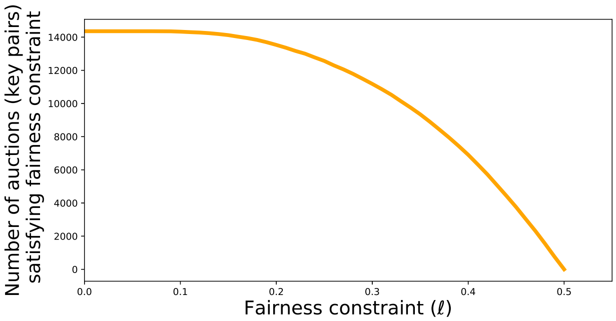


Figure 2. Implicit fairness of auctions. The x-axis depicts the fairness of the algorithm, measured by (ℓ, u)-fairness constraints. We report number of auctions which satisfy each fairness level.

A. Figures

Figure 2 represents the implicit fairness of the auctions derived from the real-world data sets. In particular we find that 3282 out of 14380 auctions have a selection lift $\mathrm{slift} < 0.3$.

Figure 3 shows the correlation between keywords. Where each group can be thought of as a category of keyword.

Figure 4 plots the loss $\mathcal{L}$ as a function of shifts α, we can observe that it is a non-convex function of the shift.

Figure 5 plots the coverage q as a function of shifts α, we can observe that it is a non-convex function of the shift.

Figure 6 shows that the reparameterization of the revenue is a concave function of the coverage q.

- Figure 6(a) plots revenue as a function of the shift, and shows its non-concavity.
- Figure 6(b) plots reparameterization of revenue as a function of the coverage, and illustrates that it is a concave function of the coverage.

B. Background and Notation

In this section, we provide some key definitions. For a detailed discussion, we refer the reader to the excellent treatises (Hartline, 2017; Nisan et al., 2007) on Mechanism design.

Definition 1. ***(Truthful mechanism).*** *Given the valuation $v_i \in \mathbb{R}$ of a bidder $i \in [n]$, and the bid $b_k \in \mathbb{R}$ of all other bidders $k \in [n]\backslash\{i\}$, a mechanism $\mathcal{M}$ is said to be truthful iff $v_i \in \mathrm{argmax}_{b_i \in \mathbb{R}} \left(x_i(b_1, \ldots, b_n)b_i - p_i(b_1, \ldots, b_n)\right)$.*

The above definition implies that for any truthful mechanism, an advertiser's optimal strategy is to bid their true valuation. Further, the can be shown that the allocation rule $x(b_1, b_2, \ldots, b_n)$, of any truthful mechanism must be monotone in b_i for all $i \in [n]$.

(Myerson, 1979) proved for any mechanism $\mathcal{M}$ there exists a truthful mechanism $\tau(\mathcal{M})$ such that $\tau(\mathcal{M})$ offers the same revenue to the seller and the same utility to each bidder as $\mathcal{M}$. As such, we restrict ourselves to truthful mechanisms. Furthermore, it is a well known fact (Nisan et al., 2007) that for any truthful mechanism its payment rule p, is uniquely defined by its allocation rule x. Hence, for any truthful mechanism our only concern is the allocation rule x.

B.1. Myerson's Optimal Mechanism

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* if $\phi(v)$ is non-decreasing in v. Likewise, we say $\mathcal{P}$ is *strictly regular* if $\phi(v)$ is strictly increasing in v.

Myerson's mechanism is defined as the VCG mechanism (Clarke, 1971; Groves, 1973; Vickrey, 1961) 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 in expectation over the bids. Since VCG is surplus maximizing, if Myerson's mechanism is truthful then it maximizes the revenue. It can be shown that if the bids have a regular distribution, then Myerson's mechanism is truthful, and therefore, revenue maximizing.

C. Why Is the TV-Distance Small?

To calculate the TV-distance we consider the distribution of winners selected by the auction mechanism, i.e., the distribution of the number of users an advertiser reaches.

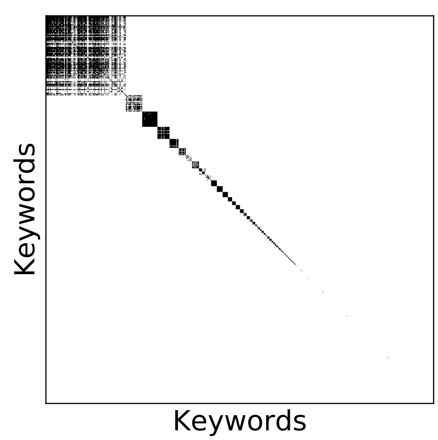


Figure 3. Correlation between keywords. The axes depict keywords, reordered to emphasize their correlation. A pair of keywords is colored white if the keywords share at least 2 advertisers. Each block can be interpreted as category of keyword (e.g., Science, Sports or Travel).

