shift $\alpha_1 \in \mathbb{R}^{n\times m}$.[4]

At the step k, it calculates the gradient of $\mathrm{rev}(q_k)$, by solving the linear system $J_q(\alpha_k)^\top \nabla\mathrm{rev}(q_k) = \nabla\mathrm{rev}_{\mathrm{shift}}(\alpha_k)$. In order to solve the above linear system, we need to calculate $J_q(\alpha_k)^\top$ and $\nabla\mathrm{rev}_{\mathrm{shift}}(\alpha_k)$. This can be done in $O(n^2m)$ steps, if α_k is known (see Section 4.1). Therefore, the algorithm requires a "good" approximation of α at each step. It maintains this, by updating the previous approximation, α_{k-1}, at the k-th iteration. It does so by using another algorithm (Algorithm 2) to approximately solve the optimal shift problem.

After calculating the gradient, it takes a gradient step and projects the current iterate on $\mathcal{Q}$. This takes $O((nm)^\omega)$ time (see Section 4.2), where ω is the fast matrix multiplication coefficient. It continues this process for $\widetilde{O}(1/\varepsilon^2)$ iterations to get an ε-accurate α. This approximation of $\alpha^\star$ determines a solution to the *infinite-dimensional fair advertising problem*. By ensuring Algorithm 2 is ε^2 accurate, we can bound the total error introduced by approximations of α, and preserve the convergence rate of Algorithm 1.

Next we give the details of the projecting on $\mathcal{Q}$, and calculating the gradient $\nabla\mathrm{rev}$.

4.1 Calculating and Bounding $\nabla\mathrm{rev}(\cdot)$

Let $J_q(\alpha)$ be the Jacobian of the vectorized coverage, $\mathrm{vec}(q(\alpha)) \in \mathbb{R}^{(n-1)m}$, with respect to the vectorized shift, $\mathrm{vec}(\alpha) \in \mathbb{R}^{(n-1)m}$. Here, we fix the shift of one advertiser i for each user type

[4] We note that one advertiser's shift is fixed to 0, see Remark 4.1

$j \in [m]$.[5] Therefore, the Jacobian is a $(n-1)m \times (n-1)m$ matrix.

$$J_q(\alpha) = \begin{bmatrix} \frac{\partial q_{11}(\alpha)}{\alpha_{11}} & \dots & \frac{\partial q_{11}(\alpha)}{\alpha_{(n-1)1}} & \dots & \frac{\partial q_{11}(\alpha)}{\alpha_{(n-1)m}} \\ \frac{\partial q_{21}(\alpha)}{\alpha_{11}} & \dots & \frac{\partial q_{21}(\alpha)}{\alpha_{(n-1)1}} & \dots & \frac{\partial q_{21}(\alpha)}{\alpha_{(n-1)m}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial q_{(n-1)1(\alpha)}}{\alpha_{11}} & \dots & \frac{\partial q_{(n-1)1(\alpha)}}{\alpha_{(n-1)1}} & \dots & \frac{\partial q_{(n-1)1(\alpha)}}{\alpha_{(n-1)m}} \end{bmatrix}$$

To obtain $\nabla \mathrm{rev}(q)$, we use the fact that $J_q(\alpha)$ is always invertible (Lemma 4.2). Then, if we know $\alpha = q^{-1}(\delta)$ for some $\delta \in [0,1]^{n\times m}$, we can obtain the $\nabla \mathrm{rev}(q)$ by solving the following linear program.

$$(J_q(\alpha))^\top \nabla \mathrm{rev}(\delta) = \nabla \mathrm{rev}_{\text{shift}}(\alpha) \qquad \text{(Gradient oracle, 20)}$$

Remark 4.1. The Jacobian, $J_q(\alpha)$, is invertible only when we fix the shift of one advertiser for each user type $j \in [m]$. Intuitively, if we increase the α_{ij} for all advertiser and one user type, the coverage remains invariant. As such, we cannot hope for the $J_q(\alpha)$ to be invertible without fixing one α_{ij} for each $j \in [m]$.

Lemma 4.2. (Jacobian is invertible). For all $\alpha \in \mathbb{R}^{(n-1)\times m}$, if all advertisers have non-zero coverage for all user types $j \in [m]$, then the Jacobian $J_q(\alpha) \in \mathbb{R}^{(n-1)\cdot m \times (n-1)\cdot m}$, is invertible.

Proof. The coverage remains invariant if the bids of all advertisers are uniformly shifted for any given user type j. Therefore, for all $j \in [m]$ we have

$$\sum_{t\in[n]} \frac{\partial q_{ij}}{\partial \alpha_{tj}} = 0. \tag{21}$$

Since, increasing the shift α_{ij}, does not increase the coverage q_{kj} for any $k \neq i$, we have that

$$\frac{\partial q_{kj}}{\partial \alpha_{ij}} \leq 0 \text{ and } \frac{\partial q_{ij}}{\partial \alpha_{ij}} \geq 0. \tag{22}$$

Now, from Equation (21) we have

$$\forall\ i \in [n], j \in [m],\ \frac{\partial q_{ij}}{\partial \alpha_{ij}} = \sum_{t\in[n]\setminus\{i\}} \left|\frac{\partial q_{ij}}{\partial \alpha_{tj}}\right|. \tag{23}$$

Further since the n-th advertiser has non-zero coverage, i.e., there is non-zero probability that advertiser n bids higher than all other advertisers, changing α_{nj} must affect all other advertisers. In other words, for all $i \in [n-1]$ $\frac{\partial q_{ij}}{\partial \alpha_{nj}} \neq 0$. Using this we have,

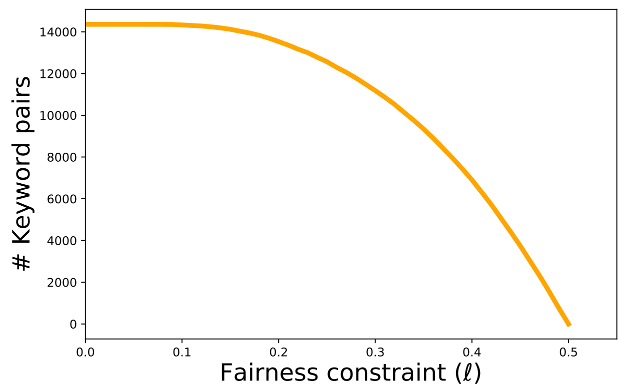
$$\forall\ i \in [n], j \in [m],\ \frac{\partial q_{ij}}{\partial \alpha_{ij}} > \sum_{t\in[n-1]\setminus\{i\}} \left|\frac{\partial q_{ij}}{\partial \alpha_{tj}}\right|. \tag{24}$$

By observing that q_{ij}, on user type j, is independent of the α_{st}, of any user type t, such that, $t \neq j$, i.e.,

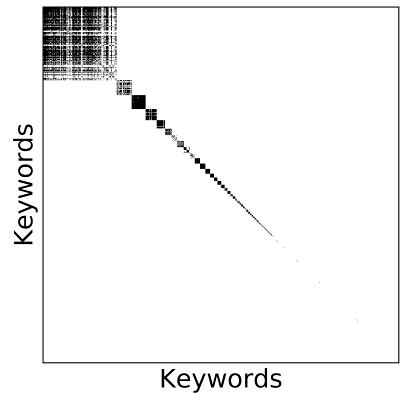
$$\forall\ i, s \in [n],\ j, t \in [m],\ s.t.\ j \neq t,\ \frac{\partial q_{ij}}{\partial \alpha_{st}} = 0, \tag{25}$$

and using Equation (23), we get that the Jacobian, $J_q(\alpha)$ is strictly diagonally dominant. Now, by the properties of strictly dominant matrices it is invertible. □

[5] In fact, we can relax this condition, by removing advertisers who have zero probability of winning.



(a) *Implicit Fairness of Keyword Pairs.* The x-axis represents fairness constraint ℓ (lower bound). We report number of keyword pairs satisfing each fairness level. We observe that 3282 auctions do not satisfy $\ell = 0.3$ fairness constraint.



(b) *Correlation among Keywords.* The axes depict keywords reordered to emphasize their correlation. Pairs sharing more than 1 advertisers are colored black. Each block can be interpreted as category (e.g., Science, Sports or Travel).

Figure 2

Remark 4.3. Since for any $i \in [n]$, q_{ij} is independent of α_{st} for any $s \in [n]$ (25). We claim that the Jacobian is sparse, and consists of only $n^2 m$ non-zero elements, which form m diagonal matrices of size $n \times n$, along the main diagonal of the Jacobian. This allows us to solve the linear system in Equation (20) in $O(n^\omega m)$ steps, where ω is the fast matrix multiplication coefficient.

4.2 Projection on Constraint Polytope ($\mathcal{Q}$)

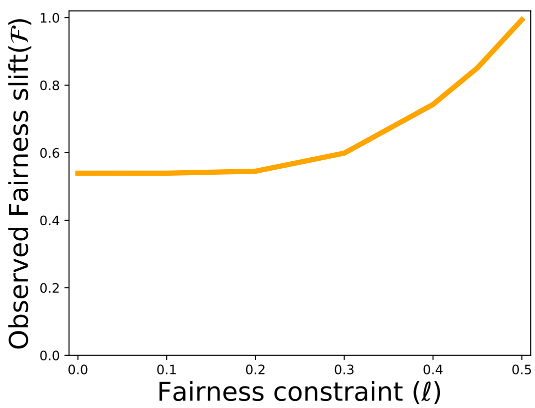
Given any point $q \in [0,1]^{n \times m}$, by determining the constraints it violates, we can express the projection on the constraint polytope, $\mathcal{Q}$, as a quadratic program with equality constraints. Using this we can construct a projection oracle $\text{proj}_{\mathcal{Q}}$, which given a point $q \in [0,1]^{n \times m}$, can project onto $\mathcal{Q}$ in $O((nm)^\omega)$ arithmetic operations, where ω is the fast matrix multiplication coefficient.

5 Empirical Study

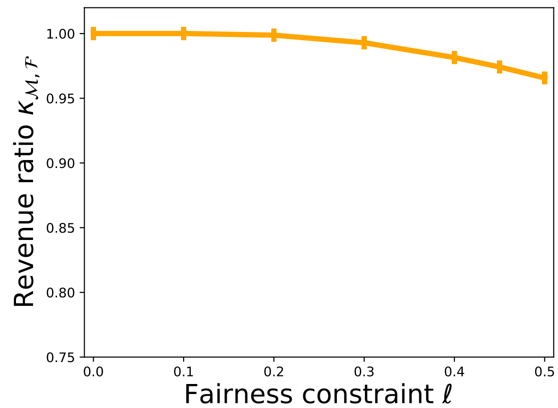
We evaluate our approach empirically on the Yahoo! A1 dataset [29]. We vary the strength of the fairness constraint for all advertisers, and find an optimal fair mechanism $\mathcal{F}$ using Algorithm 1 and compare it against the optimal unconstrained (and hence potentially unfair) mechanism $\mathcal{M}$, which is given by Myerson [20].

We first consider the impact of the fairness constraints on the revenue of the platform. Let $\text{rev}_{\mathcal{N}}$ denote the revenue of mechanism $\mathcal{N}$. We report the revenue ratio $\kappa_{\mathcal{M},\mathcal{F}} := {}^{\text{rev}_{\mathcal{F}}}/_{\text{rev}_{\mathcal{M}}}$. Note that the revenue of $\mathcal{F}$ can be at most that of $\mathcal{M}$, as it solves a constrained version of the same problem; thus $\kappa_{\mathcal{M},\mathcal{F}} \in [0,1]$.

We then consider the impact of the fairness constraints on the advertisers. Towards this, we consider the distribution of winners amongst advertisers in an auction given by $\mathcal{M}$ and an auction given by $\mathcal{F}$. We report the total variation distance $d_{TV}(\mathcal{M},\mathcal{F}) := {}^1/_2 \sum_{i\in[n]} |\sum_{j\in[m]} q_{ij}(\mathcal{M}) - q_{ij}(\mathcal{F})| \in [0,1]$ between the two distributions, as a measure of how much the winning distribution changes due to the fairness constraints.



(a) *Fairness.* We report the fairness slift($\mathcal{F}$) achieved by our fair ($\mathcal{F}$) mechanism for varying level of fairness.



(b) *Fairness and Revenue.* We report the revenue ratio $\text{rev}_{\mathcal{M},\mathcal{F}}$ between the fair ($\mathcal{F}$) and the unconstrained ($\mathcal{M}$) mechanisms.

Figure 3: The x-axis represents fairness constraint ℓ (lower bound). Error bars represent the standard error of the mean .

Lastly, we consider the fairness of the resultant mechanism $\mathcal{F}$. To this end, we measure selection lift (slift) achieved by $\mathcal{F}$, $\text{slift}(\mathcal{F}) \coloneqq \min_{i\in[n],j\in[m]}(q_{ij}/1-q_{ij}) \in [0,1]$. Where $\text{slift}(\mathcal{F}) = 1$, represents perfect fairness among the two user types.

5.1 Dataset

For the empirical results we use the Yahoo! A1 dataset [29], which contains bids placed by advertisers on the top 1000 keywords on *Yahoo! Online Auctions* between June 15, 2002 and June 14, 2003. The dataset has 10475 advertisers, and each advertiser places bids on a subset of keywords; there are approximately $2 \cdot 10^7$ bids in the dataset.

For each keyword k, let A_k be the set of advertisers that bid on it. We infer the distribution, $\mathcal{P}$, of valuation of an advertiser for a keyword by the bids they place on the keyword. In order to retain sufficiently rich valuation profiles for each advertiser, we remove advertisers who place less than 1000 bids on k, whose valuations have variance lower than $3 \cdot 10^{-3}$ from A_k, or who win the auction less than 5% of the time. This retains more than $1.5 \cdot 10^7$ bids.

The actual keywords in the dataset are anonymized; hence, in order to determine whether two keywords k_1 and k_2 are related, we consider whether they share more that one advertisers, i.e., $A_{k_1} \cap A_{k_2} > 1$. This allows us to identify keywords that are related (see Figure 2(b)), and hence for which spillover effects may be present as described in [17]. Drawing that analogy, one can think of each keyword in the pair as a different type of user for which the same advertisers are competing, and the goal would be for the advertiser to win an equal proportion of each user.
There are 14,380 such pairs. However, we observe that spillover does not affect all keyword pairs (see Figure 2(a)).

To test the effect of imposing fairness constrains in a challenging setting, we consider only the auctions which are not already fair; in particular there are 3282 keyword pairs which are less than $\ell = 0.3$ fair.

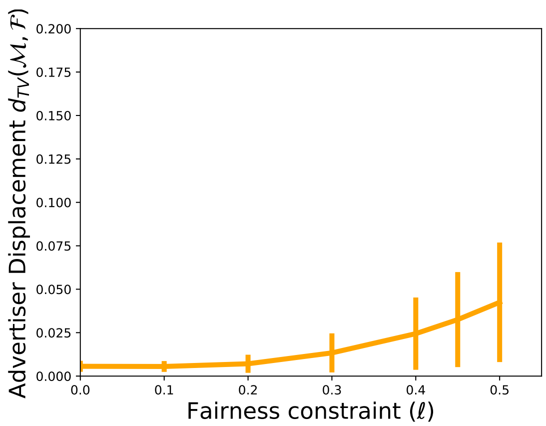


Figure 4: *Effect of fairness on advertisers.* The x-axis represents fairness constraint ℓ (lower bound). We report the $d_{TV}(\mathcal{M}, \mathcal{F})$ of the distribution of ads allocated by the fair ($\mathcal{F}$) and ($\mathcal{M}$). Error bars represent the standard error of the mean .

5.2 Experimental Setup

As we only consider pairs of keywords in this experiment, a lower bound constraint $\ell_{11} = \delta$ is equivalent to an upper bound constraint $u_{12} = 1 - \delta$. Hence, it suffices to consider lower bound constraints. We set $\ell_{i1} = \ell_{i2} = \ell \ \forall\ i \in [2]$, and vary ℓ uniformly from 0 to 0.5 , i.e., from the completely unconstrained case (which is equivalent to Myerson's action) to completely constrained case (which requires each advertiser to win each keywords in the pair with exactly the same probability). We report $\kappa_{\mathcal{N},\mathcal{M}}$ and $d_{TV}(\mathcal{N}, \mathcal{M})$ averaged over all auctions after 10^4 iterations in Figure 4; error bars represent the standard error of the mean over 10^4 iterations and 3282 auctions respectively.

5.3 Empirical Results

Fairness. Since the auctions are unbalanced to begin with, we expect the selection lift to increase with the fairness constraint. We observe a growing trend in the selection lift, eventually achieving perfect fair for $\ell = 0.5$.

Revenue Ratio. We do not expect to outperform the optimal unconstrained mechanism. However, we observe that even in the perfectly balanced setting with $\ell = 0.5$ our mechanisms lose less than 5% of the revenue.

Advertiser Displacement. Since the auctions are unbalanced to begin with, we expect TV-distance to grow with the fairness constraint. We observe this growing trend in the TV-distance on lowering the risk-difference. Even for zero risk-difference ($\ell = 0.5$) our mechanisms obtain a TV-distance is smaller than 0.05.

6 Proofs

6.1 Proof of Theorem 3.1

Proof. Let us introduce three Lagrangian multipliers, a vector $\alpha_j \in \mathbb{R}^n_{\geq 0}$, a vector $\beta_j \in \mathbb{R}^n_{\geq 0}$ and a continuous function $\gamma_j(\cdot)\colon \mathrm{supp}(\phi_j) \to \mathbb{R}_{\geq 0}\ \forall\ j \in [m]$, for the lower bound, upper bound, and single item constraints respectively. Then calculating the Lagrangian function we have,

$$L \coloneqq \sum_{j\in[m]} \Pr_{\mathcal{U}}[j] \sum_{i\in[n]} \int_{\mathrm{supp}(\phi_j)} \phi_{ij} x_{ij}(\phi_j) df_j(\phi_j) + \sum_{j\in[m]} \int_{\mathrm{supp}(\phi_j)} \gamma_j(\phi_j)\big(1 - \sum_{i\in[n]} x_{ij}(\phi_{ij})\big) df_j(\phi_j)$$
$$+ \sum_{\substack{i\in[n]\\ j\in[m]}} \alpha_{ij} \bigg(\int_{\mathrm{supp}(\phi_j)} x_{ij}(\phi_j) df_j(\phi_j) - \ell_{ij} \sum_{t\in[m]} \int_{\mathrm{supp}(\phi_t)} x_{it}(\phi_t) df_t(\phi_t) \bigg) + \sum_{\substack{i\in[n]\\ j\in[m]}} \beta_{ij} \bigg(u_{ij} \sum_{t\in[m]} \int_{\mathrm{supp}(\phi_t)} x_{it}(\phi_{it}) df_t(\phi_t) - \int_{\mathrm{supp}(\phi_j)} x_{ij}(\phi_{ij}) df_j(\phi_j) \bigg).$$

The second integral is well defined by from the continuity of $\gamma_j(\cdot)$ and monotonic nature of $x_j(\cdot)$. In order for the supremum of the Lagrangian over $x_{ij}(\cdot) \geq 0$ to be bounded, the coefficient of $x_{ij}(\cdot)$ must be non-positive. Therefore we require,

$$\int_{g\subseteq \mathrm{supp}(\phi_j)} \alpha_{ij} - \beta_{ij} + \Pr_{\mathcal{U}}[j]\phi_{ij} - \sum_{t\in[m]} (\alpha_{it}\ell_{it} - \beta_{it}u_{it}) - \gamma_j(\phi_j) df_j(\phi_j) \leq 0 \ \forall\ g \subseteq \mathrm{supp}(\phi_j), i \in [n],\ j \in [m].$$

Since $x_{ij}(\cdot)$ and $\gamma_j(\cdot)$ are continuous, the former is equivalent to

$$\alpha_{ij} - \beta_{ij} + \Pr_{\mathcal{U}}[j]\phi_{ij} - \sum_{t\in[m]} (\alpha_{it}\ell_{it} - \beta_{it}u_{it}) - \gamma_j(\phi_j) \leq 0 \quad \forall\ \phi_j, i \in [n],\ j \in [m].$$

If this holds, we can express the supremum of L as,

$$\sup_{x_{ij}(\cdot)\geq 0} L = \sum_{j\in[m]} \int_{\mathrm{supp}(\phi_j)} \gamma_j(\phi_j) df_j(\phi_j).$$

Now we can express the *dual optimization problem as:*

(Dual of the infinite-dimensional fair advertising problem). For all $j \in [m]$, find a optimal $\alpha_j \in \mathbb{R}^n_{\geq 0},\ \beta_j \in \mathbb{R}^n_{\geq 0}$ and $\gamma_j(\cdot)\colon \mathrm{supp}(\phi_j) \to \mathbb{R}_{\geq 0}$ for

$$\min_{\alpha_j,\ \beta_j \geq 0\ \gamma_j(\cdot)\geq 0} \sum_{j\in[m]} \int_{\mathrm{supp}(\phi_j)} \gamma_j(\phi_j) df_j(\phi_j) \tag{26}$$
$$\text{s.t. } \alpha_{ij} - \beta_{ij} + \Pr_{\mathcal{U}}[j]\phi_{ij} - \sum_{t\in[m]} (\alpha_{it}\ell_{it} - \beta_{it}u_{it}) \leq \gamma_j(\phi_j) \forall\ i \in [n], j \in [m], \phi_j. \tag{27}$$

Since the primal is linear in $x_{ij}(\cdot)$, and the constraints are feasible strong duality holds. Therefore, the dual optimal is primal optimal.