This distribution is different from coverage which separates the audience by their types. We report the total variation distance

$$d_{TV}(\mathcal{M},\mathcal{F}) \coloneqq 1/2 \sum_{i=1}^{n} |\sum_{j=1}^{m} q_{ij}(\mathcal{M}) - q_{ij}(\mathcal{F})| \in [0,1] \tag{28}$$

between the two distributions, as a measure of how much the winning distribution changes due to the fairness constraints.

Consider the distribution of advertiser i's coverage as the vector $\{q_{ij}(\mathcal{M})\}_{j\in[m]} \in [0,1]^m$. Its projection on the perfectly-fair polytope is

$$\frac{1}{m}\left(\sum_{j=1}^{m} q_{ij}(\mathcal{M})\right) \cdot (1,1,\cdots,1) \in [0,1]^n. \tag{29}$$

Since the coverage is uniform, it satisfies the perfect fairness constraints. Further, we can observe that this projection has a 0 total variation distance d_{TV} to $\{q_{ij}(\mathcal{M})\}_{j\in[m]}$ using Eq. 28.

If the solution $q_{ij}(\mathcal{F})$ of the optimal fair mechanism is close to this projection, then the resulting $d_{TV}(\mathcal{M},\mathcal{F})$ is small. Moving the coverage $q_{ij}(\mathcal{F})$ away from the projection involves a trade-off between increasing the total change in coverage, and decreasing the change for some types the advertiser values more.

Therefore, if the average bid of an advertiser does not vary significantly between the types, then $q_{ij}(\mathcal{F})$ is close to the projection. Importantly, this does not imply that the coverages $q_{ij}(\mathcal{M})$ of the unconstrained mechanism are balanced. To gain some intuition, consider two advertisers with similar budgets, but one advertiser places a bid of $1+\varepsilon$ for men and $1-\varepsilon$ for women, while the other places a bid of 1 for men and women. Even though the first advertiser's for men is only 2ε higher than their bid for women, they would be able to reach men, i.e., $q_1 = (1,0)$. Whereas, the platform only loses ε fraction of its revenue by changing q_1 to its projection $(1/2, 1/2)$.

D. Supplementary Proofs

D.1. Theorem 1

Proof of Theorem 1. 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

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

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 that for all $g \subseteq \mathrm{supp}(\phi_j),\ i \in [n]$, and $j \in [m]$

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

Since $x_{ij}(\cdot)$ and $\gamma_j(\cdot)$ are continuous, we can equivalently require for all $\phi_j,\ i \in [n]$, and $j \in [m]$

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

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: Find a optimal $\alpha_j \in \mathbb{R}^n_{\geq 0}$, $\beta_j \in \mathbb{R}^n_{\geq 0}$ and $\gamma_j(\cdot)\colon \mathrm{supp}(\phi_j) \to \mathbb{R}_{\geq 0}$

Dual of the infinite-dimensional fair advertising problem

$$\min_{\substack{\alpha_j\geq 0\\ \beta_j\geq 0\\ \gamma_j(\cdot)\geq 0}} \sum_{j\in[m]} \int_{\mathrm{supp}(\phi_j)} \gamma_j(\phi_j) df_j(\phi_j) \tag{30}$$

$$\text{s.t.}\quad \alpha_{ij} - \beta_{ij} + \Pr_{\mathcal{U}}[j]\phi_{ij} - \sum_{t\in[m]} (\alpha_{it}\ell_{it} - \beta_{it}u_{it}) \leq \gamma_j(\phi_j)\ \ \forall\, i \in [n], j \in [m], \phi_j. \tag{31}$$

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 $j \in [m]$

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

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., $\gamma'_j(\phi'_j) := \gamma_j(\phi_j)$, then the problem can be expressed as follows.

Dual with shifted virtual valuations

$$\min_{\gamma_j(\cdot)\geq 0} \sum_{j\in[m]} \int_{\mathrm{supp}(\phi_j)} \gamma_j(\phi'_j) df_j(\phi'_j) \tag{32}$$

$$\text{s.t.} \quad \Pr_{\mathcal{U}}[j]\phi'_{ij} \leq \gamma_j(\phi'_j) \quad \forall\, i \in [n], j \in [m], \phi'. \tag{33}$$

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 $\alpha' = 1/\Pr_{\mathcal{U}}[j] \cdot \left(\alpha^\star_{ij} - \beta^\star_{ij} + \sum_{t\in[m]}(\alpha^\star_{it}\ell_{it} - \beta^\star_{it}u_{it})\right)$ is a optimal fair mechanism.

Unconstrained Primal

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

$$\text{s.t.} \sum_{i\in[n]} x_{ij}(\phi_j) \leq 1 \quad \forall\, j \in [m], \phi_j \in \mathrm{supp}(\phi_j).$$

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. ∎

D.2. 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

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

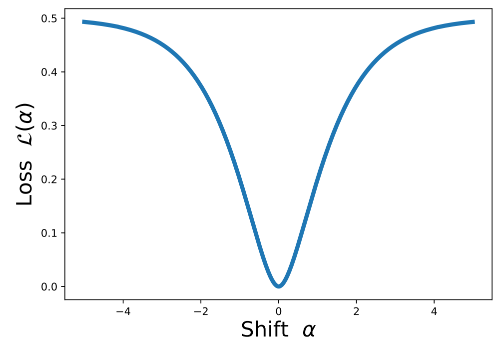


Figure 4. Loss as a function of shifts. (Non-convex) The loss $\mathcal{L}(\alpha)$, for two advertisers with exponential valuations, and $\delta = [0.5, 0.5]$.

Differentiating $\mathrm{rev}_{\mathrm{shift}}$ we can observe it is not a concave function of the shift α (see Figure 6(a)). Indeed if we consider $\frac{d^2\mathrm{rev}_{\mathrm{shift}}}{d\alpha^2} = 1/2\cdot(\alpha+1)e^{-\alpha}$, it is positive for all $\alpha > 1$. Consider the coverage $q(\alpha)$ of advertiser 1

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

Similarly we can observe that q is not a convex function of α (see Figure 5). Using $q(\alpha)$ to formulate the loss $\mathcal{L}(\alpha)$ we can easily observe that it is non-convex as well (see Figure 4). Let us re-parameterize the revenue $\mathrm{rev}_{\mathrm{shift}}$ in terms of q as $\mathrm{rev}(\cdot)$. Then we have

$$\mathrm{rev}(1-q) = 1 + (1-q)(1-\log(2-2q))) \tag{34}$$

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

We can observe that revenue is a concave function of the coverage (see Figure 6(b)).

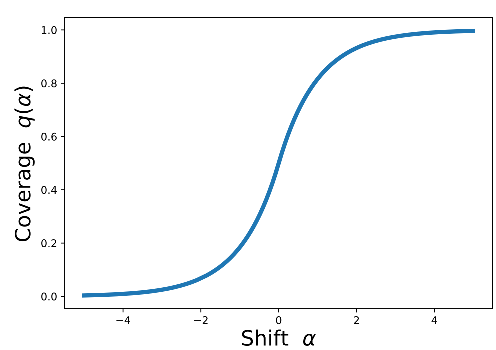
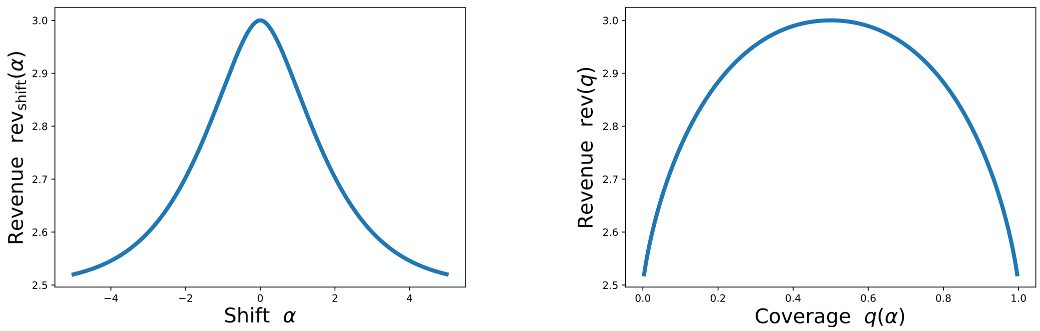


Figure 5. Coverage as a function of shift. (Non-convex) Coverage for one of the two advertisers with exponentially distributed bids, on two user types. We vary the shift of one of the advertisers from -5 to 5, and report its coverage as a function of the shift.



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

Figure 6. 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.