For any feasible constraints we have for all $i \in [n]$ $\sum_{j\in[m]} \ell_{ij} \leq 1$ and $\sum_{j\in[m]} u_{ij} \geq 1$. Therefore the coefficient of α_{ij}, $(1 - \sum_{j\in[m]} \ell_{ij}) \geq 0$, and that of β_{ij}, $(\sum_{j\in[m]} u_{ij} - 1) \geq 0$. Since α and β are non-negative, a optimal solution to the dual is finite. Let $\alpha^\star, \beta^\star$ be a optimal solutions to the dual, and $x^\star_{ij}(\cdot)$ be a optimal solution to the primal. Fixing α and β to their optimal values $\alpha^\star$ and $\beta^\star$ in the dual, let us define new virtual valuations ϕ'_{ij}, for all $i \in [n]$ and $j \in [m]$

$$\phi'_{ij} \coloneqq \phi_{ij} + \frac{1}{\Pr_{\mathcal{U}}[j]} \big(\alpha^\star_{ij} - \beta^\star_{ij} - \sum_{t\in[m]} (\alpha^\star_{it}\ell_{it} - \beta^\star_{it}u_{it}) \big).$$

Then the leftover problem has only one Lagrangian multiplier, $\gamma_j(\cdot)$. Let $\gamma'_j(\cdot)$ be the affine transformation of γ_j defined on virtual valuations, i.e., $\gamma'_j(\phi'_j) \coloneqq \gamma_j(\phi_j)$, then the problem can be expressed as follows.

(Dual with shifted virtual valuations). For all $j \in [m]$, find the optimal $\gamma_j'(\cdot)\colon \mathrm{supp}(\phi_j') \to \mathbb{R}_{\geq 0}$ for

$$\min_{\gamma_j(\cdot)\geq 0} \sum_{j\in[m]} \int_{\mathrm{supp}(\phi_j)} \gamma_j(\phi_j') df_j(\phi_j') \tag{28}$$

$$\text{s.t.} \quad \Pr_{\mathcal{U}}[j]\phi_{ij}' \leq \gamma_j(\phi_j') \quad \forall\ i \in [n], j \in [m], \phi'. \tag{29}$$

This is the dual of the following unconstrained revenue maximizing problem. Myerson's mechanism is the revenue maximizing solution to the unconstrained optimization problem. Further, by linearity and feasibility of constraints strong duality holds. Therefore the α'-shifted mechanism, for $\alpha' = {}^1\!/\!\Pr_{\mathcal{U}}[j] \cdot \big(\alpha_{ij}^\star - \beta_{ij}^\star + \sum_{t\in[m]}(\alpha_{it}^\star \ell_{it} - \beta_{it}^\star u_{it})\big)$ is a optimal fair mechanism.

(Unconstrained primal for the infinite-dimensional fair advertising problem). For all $j \in [m]$, find the optimal allocation rule $x_j(\cdot)\colon \mathbb{R}^n \to [0,1]^n$ for,

$$\begin{aligned} &\max_{x_{ij}(\cdot)\geq 0} \mathrm{rev}_{\mathcal{M}}(x_1, x_2, \ldots, x_m) \\ &\text{s.t.} \sum_{i\in[n]} x_{ij}(\phi_j) \leq\ 1 \qquad\qquad \forall\ j \in [m], \phi_j \in \mathrm{supp}(\phi_j). \end{aligned}$$

Further, Myerson's mechanism is truthful if the distribution of valuations are regular and independent. Since α-shifted mechanism applies a constant shift to all valuation, it follows under the same assumptions that any α-shifted mechanism is also truthful, and therefore has a unique payment rule defined by its allocation rule. $\square$

6.2 Proof of Theorem 3.3

Supporting Lemmas. Towards the proof of Theorem 3.3 we require the following two Lemmas. The first lemma shows that $\mathrm{rev}(\cdot)$ is Lipschitz continuous. Its proof is presented in Section 6.3.

Lemma 6.1. (Revenue is Lipschitz). For all coverages q_1, q_2, if assumption (16), (17) and (19) are satisfied, then

$$|\mathrm{rev}(q_1) - \mathrm{rev}(q_2)| \leq \left(\frac{\mu_{\max}\rho}{\mu_{\min}\eta}\right) n^2 \|q_1 - q_2\|_F. \tag{30}$$

The next lemma is an algorithm to solve the *optimal shift problem.* Its proof is presented in Section 6.4

Lemma 6.2. (An algorithm to solve the optimal shift problem). There is an algorithm (Algorithm 2) which outputs $\alpha \in \mathbb{R}^{n\times m}$, such that, if assumptions (16), (17) and (18) are satisfied, then α is an ε-optimal solution for the optimal shift problem, i.e., $\mathcal{L}(\alpha) < \varepsilon$, in

$$\log\left(\frac{m\mathcal{L}(\alpha_1)}{\varepsilon}\right) \frac{n^3(L + n^2\mu_{\max}^2)}{(\eta\mu_{\min})^2} \text{ steps.}$$

Where the arithmetic operations in each step are bounded by calculating the one $\nabla\mathcal{L}$.

Proof of Theorem 3.3. Algorithm 1 starts with an initial coverage q_0, and performs a projected gradient descent on the polytope $\mathcal{Q}$. Since $\mathcal{Q}$ is convex, the projection doesn't increase the distance between q, and the optimal solution in $\mathcal{Q}$.

$$\forall\ q,\ \|\mathrm{proj}_{\mathcal{Q}}(q) - q^\star\|_2 \leq \|q - q^\star\|_2 \tag{31}$$

We calculate the shift, $\alpha_k = q^{-1}(q_k)$ using Algorithm 2. This introduces some error, $\xi > 0$, at each iteration. We fix the optimal value of ξ later in the proof. Let $z_{k+1} \in [0,1]^{n\times m}$, be the coverage reached by the gradient update at iteration $k+1$, i.e., $z_{k+1} = q_k + \gamma\nabla\text{rev}(q_k)$, and q_{k+1} be the coverage after calculating α_k , i.e., $q_{k+1} = q(\alpha_{k+1})$, where $\alpha_k = q^{-1}(\text{proj}_{\mathcal{Q}}(q_{k+1}))$. Due to the error ξ, q_{k+1} deviates from $\text{proj}_{\mathcal{Q}}(z_{k+1})$. By definition of xi, we have the following bound on the deviation.

$$\|\text{proj}_{\mathcal{Q}}(z_{k+1}) - q_{k+1}\|_2^2 \leq \xi \tag{Error from Algorithm 2, 32}$$

Combining Equation (31) and (32), with the triangle inequality we get,

$$\begin{aligned}\|q_{k+1} - q^\star\|_2^2 &= \|q_{k+1} - \text{proj}_{\mathcal{Q}}(z_{k+1}) + \text{proj}_{\mathcal{Q}}(z_{k+1}) - q^\star\|_2^2 \\ &\overset{(31),(32)}{\leq} \|z_{k+1} - q^\star\|_2^2 + \xi. \end{aligned} \tag{33}$$

We know the revenue $\text{rev}(\cdot)$ is a concave function of q. Consider the distance between $q^\star$ and q_k. Using the first order condition of concavity on this, we have,

$$\begin{aligned}\|z_{k+1} - q^\star\|_2^2 &= \|q_k + \gamma\nabla\text{rev}(q_k) - q^\star\|_2^2 \\ &\leq \|q_k - q^\star\|_2^2 + 2\gamma(\text{rev}(q_k - \text{rev}(q^\star) + \gamma^2\|\nabla\text{rev}(q_k)\|_2^2. \end{aligned} \tag{34}$$

Accounting for the projection onto the polytope, and error introduced by Algorithm 2, we have,

$$\begin{aligned}\|q_{k+1} - q^\star\|_2^2 &\overset{(33)}{\leq} \|z_{k+1} - q^\star\|_2^2 + \xi \\ &\overset{(34)}{\leq} \|q_k - q^\star\|_2^2 + 2\gamma(\text{rev}(q_k) - \text{rev}(q^\star)) + \gamma^2\|\nabla\text{rev}(q_k)\|_2^2 + \xi. \end{aligned} \tag{35}$$

Expanding the above recurrence, we get,

$$\|q_{k+1} - q^\star\|_2^2 \overset{(35)}{\leq} k\xi + \|q_1 - q^\star\|_2^2 + 2\sum_{i\in[k]}\gamma(\text{rev}(q_i) - \text{rev}(q^\star)) + \sum_{i\in[k]}\gamma^2\|\nabla\text{rev}(q_i)\|_2^2. \tag{36}$$

Substituting $\|q_{k+1} - q^\star\|_2^2 \geq 0$, and $\|q_1 - q^\star\|_2^2 \leq 1$ in Equation (36), we get,

$$k\xi + 1 + 2\sum_{i\in[k]}\gamma(\text{rev}(q_i) - \text{rev}(q^\star)) + \sum_{i\in[k]}\gamma^2\|\nabla\text{rev}(q_i)\|_2^2 \geq 0.$$

By rearranging, replacing the sum $\sum_{i\in[k]}\gamma(\text{rev}(q^\star) - \text{rev}(q_i))$ by its minimum element, and using $\|\nabla\text{rev}(q_i)\|_2 \leq G$, we get,

$$\text{rev}(q^\star) - \max_{i\in[k]}\big(\text{rev}(x_i)\big) \leq \frac{1 + k\xi + G^2\sum_{i\in[k]}\gamma^2}{2\sum_{i\in[k]}\gamma}.$$

Choosing $\xi := G^2\gamma^2$, and substituting $\gamma := {}^{\varepsilon}/{}_{2G^2}$, $k := \left({}^{\sqrt{2}G}/{}_{\varepsilon}\right)^2$ we get,

$$\text{rev}(q^\star) - \max_{i\in[k]}\big(\text{rev}(x_i)\big) \leq \frac{1 + k\xi + G^2\sum_{i\in[k]}\gamma^2}{2\sum_{i\in[k]}\gamma} \leq \varepsilon.$$

At each step we perform a small update to q_k, and query Algorithm 2 for the shift, as such, Algorithm 2 is always warm-started. For all k, $\|z_{k+1} - q_k\|_2^2 < G\gamma$. Therefore, by Lemma 6.2, the total steps required to update α are,

$$\begin{aligned}
&\textstyle\sum_{i\in[k]} \log\left(\frac{mG\gamma}{\xi}\right)\frac{n^3(L+n^2\mu_{\max}^2)}{(\eta\mu_{\min})^2} \\
&= (\frac{\sqrt{2}G}{\varepsilon})^2 \log\left(\frac{2mG}{\varepsilon}\right)\frac{n^3(L+n^2\mu_{\max}^2)}{(\eta\mu_{\min})^2}.
\end{aligned}$$

Now, the sum of the total gradient steps by Algorithm 1, and the total gradient steps by all calls of Algorithm 2 are,

$$O\left(\frac{G^2}{\varepsilon^2}\log\left(\frac{2mG}{\varepsilon}\right)\frac{n^3(L+n^2\mu_{\max}^2)}{(\eta\mu_{\min})^2}\right).$$

Using $G = {}^{\mu_{\max}\rho}/_{\mu_{\min}\eta}n^2$, from Lemma 6.1, we have that Algorithm 1 gets an ε-approximation of optimal revenue in

$$\widetilde{O}\left(\frac{n^7\log m}{\varepsilon^2}\frac{(\mu_{\max}\rho)^2}{(\mu_{\min}\eta)^4}(L+n^2\mu_{\max}^2)\right) \text{ steps.}$$

Where $\widetilde{O}$ hides log factors in $n, \rho, \eta, \mu_{\max}, {}^1/\varepsilon$ and ${}^1/\mu_{\min}$. □

6.3 Proof of Lemma 6.1

We use the following lemmas in the proof of Lemma 6.1. The two lemmas split the Lipshitz continuity of rev$(\cdot)$ into the Lipschitz continuity of $\text{rev}_{\text{shift}}(\cdot)$ and $\alpha_{ij} = q_{ij}^{-1}(\cdot)$ respectively. Their proofs are follow in Section 6.3.1 and Section 6.3.2 respectively.