D.3. Omitted Details of Linear System in Equation (16)

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

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

To obtain $\nabla \text{rev}(q)$, we use the fact that $J_q(\alpha)$ is always invertible (Lemma 1). Then, if we know $\alpha = q^{-1}(\delta)$ for some $\delta \in [0,1]^{n \times m}$, we can express $\nabla \text{rev}(q)$ as follows,

$$\forall\, i \in [n],\ j \in [m],\ \frac{\partial \text{rev}_{\text{shift}}(\alpha)}{\partial \alpha_{ij}} = \sum_{k \in [n]} \frac{\partial \text{rev}(\alpha)}{\partial q_{kj}} \frac{\partial q_{kj}}{\alpha_{ij}}.$$

Equivalently, we can write the above as the following linear system

$$(J_q(\alpha))^\top \nabla \text{rev}(\delta) = \nabla \text{rev}_{\text{shift}}(\alpha). \quad \text{(Gradient oracle, 35)}$$

D.4. Proof of Lemma 1

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

$$\sum_{t \in [n]} \frac{\partial q_{ij}}{\partial \alpha_{tj}} = 0. \tag{36}$$

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

$$\frac{\partial q_{kj}}{\partial \alpha_{ij}} \leq 0 \text{ and } \frac{\partial q_{ij}}{\partial \alpha_{ij}} \geq 0. \tag{37}$$

Now, from Eq. (36) we have

$$\forall\, i \in [n], j \in [m],\ \frac{\partial q_{ij}}{\partial \alpha_{ij}} = \sum_{t \in [n] \setminus \{i\}} \left| \frac{\partial q_{ij}}{\partial \alpha_{tj}} \right|. \tag{38}$$

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,

$$\forall\, i \in [n], j \in [m],\ \frac{\partial q_{ij}}{\partial \alpha_{ij}} > \sum_{t \in [n-1] \setminus \{i\}} \left| \frac{\partial q_{ij}}{\partial \alpha_{tj}} \right|. \tag{39}$$

By observing the coverage on user type j is independent of the shift of user type t for all $t \neq j$, i.e.,

$$\forall\, i, s \in [n],\ j, t \in [m],\ s.t., j \neq t,\ \frac{\partial q_{ij}}{\partial \alpha_{st}} = 0, \tag{40}$$

and using Equation (38), we get that the Jacobian, $J_q(\alpha)$ is strictly diagonally dominant. Now, by the properties of strictly dominant matrices it is invertible. ■

Remark 8. *Since for any $i \in [n]$, q_{ij} is independent of α_{st} for any $s \in [n]$* (40)*. 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 Eq. (16) in $O(n^\omega m)$ steps, where ω is the fast matrix multiplication coefficient.*

D.5. Proof of Lemma 2

We use Lemma 4 and Lemma 5 in the proof of Lemma 2. The two lemmas split the Lipschitz continuity of $\mathrm{rev}(\cdot)$ into the Lipschitz continuity of $\mathrm{rev}_{\mathrm{shift}}(\cdot)$ and $\alpha_{ij} = q_{ij}^{-1}(\cdot)$ respectively. Their proofs are follow in Section D.7 and Section D.9 respectively.

Lemma 4. ***(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 $\rho \ \forall\ i \in [n],\ j \in [m]$, then $\mathrm{rev}_{\mathrm{shift}}(\alpha)$ is $(\mu_{\max}\rho n^{\frac{3}{2}})$-Lipschitz continuous.*

Lemma 5. ***(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 2. Let $\alpha,\ \beta \in \mathbb{R}^{(n-1)\times m}$ be the shifts achieving q_1 and q_2 respectively. Then by Lemma 4 and Lemma 5 we have,

$$|\mathrm{rev}(q(\alpha)) - \mathrm{rev}(q(\beta))| \overset{\text{Lemma 4}}{\leq} \mu_{\max}\rho n^{\frac{3}{2}}\|\alpha - \beta\|_F \tag{41}$$

$$\|\alpha - \beta\|_F \overset{\text{Lemma 5}}{<} \frac{\sqrt{n}}{\eta\mu_{\min}}\|q(\alpha) - q(\beta)\|_2. \tag{42}$$

By combining Eq. (41) and Eq. (42) we get the required result

$$|\mathrm{rev}(q_1) - \mathrm{rev}(q_2)| \overset{(41),(42)}{\leq} \frac{\mu_{\max}\rho}{\mu_{\min}\eta}n^2\|q_1 - q_2\|_2. \tag{43}$$

■

D.6. Proof of Lemma 3

To get an efficient complexity with a gradient-based algorithm we want to avoid small gradients "far" from the optimal. Lemma 6 shows that if $\mathcal{L}(\alpha)$ greater than ε, then the Frobenius norm $\|\mathcal{L}(\alpha)\|_F$ of $\mathcal{L}(\alpha)$ is greater than $\sqrt{\varepsilon}$. The proof of Lemma 6 is provided in Section D.10.

Lemma 6. ***(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, in Lemma 7 we show that the gradient, $\nabla\mathcal{L}(\alpha)$, is $O(n(L + n^2\mu_{\max}^2))$-Lipschitz continuous. Therefore, at each step where $\mathcal{L}(\alpha) \geq \xi$, we improve the loss by a factor of $1 - \beta\xi$, where β does not depend on ξ. This gives us a complexity bound of $O(\log 1/\varepsilon)$. The proof of Lemma 7 is presented in Section D.11.

Lemma 7. ***(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.*

Proof of Lemma 3. 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 Eq. (40) 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 2 we have that $\mathcal{L}_j$ is $O(n(L + n^2\mu_{\max}^2))$-Lipschitz continuous. Let $L' := O(n(L + n^2\mu_{\max}^2))$, 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 $\lambda := \frac{2}{n-1}\eta\mu_{\min}$, then from Lemma 6 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

$$\mathcal{L}(\alpha) = \sum_{j=1}^{m}\mathcal{L}_j(\alpha_j) < \varepsilon.$$

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

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

■

D.7. Proof of Lemma 4

Proof. We first consider the revenue for one user type j, $\mathrm{rev}_{\mathrm{shift},\ j}(\alpha)$, 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

$$\text{rev}_{\text{shift},\ j}(\alpha) := \sum_{i\in[n]} \Pr_{\mathcal{U}}[j] \int_{\text{supp}(f_{ij})} y f_{ij}(y) \prod_{k\in[n]\setminus\{i\}} F_{kj}(y+\alpha_{ij}-\alpha_{kj})dy. \tag{Revenue from type j, 44}$$

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

$$\text{rev}_{\text{shift}}(\alpha) = \sum_{j=1}^{m} \text{rev}_{\text{shift},\ j}(\alpha)$$

We can express $\nabla \text{rev}_{\text{shift},\ j}(\alpha)$ as shown in Figure D.7.

We can observe that every term in the gradient (Eq. (45), Eq. (46)) 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 (Eq. (45)) involves at most $2n$ terms from Eq. (47) for some $i, k, \ell \in [n]$ and $j \in [m]$,

$$\int_{\text{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y+\alpha_{ij}-\alpha_{kj}) \prod_{\ell\neq i,k} F_{\ell j}(y+\alpha_{ij}-\alpha_{\ell j})dy. \tag{47}$$

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

$$\begin{aligned}
&\left|\int_{\text{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y+\alpha_{ij}-\alpha_{kj}) \prod_{\ell\neq i,k} F_{\ell j}(y+\alpha_{ij}-\alpha_{\ell j})dy\right| \\
&\overset{(18)}{\leq} \mu_{\max} \left|\int_{\text{supp}(f_{ij})} y f_{ij}(y)dy\right| \qquad \text{(Using } F_{ij}(\cdot)\leq 1\text{)} \\
&\overset{(20)}{\leq} \mu_{\max}\rho. \qquad (48)
\end{aligned}$$

Consider,

$$\int_{\text{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y+\alpha_{ij}-\alpha_{kj}) \prod_{\ell\neq i,k} F_{\ell j}(y+\alpha_{ij}-\alpha_{\ell j})dy \tag{49}$$

$$\int_{\text{supp}(f_{kj})} y f_{kj}(y) f_{ij}(y+\alpha_{kj}-\alpha_{ij}) \prod_{\ell\neq i,k} F_{\ell j}(y+\alpha_{kj}-\alpha_{\ell j})dy \tag{50}$$

Then rewriting the gradient, from Figure 7, we have

$$\begin{aligned}
\left|\frac{\partial \text{rev}_{\text{shift},\ j}(\alpha)}{\partial \alpha_{ij}}\right| &= \Pr_{\mathcal{U}}[j] \sum_{k\in[n-1]\setminus\{i\}} ((49)-(50)) \\
&\overset{(48)}{\leq} \Pr_{\mathcal{U}}[j] \sum_{k\in[n-1]\setminus\{i\}} \mu_{\max}\rho \\
&\leq (n-2) \Pr_{\mathcal{U}}[j]\rho\mu_{\max}. \qquad (51)
\end{aligned}$$

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

$$\begin{aligned}
\|\nabla \text{rev}_{\text{shift},\ j}(\alpha)\|_F^2 &= \sum_{\substack{i\in[n-1]\\k\in[m]}} \left|\frac{\partial \text{rev}_{\text{shift},\ j}(\alpha)}{\partial \alpha_{ik}}\right|^2 \\
&\overset{(46)}{=} \sum_{i\in[n-1]} \left|\frac{\partial \text{rev}_{\text{shift},\ j}(\alpha)}{\partial \alpha_{ij}}\right|^2 \qquad (52) \\
&\overset{(51)}{\leq} \Pr_{\mathcal{U}}[j](n-1)((n-2)\rho\mu_{\max})^2 \qquad (53)
\end{aligned}$$