Lemma 6.3. (Revenue is Lipschitz continuous in shifts). For all $\alpha \in \mathbb{R}^{(n-1)\times m}$, if pdf, $f_{ij}(\phi)$ of the virtual valuations is bounded above by $\mu_{\max}$, and ϕ_{ij} is bounded above by ρ $\forall\, i \in [n],\ j \in [m]$, then $\text{rev}_{\text{shift}}(\alpha)$ is $(\mu_{\max}\rho n^{\frac{3}{2}})$-Lipschitz continuous.

Lemma 6.4. (Shifts is Lipschitz continuous in coverage). For all $\alpha, \beta \in \mathbb{R}^{(n-1)\times m}$, such that $q_{ij}(\beta + t(\alpha - \beta)) > \eta$, if the probability density function, $f_{ij}(\cdot)$, of virtual valuations is bounded by $\mu_{\min}$ and $\mu_{\max}$ $\ \forall\ t \in [0,1],\ i \in [n],\ j \in [m]$, then

$$\|\alpha - \beta\|_F < \frac{\sqrt{n}}{\eta\mu_{\min}}\|q(\alpha) - q(\beta)\|_2.$$

Proof of Lemma 6.1. Let $\alpha,\, \beta \in \mathbb{R}^{(n-1)\times m}$ be the shifts achieving q_1 and q_2 respectively. Then by Lemma 6.3 and Lemma 6.4 we have,

$$|\text{rev}(q(\alpha)) - \text{rev}(q(\beta))| \overset{\text{Lemma 6.3}}{\leq} \mu_{\max}\rho n^{\frac{3}{2}}\|\alpha - \beta\|_F \tag{37}$$

$$\|\alpha - \beta\|_F \overset{\text{Lemma 6.4}}{<} \frac{\sqrt{n}}{\eta\mu_{\min}}\|q(\alpha) - q(\beta)\|_2. \tag{38}$$

By combining Equation (37) and Equation (38) we get the required result,

$$|\text{rev}(q_1) - \text{rev}(q_2)| \overset{(37),(38)}{<} \frac{\mu_{\max}\rho}{\mu_{\min}\eta}n^2\|q_1 - q_2\|_2. \tag{39}$$

□

For all $j, k \in [m],\ i \in [n-1], \text{s.t.} j \neq k$

$$\frac{\partial \text{rev}_{\text{shift},\ j}(\alpha)}{\partial \alpha_{ij}} = \Pr_{\mathcal{U}}[j] \sum_{k \neq i} \int_{\text{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy \tag{41}$$

$$- \Pr_{\mathcal{U}}[j] \sum_{k \neq i} \int_{\text{supp}(f_{kj})} y f_{kj}(y)\ f_{ij}(y + \alpha_{kj} - \alpha_{ij}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{kj} - \alpha_{\ell j}) dy$$

$$\frac{\partial \text{rev}_{\text{shift},\ j}(\alpha)}{\partial \alpha_{ik}} = 0 \tag{42}$$

Figure 5: *Gradient of* $\text{rev}_{\text{shift},j}(\cdot)$. Equations from the proof of Lemma 6.3.

6.3.1 Proof of Lemma 6.3

Proof. We first consider the revenue for one user type j, $\text{rev}_{\text{shift},\ j}(\alpha)$, and then combine the result across all user type to show that $\text{rev}_{\text{shift}}(\alpha)$ is Lipschitz continuous. Formally, we define $\text{rev}_{\text{shift},\ j}(\alpha)$ as,

$$\text{rev}_{\text{shift},\ j}(\alpha) \coloneqq \sum_{i \in [n]} \Pr_{\mathcal{U}}[j] \int_{\text{supp}(f_{ij})} y f_{ij}(y) \prod_{k \in [n] \setminus \{i\}} F_{kj}(y + \alpha_{ij} - \alpha_{kj}) dy. \quad \text{(Revenue from user type } j\text{, 40)}$$

Then the total revenue $\text{rev}_{\text{shift}}(\alpha)$ is just a sum of $\text{rev}_{\text{shift},\ j}(\alpha)$ for all user types.

$$\text{rev}_{\text{shift}}(\alpha) = \sum_{j \in [m]} \text{rev}_{\text{shift},\ j}(\alpha)$$

We can express $\nabla \text{rev}_{\text{shift},\ j}(\alpha)$ as shown in Figure 5. We can observe that every term in the gradient (Equation (41), Equation (42)) is a linear function of $f_{ij}(\cdot)$ and $F_{ij}(\cdot)$ for some $i \in [n]$ and $j \in [m]$. Since, each term in the gradient (Equation (41)) involves at most $2n$ terms of the form of Equation (43) for some $i, k, \ell \in [n]$ and $j \in [m]$,

$$\int_{\text{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy. \tag{43}$$

Bounding this term, for all $i, k, \ell \in [n]$ and $j \in [m]$ by $\mu_{\max}\rho$ would give us a bound on $\nabla \text{rev}_{\text{shift}}(\alpha)$. To this end, consider

$$\left| \int_{\text{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy \right| \overset{(17)}{\leq} \mu_{\max} \left| \int_{\text{supp}(f_{ij})} y f_{ij}(y) dy \right| \quad \text{(Using } F_{ij}(\cdot) \leq 1\text{)}$$

$$\overset{(19)}{\leq} \mu_{\max}\rho. \tag{44}$$

For all $k \in [n]$, let terms $t_1(k)$ and $t_2(k)$ be defined as follows

$$t_1(k) \coloneqq \int_{\text{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy \tag{45}$$

$$t_2(k) \coloneqq \int_{\text{supp}(f_{kj})} y f_{kj}(y) f_{ij}(y + \alpha_{kj} - \alpha_{ij}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{kj} - \alpha_{\ell j}) dy. \tag{46}$$

Then rewriting the gradient from Figure 5, we have

$$\begin{aligned}
\left|\frac{\partial \mathrm{rev}_{\mathrm{shift},\ j}(\alpha)}{\partial \alpha_{ij}}\right| &= \Pr_{\mathcal{U}}[j] \sum_{k\in[n-1]\setminus\{i\}} \big(t_1(k)-t_2(k)\big) \\
&\overset{(44)}{\leq} \Pr_{\mathcal{U}}[j] \sum_{k\in[n-1]\setminus\{i\}} \mu_{\max}\rho \\
&\leq (n-2)\Pr_{\mathcal{U}}[j]\rho\mu_{\max}. \qquad (47)
\end{aligned}$$

Now calculating the Frobenius norm of $\mathrm{rev}_{\mathrm{shift,\ j}}(\alpha)$ we get

$$\begin{aligned}
\|\nabla \mathrm{rev}_{\mathrm{shift},\ j}(\alpha)\|_F^2 &= \sum_{\substack{i\in[n-1]\\ k\in[m]}} \left|\frac{\partial \mathrm{rev}_{\mathrm{shift},\ j}(\alpha)}{\partial \alpha_{ik}}\right|^2 \\
&\overset{(42)}{=} \sum_{i\in[n-1]} \left|\frac{\partial \mathrm{rev}_{\mathrm{shift},\ j}(\alpha)}{\partial \alpha_{ij}}\right|^2 \qquad (48) \\
&\overset{(47)}{\leq} \Pr_{\mathcal{U}}[j](n-1)((n-2)\rho\mu_{\max})^2. \qquad (49)
\end{aligned}$$

Now, we proceed to bound $\nabla \mathrm{rev}_{\mathrm{shift}}(\alpha)$,

$$\begin{aligned}
\|\nabla \mathrm{rev}_{\mathrm{shift}}(\alpha)\|_F^2 &= \sum_{\substack{i\in[n-1]\\ j\in[m]}} \left|\sum_{k\in[m]} \frac{\partial \mathrm{rev}_{\mathrm{shift},\ k}(\alpha)}{\partial \alpha_{ij}}\right|^2 \\
&\overset{(42)}{=} \sum_{\substack{i\in[n-1]\\ j\in[m]}} \left|\frac{\partial \mathrm{rev}_{\mathrm{shift},\ j}(\alpha)}{\partial \alpha_{ij}}\right|^2 \\
&\overset{(49)}{=} \sum_{j\in[m]} \|\mathrm{rev}_{\mathrm{shift},\ j}(\alpha)\|_F^2 \\
&\leq (n-1)((n-2)\rho\mu_{\max})^2 \sum_{j\in[m]} \Pr_{\mathcal{U}}[j] \\
&\leq (n-1)((n-2)\rho\mu_{\max})^2. \qquad (50)
\end{aligned}$$

Therefore, it follows that $\|\nabla \mathrm{rev}_{\mathrm{shift}}(\alpha)\|_F \leq n^{\frac{3}{2}}\rho\mu_{\max}$. □

6.3.2 Proof of Lemma 6.4

Technical lemmas. We use the following lemmas in the proof of Lemma 6.4. The first lemma is a lower bound on the derivative $q_{ij}(\alpha)$, and follows from assumptions (16) and (17).

Lemma 6.5. (Lower bound of derivative of $q_{ij}(\alpha)$). If for all $i \in [n],\ j \in [m]$ the probability density function, $f_{ij}(\cdot)$, of virtual valuations is bounded below by $\mu_{\min}$, and every advertiser has at least η coverage on every user type $j \in [m]$, then the absolute value of each gradient $\left|\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{sj}}\right|$ is lower bounded by $\eta\mu_{\min}$, i.e.,

$$\left|\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{sj}}\right| \geq \eta\mu_{\min}\ \forall\ i,s \in [n],\ j \in [m], \alpha \in \mathbb{R}^{n\times m}.$$

For all $i \in [n-1]$ and $j \in [m]$

$$q_{ij}(tu+\alpha) = \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) F_{nj}(y + tu_i + \alpha_{ij}) \prod_{k \neq i,n} F_{kj}(y + t(u_i - u_k) + \alpha_{ij} - \alpha_{kj}) dy \tag{54}$$

$$\frac{\partial q_{ij}(tu+\alpha)}{\partial t} = u_i \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{nj}(y + tu_i + \alpha_{ij}) \prod_{k \neq i,n} F_{kj}(y + t(u_i - u_k) + \alpha_{ij} - \alpha_{kj}) dy \tag{55}$$

$$+ (u_i - u_k) \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) F_{nj}(y + tu_i + \alpha_{ij}) \sum_{k \neq i,n} f_{kj}(y + t(u_i - u_k) + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq i,k,n} F_{\ell j}(y + t(u_i - u_\ell) + \alpha_{ij} - \alpha_{\ell j}) dy$$

Figure 6: *Directional derivative of* $q_{ij}(\cdot)$. Equations from the proof of Lemma 6.6.