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

$$\begin{aligned}
\|\nabla \text{rev}_{\text{shift}}(\alpha)\|_F^2 &= \sum_{\substack{i\in[n-1]\\j\in[m]}} \left|\sum_{k\in[m]} \frac{\partial \text{rev}_{\text{shift},\ k}(\alpha)}{\partial \alpha_{ij}}\right|^2 \\
&\overset{(46)}{=} \sum_{\substack{i\in[n-1]\\j\in[m]}} \left|\frac{\partial \text{rev}_{\text{shift},\ j}(\alpha)}{\partial \alpha_{ij}}\right|^2 \\
&\overset{(52)}{=} \sum_{j\in[m]} \|\text{rev}_{\text{shift},\ j}(\alpha)\|_F^2 \\
&\overset{(53)}{\leq} \sum_{j\in[m]} \Pr_{\mathcal{U}}[j](n-1)((n-2)\rho\mu_{\max})^2 \\
&\leq (n-1)((n-2)\rho\mu_{\max})^2 \sum_{j\in[m]} \Pr_{\mathcal{U}}[j] \\
&\leq (n-1)((n-2)\rho\mu_{\max})^2. \qquad (54)
\end{aligned}$$

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

D.8. Lemma 8

Lemma 8. *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 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.,*

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

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

$$q_{ij}(\alpha) = \int_{\text{supp}(f_{ij})} f_{ij}(y) \prod_{k\in[n]\setminus\{i\}} F_{kj}(y+\alpha_{ij}-\alpha_{kj})dy \geq \eta. \tag{55}$$

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

$$\begin{aligned}
&\left|\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{sj}}\right| \\
&= \left| \int_{\text{supp}(f_{ij})} f_{ij}(y) f_{sj}(y+\alpha_{sj}-\alpha_{ij}) \prod_{k\neq i,s} F_{kj}(y+\alpha_{ij}-\alpha_{kj}) dy \right| \\
&\geq \mu_{\min} \left| \int_{\text{supp}(f_{ij})} f_{ij}(y) \prod_{k\neq i,s} F_{kj}(y+\alpha_{ij}-\alpha_{kj}) dy \right| \\
&\qquad\qquad (\text{Using } f_{ij}(\phi_{ij}) \geq \mu_{\min}) \\
&\geq \mu_{\min} \left| \int_{\text{supp}(f_{ij})} f_{ij}(y) \prod_{k\neq i} F_{kj}(y+\alpha_{ij}-\alpha_{kj}) dy \right| \\
&\qquad\qquad (\text{Using } F_{ij}(\phi_{ij}) \leq 1) \\
&\overset{(55)}{\geq} \eta\mu_{\min}. \qquad (56)
\end{aligned}$$

■

D.9. Proof of Lemma 5

In the Lemma 9 we extend the lower bound from Lemma 8 to the directional derivative of $q_{ij}(\alpha)$.

Lemma 9. (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 $i \in \text{argmax}_{k\in[n-1]} |u_k|$ and for all $t \in [0, t_{\max}]$*

$$\text{sign}(u_i)\frac{\partial q_{ij}(tu+\alpha_j))}{\partial t} > \frac{\eta\mu_{\min}}{\sqrt{n}}. \qquad (57)$$

Proof. Consider $i \in \text{argmax}_{k\in[n-1]} |u_k|$. 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 Eq. (10) we can express $q_{ij}(tu+\alpha)$ and its gradient as shown in Figure 8. Since $i \in \text{argmax}_{k\in[n-1]} |u_k|$,

$$\begin{aligned}
|u_i| &\geq |u_k| \\
\text{sign}(u_i)u_i &\geq \max(u_k, -u_k) \\
(u_i - u_k)\text{sign}(u_i) &> 0. \qquad (60)
\end{aligned}$$

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 (59) with $\text{sign}(u_i)$ and using Equation (60) and the fact that the integrals involved are positive to lower bound the equation we get

$$\begin{aligned}
\text{sign}(u_i)\frac{\partial q_{ij}(tu+\alpha_j)}{\partial t} &\overset{(60)}{\geq} \text{sign}(u_i)u_i \left.\frac{\partial q_{ij}}{\partial \alpha_{nj}}\right|_{\alpha_j+tu} \\
&\overset{\text{Lemma 8}}{\geq} |u_i|\eta\mu_{\min} \\
&\geq \frac{\eta\mu_{\min}}{\sqrt{n-1}} \\
&\qquad (\text{Using } |u_i| > 1/\sqrt{n-1}) \\
&> \frac{\eta\mu_{\min}}{\sqrt{n}}.
\end{aligned}$$