Proof. Each advertiser has at least η coverage on every type of user we have for all $i \in [n],\ j \in [m]$, i.e.,

$$q_{ij}(\alpha) = \int_{\mathrm{supp}(f_{ij})} f_{ij}(\phi) \prod_{k \in [n] \setminus \{i\}} F_{kj}(\phi + \alpha_{ij} - \alpha_{kj}) d\phi \geq \eta. \tag{51}$$

Now, considering $\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{sj}}$ we get,

$$\begin{aligned}
\left| \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{sj}} \right| &= \left| \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{sj}(y + \alpha_{sj} - \alpha_{ij}) \prod_{k \neq i,s} F_{kj}(y + \alpha_{ij} - \alpha_{kj}) dy \right| \\
&\geq \mu_{\min} \left| \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) \prod_{k \neq i,s} F_{kj}(y + \alpha_{ij} - \alpha_{kj}) dy \right| && (\text{Using } f_{ij}(\phi_{ij}) \geq \mu_{\min}) \\
&\geq \mu_{\min} \left| \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) \prod_{k \neq i} F_{kj}(y + \alpha_{ij} - \alpha_{kj}) dy \right| && (\text{Using } F_{ij}(\phi_{ij}) \leq 1) \\
&\overset{(51)}{\geq} \eta \mu_{\min}.
\end{aligned} \tag{52}$$

□

In the next lemma we extend the lower bound to the directional derivative of $q_{ij}(\alpha)$.

Lemma 6.6. (Lower bound of directional derivative of $q_{ij}(\alpha)$). Given a shift $\alpha_j \in \mathbb{R}^{n-1}$, $t_{\max} > 0$, and a direction vector $u \in \mathbb{R}^{n-1}$, s.t. $\|u\|_2 = 1$, if the probability density function, $f_{ij}(\cdot)$, of virtual valuations is bounded below by $\mu_{\min}$ and bounded above by $\mu_{\max}\ \forall\, i \in [n],\ j \in [m]$, and $q_{ij}(tu + \alpha_j) > \eta$ for all $t \in [0, t_{\max}]$, then for $i \in \mathrm{argmax}_{k \in [n-1]} |u_k|$ and for all $t \in [0, t_{\max}]$

$$\mathrm{sign}(u_i) \frac{\partial q_{ij}(tu + \alpha_j))}{\partial t} > \frac{\eta \mu_{\min}}{\sqrt{n}}. \tag{53}$$

Proof. Consider $i \in \mathrm{argmax}_{k \in [n-1]} |u_k|$. Advertiser i's bids are being increased faster than or equal to any other advertiser's. Recalling that the shift of advertiser n, $\alpha_{ij} = 0$ for all user types $j \in [m]$, using Equation (10) from the paper we can express $q_{ij}(tu+\alpha)$ and its gradient as shown in Figure 6.

Since $i \in \operatorname{argmax}_{k\in[n-1]} |u_k|$, we have

$$\begin{aligned}|u_i| &\geq |u_k| \\ \operatorname{sign}(u_i)u_i &\geq \max(u_k, -u_k) \\ (u_i - u_k)\operatorname{sign}(u_i) &> 0. \end{aligned} \tag{56}$$

Since $\|u\|_2 = \sum_{i\in[n-1]} |u_i|^2 = 1$, we can lower bound $|u_i|^2$, the maximum coordinate of $u \in \mathbb{R}^{n-1}$ by magnitude by $\frac{1}{n-1}$, i.e., $|u_i| \geq 1/\sqrt{n-1}$. Multiplying Equation (55) with $\operatorname{sign}(u_i)$ and using Equation (56) and the fact that the integrals involved are positive to lower bound the equation we get

$$\begin{aligned}\operatorname{sign}(u_i)\frac{\partial q_{ij}(tu+\alpha_j)}{\partial t} &\overset{(56)}{\geq} \operatorname{sign}(u_i)u_i\frac{\partial q_{ij}}{\partial \alpha_{nj}}\bigg|_{\alpha_j+tu} \\ &\overset{\text{Lemma 6.5}}{\geq} |u_i|\eta\mu_{\min} \\ &\geq \frac{\eta\mu_{\min}}{\sqrt{n-1}} && (\text{Using } |u_i| > 1/\sqrt{n-1}) \\ &> \frac{\eta\mu_{\min}}{\sqrt{n}}. \end{aligned}$$

□

Proof of Lemma 6.4. Consider a user type $j \in [m]$ and the corresponding shifts $\alpha_j, \beta_j \in \mathbb{R}^n$, where α_j, β_j are the j-th columns of α and β respectively. Let $u = \alpha_j - \beta_j$, then from Lemma 6.6 we have $\exists\ i \in [n-1]$, such that,

$$\forall\ t \in [0,1], \quad \left|\frac{\partial q_{ij}(tu+\beta_j)}{\partial t}\right| > \eta\mu_{\min}. \tag{57}$$

Consider this i, then from the fundamental theorem of calculus we have

$$\begin{aligned}\|q_j(\alpha_j) - q_j(\beta_j)\|_2^2 = \sum_{k\in[n]}\left|\int_0^1 \frac{\partial q_{kj}(tu+\beta_j)}{\partial t}dt\right|^2 &\geq \left|\int_0^1 \frac{\partial q_{ij}(tu+\beta_j)}{\partial t}dt\right|^2 \\ &\overset{(57)}{>} \frac{(\eta\mu_{\min})^2}{n}\left|\int_0^1 (tu+\beta_j)dt\right| \\ &> \frac{(\eta\mu_{\min})^2}{n}\|\alpha_j - \beta_j\|_2^2. \end{aligned} \tag{58}$$

Using Equation (58) for every user type j we get that $\|q(\alpha) - q(\beta)\|_F > (\eta\mu_{\min})^2/n \cdot \|\alpha - \beta\|_F$. □

6.4 Proof of Lemma 6.2

The following remark gives insight, allowing us to use a gradient-based algorithm to solve the *optimal shift problem.*

Remark 6.7. The gradient of the loss, $2\sum_{j\in[m],\ i\in[n]}(q_{ij}(\alpha) - \delta_{ij})\nabla q_{ij}(\alpha)$, is linear combination of $\nabla q_{ij}(\alpha)$ for all $i \in [n]$ and $j \in [m]$. Since the rows of the Jacobian, $\{\nabla q_j(\alpha)\}_{j\in[m]}$, are linearly independent, the gradient is never zero unless we are at the global minimum where $\delta = q(\alpha)$. This guarantees that the loss does not have any saddle-points or local-maximum, and that any local minimum is a global minimum.

Algorithm 2 Algorithm2($\delta, \alpha_1, \xi, L, \eta, \mu_{\max}, \mu_{\min}$)

Input: A target coverage $\delta \in [0,1]^{n\times m}$, an approximate shift $\alpha_1 \in \mathbb{R}^{n\times m}$, a constant $\xi > 0$ that controls the accuracy, Lipschitz constant $L > 0$ of $f_{ij}(\cdot)$, the minimum coverage $\eta > 0$, and the lower and upper bounds, $\mu_{\min}$ and $\mu_{\max}$, of $f_{ij}(\cdot)$.

Output: An approximation $\alpha \in \mathbb{R}^{n\times m}$ of shifts for δ.

1: Initialize $\gamma \coloneqq (4nL + 2n^3\mu_{\max}^2)^{-1}$
2: Initialize
   $T \coloneqq \log\left({mn^3\mathcal{L}(\alpha_1)}/{\varepsilon}\right) \frac{(L+n^2\mu_{\max}^2)}{(\eta\mu_{\min})^2}$
3: **for** t = 1,2,...,T **do**
4: Compute $\nabla L(\alpha_t) \coloneqq \nabla[\mathcal{L}_1(\alpha_t), \ldots, \mathcal{L}_m(\alpha_t)]$
5: Update $\alpha_{t+1} \coloneqq \alpha_t - \gamma\nabla L(\alpha_t)$
6: **end for**
7: **return** α

In addition, we use the following two lemmas in the proof of Lemma 6.2. First, to get an efficient complexity bound, we want to avoid an arbitrarily small gradient while being "far" from the optimal. To this end, the first lemma shows that if $\mathcal{L}(\alpha)$ is greater than ε, then the Frobenius norm of $\nabla\mathcal{L}(\alpha)$ is greater than $\sqrt{\varepsilon}$. The following lemma's proof is provided in Section 6.4.1.

Lemma 6.8. (Lower bounding $\nabla\mathcal{L}_j(\cdot)$). Given $\alpha_j \in \mathbb{R}^{n-1}$, such that $\mathcal{L}_j(\alpha_j) > \varepsilon$ and $q_{ij}(\alpha_j) > \eta$, if the probability density function, $f_{ij}(\cdot)$, of virtual valuations is bounded below by $\mu_{\min}$ $\forall\ i \in [n],\ j \in [m]$, then $\|\nabla\mathcal{L}_j(\alpha_j)\|_2 > \frac{2}{n-1}\sqrt{\varepsilon}\eta\mu_{\min}$.

Next, we would like $\nabla\mathcal{L}(\alpha)$ to be well behaved. Here, we show that $\nabla\mathcal{L}(\alpha)$, is $O(n(L+n^2\mu_{\max}^2))$-Lipschitz continuous. Therefore, at each step of the gradient descent, we are guaranteed a decrease in the loss proportional to $\|\nabla\mathcal{L}(\alpha)\|_2^2$. The following lemma's proof is presented in Section 6.4.2.

Lemma 6.9. (Gradient of $\mathcal{L}(\cdot)$ is Lipschitz). If the probability density function, $f_{ij}(\phi)$, of the virtual valuations, ϕ_{ij} is L-Lipschitz continuous and bounded above by $\mu_{\max}$, then $\nabla\mathcal{L}_j(\alpha_j)$ is $O(n(L+n^2\mu_{\max}^2))$-Lipschitz.

Now, at each step, if the loss is greater than ξ, we get an improvement by a factor of $1-\beta\xi$, where β does not depend on ξ. This gives us a complexity bound of $O(\log {1}/{\varepsilon})$.

Proof of Lemma 6.2. At each iteration of the algorithm we calculate $\nabla\mathcal{L}_j(\alpha)$ for all $j \in [m]$, i.e., we calculate $\nabla\mathcal{L}(\alpha)$. We note that this bounds the arithmetic calculations at one iteration.
We recall from Equation (25) that the shift for one user type do not affect the coverage for the other. Therefore we can independently find a optimal shift α_j for all each user type $j \in [m]$.
From Lemma 6.1 we have that $\mathcal{L}_j$ is $O(n(L+n^2\mu_{\max}^2))$-Lipschitz continuous. Let $L' \coloneqq O(n(L+n^2\mu_{\max}^2))$, for brevity. We can get an upper bound to $\mathcal{L}_j(\alpha_k)$ from the first order approximation of $\mathcal{L}_j$ at α_k, further using the update rule $\alpha_{k+1} = \alpha_k - \frac{1}{L'}\nabla\mathcal{L}_j(\alpha_k)$ we have,

$$\mathcal{L}_j(\alpha_{k+1}) \leq \mathcal{L}_j(\alpha_k) - \frac{1}{2L'}\|\nabla\mathcal{L}_j(\alpha_k)\|_2^2.$$

Let $\lambda \coloneqq \frac{2}{n-1}\eta\mu_{\min}$, then from Lemma 6.8 we have that $\nabla\mathcal{L}_j(\alpha)$ is lower bounded by $\sqrt{\mathcal{L}_j(\alpha_k)}\lambda$. Using this to lower bound the gradient we get,

$$\mathcal{L}_j(\alpha_k) - \mathcal{L}_j(\alpha_{k+1}) \geq \frac{1}{2L'}\|\nabla\mathcal{L}_j(\alpha_k)\|_2^2$$
$$\mathcal{L}_j(\alpha_{k+1}) \leq \mathcal{L}_j(\alpha_k) - \frac{\mathcal{L}_j(\alpha_k)\lambda^2}{2L'}.$$

By the above recurrence we get,

$$\mathcal{L}_j(\alpha_k) \leq \mathcal{L}_j(\alpha_0)\left(1 - \frac{\lambda^2}{2L'}\right)^k.$$

Setting $k := \log \frac{m\mathcal{L}(\alpha_0)}{\varepsilon} \frac{-1}{\log\left(1-\frac{\lambda^2}{2L'}\right)}$ we get that for all $j \in [m]$, $\mathcal{L}_j(\alpha_k) < \varepsilon/m$. Therefore,

$$\mathcal{L}(\alpha) = \sum_{j\in[m]} \mathcal{L}_j(\alpha_j) < \varepsilon.$$

Substituting $L' = O(n(L + n^2\mu_{\max}^2))$ we get that the algorithm outputs α, such that $\mathcal{L}(\alpha) < \varepsilon$ in

$$\log\left(\frac{m\mathcal{L}(\alpha_1)}{\varepsilon}\right)\frac{n^3(L+n^2\mu_{\max}^2)}{(\eta\mu_{\min})^2}\text{steps}.$$

□

6.4.1 Proof of Lemma 6.8

In the proof of Lemma 6.8 we use Lemma 6.10, which shows that any linear combination of $\nabla q_{ij}(\alpha)$ for all $i \in [n]$, with reasonably "large" weights is lower bounded. We note that Lemma 6.10 does not follow from linear independence of $\nabla q_{ij}(\alpha)\ \forall\ i \in [n]$ (Lemma 4.2), because linear combinations of linearly independent vectors can be arbitrary small while having "large" weights.

Lemma 6.10. Given $x \in \mathbb{R}^{n-1}$, such that $\|x\|_1 > 1$, if for all $i \in [n],\ j \in [m]$ the probability density function, $f_{ij}(\cdot)$, of virtual valuations is bounded below by $\mu_{\min}$, and $q_{ij}(\alpha_j) > \eta$ coverage on every user type $j \in [m]$, then,

$$\Big\| \sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha_j) \Big\|_2 > \frac{\eta\mu_{\min}}{n-1}\ \forall\ \alpha_j \in \mathbb{R}^{n-1}.$$

Proof. Without loss of generality consider a reordering of $(x_1, x_2, \ldots, x_n)$, s.t. for some $p \leq n-1$,

$$x_i \geq 0\ \forall\ i \leq p \tag{59}$$
$$x_i < 0\ \forall\ i > p. \tag{60}$$

Case A. $\sum_{i\in[p]} x_i < 1/2$ **:**
We can replace x by $-x$, since this does not change the norm $\big\|\sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha_j)\big\|_2$. Now replacing p by $(n-p-1)$ we get *Case B.*

Case B. $\sum_{i\in[p]} x_i \geq 1/2$ **:**
The coverage remains invariant if the bids of all advertisers are uniformly shifted for any given user type j. $(\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{nj})$. Therefore we have for all $i \in [n-1]$,

$$\begin{aligned}\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{ij}} + \sum_{k\in[n-1]\setminus\{i\}} \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} &\overset{(21)}{=} -\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{nj}} \\ &\overset{(22)}{=} \left|\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{nj}}\right| \\ &\overset{\text{Lemma 6.5}}{\geq} \eta\mu_{\max}.\end{aligned} \tag{61}$$

Calculating the weighted sum of Equation (62) over $i \in [p]$ with weights x_i we get,

$$\sum_{i\in[p]} x_i \Bigg(\sum_{k\in[n-1]} \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \Bigg) \overset{(62)}{>} \sum_{i\in[p]} x_i \eta \mu_{\min} > \frac{\eta\mu_{\min}}{2}.$$

On rearranging the LHS we get

$$\sum_{k\in[n-1]} \left(\sum_{i\in[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \right) > \frac{\eta\mu_{\min}}{2}.$$

Therefore, by the pigeonhole principle on elements of the outer sum, $\exists\ k \in [n-1]$, s.t.

$$\sum_{i\in[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \geq \frac{1}{n-1} \sum_{k\in[n-1]} \Bigg(\sum_{i\in[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \Bigg) \tag{62}$$

$$\geq \frac{\eta\mu_{\min}}{2(n-1)}. \tag{63}$$

From Equation (22) for all $i \in [p]$ and $k > p$, $\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} < 0$. Therefore, $k \leq p$ in Equation (63). From this we get

$$\begin{aligned}
\sum_{i\in[n-1]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} &= \sum_{i\in[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} + \sum_{i\in[n-1]\setminus[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \\
&\overset{(63)}{\geq} \frac{\eta\mu_{\min}}{2(n-1)} + \sum_{i\in[n-1]\setminus[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \\
&\overset{\text{Lemma 6.5}}{\geq} \frac{\eta\mu_{\min}}{2(n-1)} + \eta\mu_{\min} \sum_{i\in[n-1]\setminus[p]} (-x_i) \\
&\overset{(60)}{\geq} \frac{\eta\mu_{\min}}{2(n-1)}.
\end{aligned} \tag{64}$$

Therefore, $\exists\ k \in [n-1]$, such that, $\sum_{i\in[n-1]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} > \eta\mu_{\min}$. It follows that

$$\begin{aligned}
\Bigg\| \sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha) \Bigg\|_2^2 &= \sum_{t\in[m]} \sum_{k\in[n-1]} \Bigg(\sum_{i\in[n-1]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kt}} \Bigg)^2 \\
&\overset{(25)}{=} \sum_{k\in[n-1]} \Bigg(\sum_{i\in[n-1]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \Bigg)^2 \\
&\overset{(64)}{\geq} \left(\frac{\eta\mu_{\min}}{2(n-1)} \right)^2
\end{aligned} \tag{65}$$

$$\Bigg\| \sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha) \Bigg\|_2 \geq \frac{\eta\mu_{\min}}{2(n-1)}.$$

$\square$

Proof of Lemma 6.8. Since $\mathcal{L}_j(\alpha_j) \geq \varepsilon$, we have

$$\mathcal{L}_j(\alpha_j) = \sum_{i\in[n-1]} (\delta_{ij} - q_{ij}(\alpha_j))^2 \geq \varepsilon. \tag{66}$$

Further, using $\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_{i,k} 2a_i a_k$ we get

$$\mathcal{L}_j(\alpha_j) = \sum_{i\in[n-1]} (\delta_{ij} - q_{ij}(\alpha_j))^2 \leq \Big(\sum_{i\in[n-1]} |\delta_{ij} - q_{ij}(\alpha_j)| \ \Big)^2. \tag{67}$$

From these we have that,

$$\sum_{i\in[n-1]} |\delta_{ij} - q_{ij}(\alpha_j)| \overset{(67),(66)}{\geq} \sqrt{\varepsilon}. \tag{68}$$

Considering $x_i = \frac{1}{\sqrt{\varepsilon}}(\delta_{ij} - q_{ij}(\alpha_j))$ we have

$$\sum_{i\in[n-1]} |x_i| = \frac{1}{\sqrt{\varepsilon}} \sum_{i\in[n-1]} |\delta_{ij} - q_{ij}(\alpha)| > 1.$$

From Lemma 6.10 we have

$$\begin{aligned} \Big\| \sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha_j) \ \Big\|_2 &\overset{\text{Lemma 6.10}}{\geq} \frac{\eta\mu_{\min}}{n-1} \\ \Big\| \sum_{i\in[n-1]} 2(\delta_{ij} - q_{ij}(\alpha_j)) \nabla q_{ij}(\alpha_j) \ \Big\|_2 &\geq 2\sqrt{\varepsilon}\frac{\eta\mu_{\min}}{n-1}. \end{aligned}$$

□

6.4.2 Proof of Lemma 6.9

In order to show that the loss $\mathcal{L}(\cdot)$ is $O(n(L + n^2\mu_{\max}^2))$-Lipschitz continuous, we first show that ∇q_{ij} is $2n(L + n\mu_{\max}^2)$-Lipschitz continuous. To this end, we show that the elements of $\nabla^2 q_{ij}$ are bounded (Lemma 6.11), and then use Lemma 6.12 (Corollary 1.2 in [26]) to bound the magnitudes of the eigen-values.

Lemma 6.11. Given $\alpha_j \in \mathbb{R}^n$, if pdf, $f_{ij}(\phi)$ of the virtual valuations, ϕ_{ij} is L-Lipschitz continuous and bounded above by $\mu_{\max}$, then elements in the main diagonal of the Hessian, $\nabla^2 q_{ij}(\alpha_j)$ are bounded in absolute value by $n(L + n\mu_{\max}^2)$, and all other elements are bounded in absolute value by $L + n\mu_{\max}^2$, i.e.,

$$\begin{aligned} &\forall\ i \in [n], && \frac{\partial^2 q_{ij}}{\partial\alpha_{ij}\partial\alpha_{ij}} \leq n(L + n\mu_{\max}^2) \\ &\forall\ k,\ t \in [n], k \neq i \text{ or } t \neq i, && \frac{\partial^2 q_{ij}}{\partial\alpha_{kj}\partial\alpha_{tj}} \leq L + n\mu_{\max}^2. \end{aligned}$$

Proof. Consider the Hessian of $q_{ij}(\alpha_j)$ in Figure 7, which follows from differentiating Equation 10 with respect to α_j, where α_j is the j-th column of α. We note that $\frac{q_{ij}}{\alpha_{st}} = 0$ for any $t \neq j$, for all $i, s \in [n]$ and $j, t \in [m]$, and so we only need to calculate the gradient with respect to α_j.