■

Proof of Lemma 5. Consider a 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 9 we have $\exists\, i \in [n-1]$, such that

$$\forall\, t \in [0,1], \ \left|\frac{\partial q_{ij}(tu+\beta_j)}{\partial t}\right| > \eta\mu_{\min} \qquad (61)$$

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

$$\begin{aligned}
\|q_j(\alpha_j) - q_j(\beta_j)\|_2^2 &= \sum_{i\in[n]} \left| \int_0^1 \frac{\partial q_{ij}(tu+\beta_j)}{\partial t} dt \right|^2 \\
&\geq \left| \int_0^1 \frac{\partial q_{ij}(tu+\beta_j)}{\partial t} dt \right|^2 \\
&\overset{(61)}{\geq} \frac{(\eta\mu_{\min})^2}{n} \left| \int_0^1 (tu+\beta_j) dt \right| \\
&\geq \frac{(\eta\mu_{\min})^2}{n} \|\alpha_j - \beta_j\|_2^2. \qquad (62)
\end{aligned}$$

Using Equation (62) for every type $j \in [m]$ we get that $\|q(\alpha) - q(\beta)\|_F > (\eta\mu_{\min})^2/n \cdot \|\alpha - \beta\|_F$. ■

D.10. Proof of Lemma 6

In the proof of Lemma 6 we use Lemma 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 10 does not follow from linear independence of $\nabla q_{ij}(\alpha)\ \forall\, i \in [n]$ (Lemma 1), because linear combinations of linearly independent vectors can be arbitrary small while having "large" weights. We present the proof of Lemma 10 in Section D.12.

Lemma 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,*

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

Proof of Lemma 6. Since $\mathcal{L}_j(\alpha_j) \geq \varepsilon$, we have

$$\mathcal{L}_j(\alpha_j) = \sum_{i\in[n-1]} (\delta_{ij} - q_{ij}(\alpha_j))^2 \geq \varepsilon. \quad (63)$$

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

$$\begin{aligned}\mathcal{L}_j(\alpha_j) &= \sum_{i\in[n-1]} (\delta_{ij} - q_{ij}(\alpha_j))^2 \\ &\leq \left(\sum_{i\in[n-1]} \left|\delta_{ij} - q_{ij}(\alpha_j)\right|\right)^2. \quad (64)\end{aligned}$$

From these we have that

$$\sum_{i\in[n-1]} |\delta_{ij} - q_{ij}(\alpha_j)| \overset{(64),(63)}{\geq} \sqrt{\varepsilon}. \quad (65)$$

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 10 we have

$$\left\| \sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha_j) \right\|_2 \overset{\text{Lemma 10}}{\geq} \frac{\eta\mu_{\min}}{n-1}$$

$$\left\| \sum_{i\in[n-1]} 2(\delta_{ij} - q_{ij}(\alpha_j)) \nabla q_{ij}(\alpha_j) \right\|_2 \geq 2\sqrt{\varepsilon}\frac{\eta\mu_{\min}}{n-1}.$$

■

D.11. Proof of Lemma 7

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 11), and then use Lemma 12 (Corollary 1.2 in (Varga, 2011)) to bound the magnitudes of the eigen-values.

Lemma 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.,*

$$\forall\, i \in [n], \qquad \frac{\partial^2 q_{ij}}{\partial\alpha_{ij}\partial\alpha_{ij}} \leq n(L + n\mu_{\max}^2)$$

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

Proof. Consider the Hessian of $q_{ij}(\alpha_j)$ in Figure 9, 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 some combinations of $i, k, \ell \in [n]$ and $j \in [m]$

$$\int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{kj}(y+\alpha_{ij}-\alpha_{kj}) f_{\ell j}(y+\alpha_{ij} - \alpha_{\ell j}) \prod_{h\neq \ell,k,i} F_{hj}(y+\alpha_{ij}-\alpha_{hj}) dy \quad (70)$$

$$\int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f'_{kj}(y+\alpha_{ij}-\alpha_{kj}) \prod_{\ell\neq k,i} F_{\ell j}(y+\alpha_{ij}-\alpha_{\ell j}) dy. \quad (71)$$