We can observe that for all $i \in [n]$ and $j \in [m]$ every term in the Hessian is linear function of $f'_{ij}(y)$, $f_{ij}(y)$ and $F_{ij}(y)$. In particular each term in the Hessian is a sum of the following terms, for

For all distinct i, k, t in $[n]$

$$\frac{\partial^2 q_{ij}}{\partial \alpha_{ij} \partial \alpha_{ij}} = \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) \sum_{k \neq i} f'_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy \tag{69}$$
$$+ \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) \sum_{k \neq i} f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \sum_{\ell \neq i,k} f_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) \prod_{h \neq \ell,k,i} F_{hj}(y + \alpha_{ij} - \alpha_{hj}) dy$$

$$\frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{kj}} = \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f'_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy \tag{70}$$

$$\frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{ij}} = \frac{\partial^2 q_{ij}}{\partial \alpha_{ij} \partial \alpha_{kj}} = -\int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f'_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy \tag{71}$$
$$- \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \sum_{\ell \neq k,i} f_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) \prod_{h \neq \ell,k,i} F_{hj}(y + \alpha_{ij} - \alpha_{hj}) dy$$

$$\frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{tj}} = \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) f_{tj}(y + \alpha_{ij} - \alpha_{tj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy. \tag{72}$$

Figure 7: *Hessian of* $q_{ij}(\cdot)$. Equations from proof of Lemma 6.11.

some combinations of $i, k, \ell \in [n]$ and $j \in [m]$,

$$\int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) f_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) \prod_{h \neq \ell,k,i} F_{hj}(y + \alpha_{ij} - \alpha_{hj}) dy \tag{73}$$

$$\int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f'_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy. \tag{74}$$

Each term along the diagonal of the Hessian (Equation (69)), $\frac{\partial^2 q_{ij}}{\partial \alpha_{ij} \partial \alpha_{ij}}$, is a combination of $(n-1)$ terms of the form Equation (73), and n^2 terms of the form Equation (74). All other terms in the Hessian contain at most n terms of the form Equation (74), and 1 term of the form Equation (73). Bounding these terms for all $i, k, \ell \in [n]$ and $j \in [m]$ by $\mu_{\max}^2$ would give us a bound on terms of the Hessian, which in turn gives bounds on the eigen-values of the Hessain.
To this end, recall that for all $i \in [n]$, $j \in [m]$, and $y \in \mathrm{supp}(f_{ij})$

$$0 < f_{ij}(y) \leq \mu_{\max} \tag{75}$$
$$|f'_{ij}(y)| < L \tag{76}$$
$$0 \leq F_{ij}(y) \leq 1. \tag{77}$$
$$\int f_{ij}(y) dy = 1. \tag{78}$$

We can now bound Equation (73) and Equation (74) as follows

$$(73) \overset{(75),(77)}{\leq} \mu_{\max}^2 \Big| \int_{\mathrm{supp}(f_{ij})} f_{ij}(y)dy \Big| \overset{(78)}{\leq} \mu_{\max}^2 \tag{79}$$

$$(74) \overset{(76),(77)}{\leq} L \Big| \int_{\mathrm{supp}(f_{ij})} f_{ij}(y)dy \Big| \overset{(78)}{\leq} L. \tag{80}$$

Now we have for all $k,\ i \in [n]$, s.t., $k \neq i$,

- $$\begin{aligned} \left| \frac{\partial^2 q_{ij}}{\partial \alpha_{ij} \partial \alpha_{ij}} \right| &= \left| \sum_{k \neq i} (74) + \sum_{k \neq i} \sum_{\ell \neq i,k} (73) \right| \\ &\overset{(79),(80)}{\leq} (n-1)\big(L + (n-2)\mu_{\max}^2\big) \qquad \text{(Using triangle inequality)} \\ &\leq n(L + n\mu_{\max}^2) \end{aligned} \tag{81}$$
- $$\left| \frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{kj}} \right| = \big|(74)\big| \overset{(80)}{\leq} L \tag{82}$$
- $$\left| \frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{tj}} \right| = \big|(73)\big| \overset{(79)}{\leq} \mu_{\max}^2 \tag{83}$$
- $$\begin{aligned} \left| \frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{ij}} \right| &= \left| \frac{\partial^2 q_{ij}}{\partial \alpha_{ij} \partial \alpha_{kj}} \right| = \left| (74) + \sum_{\ell \neq k,i} (73) \right| \\ &\overset{(79),(80)}{\leq} L + (n-2)\mu_{\max}^2 \end{aligned} \tag{84}$$

□

Lemma 6.12. (Corollary 1.2 in [26]) For any matrix $A \in \mathbb{R}^{n \times n}$, and any eigen-value $\lambda \in \mathbb{R}$ of A,

$$\lambda \leq \max_{i \in [n]} \sum_{j \in [n]} |A_{ij}|.$$

We refer the reader to [26] for a proof of the above lemma.

Proof of Lemma 6.9. To show that $\nabla \mathcal{L}_j(\alpha_j)$ is Lipschitz continuous, we show that $q_{ij}(\alpha_j)$ is Lipshitz continuous, then use the fact that $\nabla q_{ij}(\alpha_j)$ is Lipschitz continuous from Lemma 6.12, and that δ_j and $q_{ij}(\cdot)$ have bounded sums if the loss is greater than ε. To this end we recall,

$$\mathcal{L}_j(\alpha_j) \coloneqq \sum_{i \in [n-1]} (\delta_{ij} - q_{ij}(\alpha_j))^2$$

$$\nabla \mathcal{L}_j(\alpha_j) = -2 \sum_{i \in [n-1]} (\delta_{ij} - q_{ij}(\alpha_j)) \nabla q_{ij}(\alpha_j)$$

Consider the following term for some $i, k \in [n]$ and $j \in [m]$,

$$t(k) \coloneqq \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy. \tag{85}$$

Now we can express $\left|\frac{\partial q_{ij}}{\partial \alpha_{ij}}\right|$ and $\left|\frac{\partial q_{ij}}{\partial \alpha_{kj}}\right|$ $\forall\ i,k \in [n]$ and $k \neq i$ as follows,

$$\begin{aligned}
\left|\frac{\partial q_{ij}}{\partial \alpha_{ij}}\right| &= \Big|\sum_{k\in[n]\setminus\{i\}} t(k)\Big| \leq (n-1)|t(i)| \\
&\leq (n-1)\mu_{\max}\Bigg|\int\limits_{\mathrm{supp}(f_{ij})} f_{ij}(y)dy\Bigg| && \text{(Using } f_{ij}(\phi_{ij}) \leq \mu_{\max} \text{ and } F_{ij}(\phi_{ij}) \leq 1) \\
&\overset{(78)}{\leq} (n-1)\mu_{\max}. && \text{(Using } \int_{\mathrm{supp}(f_{ij})} f_{ij}(z)dz = 1) \qquad (86) \\
\left|\frac{\partial q_{ij}}{\partial \alpha_{kj}}\right| &= \big|t(k)\big| \\
&\overset{(77)}{\leq} \mu_{\max}\Bigg|\int\limits_{\mathrm{supp}(f_{ij})} f_{ij}(y)dy\Bigg| && \text{(Using } f_{ij}(\phi_{ij}) \leq \mu_{\max} \text{ and } F_{ij}(\phi_{ij}) \leq 1) \\
&\overset{(78)}{\leq} \mu_{\max}. && \text{(Using } \int_{\mathrm{supp}(f_{ij})} f_{ij}(z)dz = 1) \qquad (87)
\end{aligned}$$

Now we can show that the gradient of $q_{ij}(\alpha_j)$ is bounded, i.e., $q_{ij}(\alpha_j)$ is Lipschitz continuous. For this consider $\|\nabla q_{ij}(\alpha_j)\|$,

$$\begin{aligned}
\|\nabla q_{ij}(\alpha_j)\|_2^2 = \sum_{k\in[n]} \left(\left|\frac{\partial q_{ij}}{\partial \alpha_{kj}}\right|^2\right) &\leq \left|\frac{\partial q_{ij}}{\partial \alpha_{ij}}\right|^2 + \sum_{k\in[n]\setminus\{i\}} \left(\left|\frac{\partial q_{ij}}{\partial \alpha_{kj}}\right|^2\right) \\
&\overset{(86),(87)}{\leq} (n-1)^2\mu_{\max}^2 + n\mu_{\max}^2 \\
&\leq n^2\mu_{\max}^2. \qquad (88)
\end{aligned}$$

Since $q_{ij}(\alpha_j)$ and δ_{ij} represent the probabilities of advertisers winning they sum to 1. Therefore, for all user type $j \in [m]$, their sum is bounded by 1, i.e., $\sum_{i\in[n]} q_{ij}(\alpha_j) \leq 1$ and $\sum_{i\in[n]} \delta_{ij} \leq 1$. Using the triangle inequality we get

$$\sum_{i\in[n-1]} |\delta_{ij} - q_{ij}(\alpha_j)| \leq \sum_{i\in[n-1]} \Big(|\delta_{ij}| + |q_{ij}(\alpha_j)|\Big) \leq 2. \qquad (89)$$

We represent the Hessian $\nabla^2 q_{ij}(\alpha_j)$ by $H(\alpha_j)$ for brevity. Then the Hessian of $\mathcal{L}(\cdot)$ is

$$\nabla^2\mathcal{L}_j(\alpha_j) = 2\sum_{i\in[n-1]} \nabla q_{ij}(\alpha_j)\nabla q_{ij}(\alpha_j)^\top - \sum_{i\in[n-1]} (\delta_{ij} - q_{ij}(\alpha_j))H(\alpha_j). \qquad (90)$$

We know from Lemma 6.12 that the eigen-values of $H(\alpha_j)$ are bounded in absolute value by $2n(L + n\mu_{\max}^2)$. We also know that the only non-zero eigen-value of $vv^\top$ for any vector v is $\|v\|_2^2$.

Let $\|X\|_\star$ be the spectral-norm of matrix X, which is defined as the maximum singular value of X. Then, since singular-values are absolute values of the eigen-values the spectral norm of $H(\alpha_j)$ and $vv^\top$ are bounded. Specifically,

$$\|H(\alpha_j)\|_\star \overset{\text{Lemma 6.12}}{\leq} 2n(L + n\mu_{\max}^2) \qquad (91)$$

$$\|q_{ij}(\alpha_j)q_{ij}(\alpha_j)^\top\|_\star \leq \|q_{ij}(\alpha_j)\|_2^2 \overset{(88)}{\leq} n^2\mu_{\max}^2. \qquad (92)$$

Now, we use the sub-additivity of the spectral-norm, represented as $\|\cdot\|_\star$.

$$\|A+B\|_\star \le \|A\|_\star + \|B\|_\star \qquad \text{(Sub-additivity of } \|\cdot\|_\star\text{, 93)}$$

This gives us the following

$$\begin{aligned}\|\nabla^2 \mathcal{L}_j(\alpha_j)\|_\star &\overset{(91)}{\le} 2 \sum_{i\in[n-1]} \|\nabla q_{ij}(\alpha_j)\nabla q_{ij}(\alpha_j)^\top\|_\star + (\delta_{ij} - q_{ij}(\alpha_j))\|H(\alpha_j)\|_\star \\ &\overset{(91),(92)}{\le} 2 \sum_{i\in[n-1]} n^2\mu_{\max}^2 + 4n(L+n\mu_{\max}^2) \sum_{i\in[n-1]} (\delta_{ij} - q_{ij}(\alpha_j)) \\ &\overset{(89)}{\le} 2n^3\mu_{\max}^2 + 4n(L+n\mu_{\max}^2).\end{aligned}$$

Therefore, $\|\nabla^2\mathcal{L}_j(\alpha_j)\|_\star \le O(n(L+n^2\mu_{\max}^2))$, and the eigen-values of $\|\nabla^2\mathcal{L}_j(\alpha_j)\|_\star$ are bounded in absolute value by $O(n(L+n^2\mu_{\max}^2))$. □

7 Conclusion and Future Work

We initiate a formal study of designing auction-based mechanisms for online advertising that can ensure advertisements are not shown disproportinately to different populations. This is especially relevant for ads for employment opportunities, housing, and other regulated markets where biases in the recipient population can be illegal and/or unethical. As has been shown recently, existing platforms suffer from various spillover effect that result in such biased distributions. Our approach places constraints on the allocations given from an ad across different sub-populations in order to ensure fair exposure of the content. It can be used flexibly placing constraints on some or all advertisers, across some or all sub-populations, and varying the tightness of the constraint depending on the level of fairness desired.

We present a truthful mechanism which attains the optimal revenue while satisfying the constraints necessary to attain such fairness, and present an efficient algorithm for finding this mechanism given the advertiser properties and fairness constraints. Empirically, we observe that our mechanisms can satisfy fairness constraints at a minor loss to the revenue of the platform, even when the constraints ensure it attains perfect fairness. Hence, fairness is not necessarily at odds with maximizing the platform's ad revenue. Furthermore, we show empirically that advertisers are not significantly impacted with respect to their winning percentages – the sub-populations their ads are shown to change to be fair, but overall they are still able to display similar numbers of ads.

This work leaves open several interesting directions. On the technical side, it would be interesting to improve Theorem 3.3 by weakening the assumptions on the distributions, or by deriving better complexity bounds. Additionally, we consider disjoint user types. It would be important to extend our results to an arbitrary group structures and improve fairness across intersectional types. Empirically, it would be important to test such algorithms in the field, and in particular, study the effect on the users and how the profile of advertisements they see is impacted by the constraints.

References

[1] Alphabet Inc. revenue. Online advertisements contribute 85% to alphabet inc.'s total revenue. source: Alphabet Inc. Q3 2018 Results.

[2] L. Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. Controlling polarization in personalization: An algorithmic framework. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 160–169, New York, NY, USA, 2019. ACM.

[3] Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, Sep 1971.

[4] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *CoRR*, abs/1408.6491, 2014.

[5] Facebook, Ad Targeting.

[6] Facebook Ads. Facebook: Advertising Policies, Discriminatory practices.

[7] Facebook, What's a bid? About the delivery system: Bid strategies.

[8] Ayman Farahat and Michael C. Bailey. How effective is targeted advertising? In Alain Mille, Fabien L. Gandon, Jacques Misselis, Michael Rabinovich, and Steffen Staab, editors, *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012*, pages 111–120. ACM, 2012.

[9] Thomas Fox-Brewster. Creepy Or Cool? Twitter Is Tracking Where You've Been, What You Like And Is Telling Advertisers. *Forbes Magazine*, May 2017.

[10] Avi Goldfarb and Catherine Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, May 2011.

[11] Google, Ad Rank.

[12] Google, Automated Bidding.

[13] Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–31, 1973.

[14] Jessica Guynn. Facebook accused of bias in job ads that let companies target men, exclude women. *USA Today*, September 2018.

[15] Jason D. Hartline. *Mechanism Design and Approximation*. July 2017. Book draft.

[16] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[17] Anja Lambrecht and Catherine E. Tucker. Algorithmic bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. March 2018.

[18] S. Muthukrishnan. Ad exchanges: Research issues. In Stefano Leonardi, editor, *Internet and Network Economics, 5th International Workshop, WINE 2009, Rome, Italy, December 14-18, 2009. Proceedings*, volume 5929 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2009.

[19] Roger B Myerson. Incentive Compatibility and the Bargaining Problem. *Econometrica*, 47(1):61–73, January 1979.

[20] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[21] NFHA. Facebook sued by civil rights groups for discrimination in online housing advertisements. *National Fair Housing Alliance*, March 2018.

[22] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[24] Latanya Sweeney. Discrimination in online ad delivery. *Commun. ACM*, 56(5):44–54, 2013.

[25] Craig Timberg and Tracy Jan. HUD secetary carson accuses Facebook of enabling housing discrimination. *The Washington Post*, August 2018.

[26] R.S. Varga. *Gervsgorin and His Circles*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2011.

[27] Iris Vermeren. Men vs. Women: Who Is More Active on Social Media? *Brandwatch*, January 2015.

[28] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

[29] Yahoo. A1 - Yahoo! Search Marketing Advertising Bidding Data, Version 1.0.

[30] Jun Yan, Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen. How much can behavioral targeting help online advertising? In Juan Quemada, Gonzalo León, Yoelle S. Maarek, and Wolfgang Nejdl, editors, *Proceedings of the 18th International Conference on World Wide Web, WWW 2009*, pages 261–270. ACM, 2009.

[31] Shuai Yuan, Ahmad Zainal Abidin, Marc Sloan, and Jun Wang. Internet advertising: An interplay among advertisers, online publishers, ad exchanges and web users. *CoRR*, abs/1206.1754, 2012.

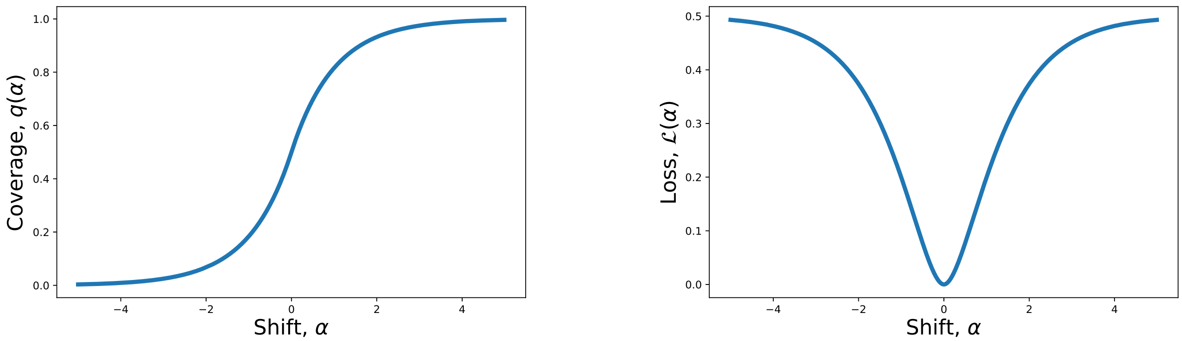
A A Simple Example of Competitive Spillover

In this section we discuss the example of competitive spillover presented the introduction more concretely.

Let the two advertisers have an equal budget of \$30. Both of them place a bid of \$1, if they target the current user and otherwise place a bid of \$0. For the sake of simplicity, let us assume that men and women visit the platform alternately. Consider an auction mechanism that shows the ad of the highest bidder. If there is more than one bidder with the same highest bid, the auction mechanism chooses one uniformly at random.

Whenever a man visits the webpage, the second advertiser places a bid of \$1, while the first advertiser places a bid of \$0. Therefore, the mechanism always shows the second advertiser's ad.Whereas when a women visits the platform, both the advertisers place a bid of \$1, and one of them shows the advertisement with a 50% probability.

Consider the point when 40 users have visited the platform, 20 men and 20 women. The second advertiser has shown 20 ads to men, and 10 ads to women. Whereas the first advertiser has only



(a) *Coverage as a function of shift. (Non-Convex)* Coverage for one of the two advertisers with exponentially distributed bids, on two user types. We vary the shift of one of the advertisers and report its coverage as a function of the shift.

(b) *Loss as a function of shifts. (Non-Convex)* The loss $\mathcal{L}(\alpha)$, for two advertisers with exponential valuations, and $\delta = (0.5, 0.5)$. We vary the shift of one of the advertisers and report its coverage as a function of the shift.

Figure 8

displays 10 ads to females. Having shown 30 ads, the second advertiser has finished the budget, and leaves the auction, while the first advertiser stays till another 20 women visit the platform.

In such a situation, the second advertiser who meant the ad to be unbiased among users, ends up under-representing women in the viewers of the ad.

B Revenue is Non-Concave in α

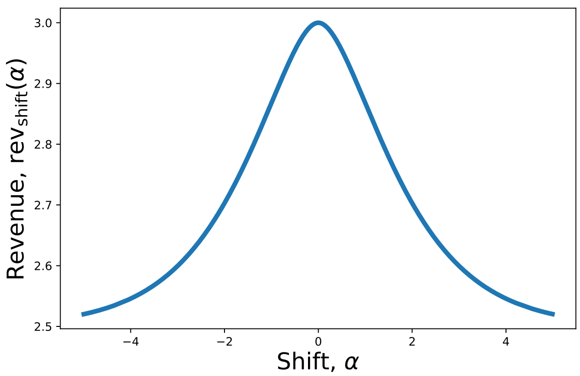
Consider two advertisers and one user type with $f_{11}(x) = e^{-x}$ and $f_{21}(x) = e^{-x}$. We fix the shift of advertiser 2 to 0, and consider a positive shift $\alpha \geq 0$ of advertiser 1. Then,

$$\begin{aligned}
\mathrm{rev}_{\mathrm{shift}}(\alpha) &= \int\limits_{\mathrm{supp}(f_{11})} y f_{11}(y) F_{21}(y+\alpha) dy + \int\limits_{\mathrm{supp}(f_{21})} y f_{21}(y) F_{11}(y-\alpha) dy \\
&= \int_0^\infty y e^{-y}(1 - e^{-(y+\alpha)}) dy + \int_\alpha^\infty y e^{-y}(1 - e^{-(y-\alpha)}) dy \\
&= 1 + 1/2 \cdot e^{-\alpha}(\alpha + 1).
\end{aligned}$$

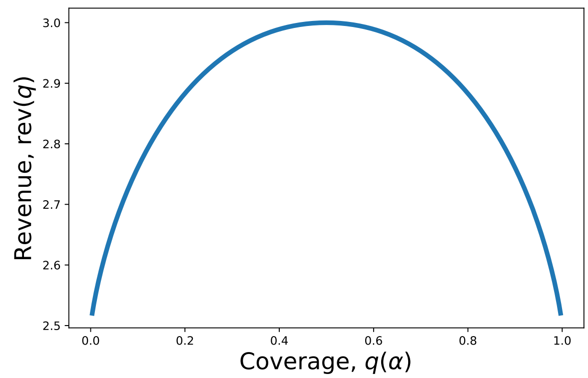
Differentiating $\mathrm{rev}_{\mathrm{shift}}$ we can observe it is not a concave function of the shift α (see Figure 9(a)). Indeed if we consider $\frac{d^2 \mathrm{rev}_{\mathrm{shift}}}{d\alpha^2} = 1/2 \cdot e^{-\alpha}(\alpha - 1)$, it is positive for all $\alpha > 1$. Consider the coverage $q(\alpha)$ of advertiser 1.

$$\begin{aligned}
q(\alpha) &= \int\limits_{\mathrm{supp}(f_{11})} y f_{11}(y) F_{21}(y+\alpha) dy \\
&= \int_0^\infty e^{-y}(1 - e^{-(y+\alpha)}) dy \\
&= 1 - 1/2 \cdot e^{-\alpha}.
\end{aligned}$$

Similarly we can observe that q is not a convex function of α (see Figure 8(a)). Using $q(\alpha)$ to formulate the loss $\mathcal{L}(\alpha)$ we can easily observe that it is non-convex as well (see Figure 8(b)). Let



(a) We report the total revenue as a function of the shift.



(b) We report the total revenue as a function of the coverage.

Figure 9: *Revenue as a function of coverage and shift.* Total revenue for two advertisers with exponentially distributed bids, on two user types. We vary the shift of one of the advertisers.

us re-parameterize the revenue, $\mathrm{rev}_{\mathrm{shift}}$ in terms of q, as $\mathrm{rev}(\cdot)$.

$$\mathrm{rev}(1-q) = 1 + (1-q)(1 - log(2-2q)))$$
$$\frac{d^2\mathrm{rev}(q)}{dq^2} = \frac{-1}{1-q} \leq 0. \qquad \text{(Using } q < 1\text{)}$$

We can observe that revenue is a concave function of the coverage (see Figure 9(b)).