Each term along the diagonal of the Hessian (Eq. (66)), $\frac{\partial^2 q_{ij}}{\partial\alpha_{ij}\partial\alpha_{ij}}$, is a combination of $(n-1)$ terms of the form Eq. (70), and n^2 terms of the form Eq. (71). All other terms in the Hessian contain at most n terms of the form Eq. (71) and 1 term of the form Eq. (70). 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 Hessian. To this end, recall that for all $i \in [n],\ j \in [m]$, and $y \in \mathrm{supp}(f_{ij})$

$$0 < f_{ij}(y) \leq \mu_{\max}, \quad (72)$$

$$0 \leq F_{ij}(y) \leq 1, \quad (73)$$

$$|f'_{ij}(y)| < L, \quad (74)$$

$$\int_{\mathrm{supp}(f_{ij})} f_{ij}(z) dz = 1. \quad (75)$$

We can now bound Equation (70) and Equation (71) as follows

$$(70) \overset{(72),(73)}{\leq} \mu_{\max}^2 \left| \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) dy \right| \overset{(75)}{\leq} \mu_{\max}^2, \quad (76)$$

$$(71) \overset{(74),(73)}{\leq} L \left| \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) dy \right| \overset{(75)}{\leq} L. \quad (77)$$

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

$$\left|\frac{\partial^2 q_{ij}}{\partial\alpha_{ij}\partial\alpha_{ij}}\right| = \left|\sum_{k\neq i}(71) + \sum_{k\neq i}\sum_{\ell\neq i,k}(70)\right| \overset{(76),(77)}{\leq} (n-1)\big(L+(n-2)\mu_{\max}^2\big) \quad \text{(Using triangle inequality)}$$
$$\leq n(L+n\mu_{\max}^2) \tag{78}$$
$$\left|\frac{\partial^2 q_{ij}}{\partial\alpha_{kj}\partial\alpha_{kj}}\right| = \big|(71)\big| \overset{(77)}{\leq} L \tag{79}$$
$$\left|\frac{\partial^2 q_{ij}}{\partial\alpha_{kj}\partial\alpha_{tj}}\right| = \big|(70)\big| \overset{(76)}{\leq} \mu_{\max}^2 \tag{80}$$
$$\left|\frac{\partial^2 q_{ij}}{\partial\alpha_{kj}\partial\alpha_{ij}}\right| = \left|\frac{\partial^2 q_{ij}}{\partial\alpha_{ij}\partial\alpha_{kj}}\right| = \left|(71) + \sum_{\ell\neq k,i}(70)\right| \overset{(76),(77)}{\leq} L+(n-2)\mu_{\max}^2 \tag{81}$$

■

Lemma 12. *(Corollary 1.2 in (Varga, 2011)) 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 (Varga, 2011) for a proof of the above lemma.

Proof of Lemma 7. To show that $\nabla\mathcal{L}_j(\alpha_j)$ is Lipschitz continuous, we show that $q_{ij}(\alpha_j)$ is Lipschitz continuous, then use the fact that $\nabla q_{ij}(\alpha_j)$ is Lipschitz continuous from Lemma 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 $j \in [m]$.

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

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

$$\left|\frac{\partial q_{ij}}{\partial\alpha_{ij}}\right| = \Big|\sum_{k\in[n]\setminus\{i\}} t(k)\Big| \leq (n-1)\cdot\big|t(i)\big| \overset{(72),(73)}{\leq} (n-1)\left|\mu_{\max}\int_{\mathrm{supp}(f_{ij})} f_{ij}(y)dy\right| \overset{(75)}{\leq} (n-1)\mu_{\max} \tag{82}$$

$$\left|\frac{\partial q_{ij}}{\partial\alpha_{kj}}\right| = \big|t(k)\big| \overset{(72),(73)}{\leq} \mu_{\max}\left|\int_{\mathrm{supp}(f_{ij})} f_{ij}(y)dy\right| \overset{(75)}{\leq} \mu_{\max} \tag{83}$$

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 $\|\nabla q_{ij}(\alpha_j)\|$

$$\|\nabla q_{ij}(\alpha_j)\|_2^2 = \sum_{k\in[n]}\left(\left|\frac{\partial q_{ij}}{\partial\alpha_{kj}}\right|^2\right) \leq \left|\frac{\partial q_{ij}}{\partial\alpha_{ij}}\right|^2 + \sum_{k\in[n]\setminus\{i\}}\left(\left|\frac{\partial q_{ij}}{\partial\alpha_{kj}}\right|^2\right) \overset{(82),(83)}{\leq} (n-1)^2\mu_{\max}^2 + n\mu_{\max}^2 \leq n^2\mu_{\max}^2$$
$$\|\nabla q_{ij}(\alpha_j)\|_2 \leq n\mu_{\max} \tag{84}$$

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

$$\sum_{i=1}^{n-1}|\delta_{ij} - q_{ij}(\alpha_j)| \leq \sum_{i=1}^{n-1}|\delta_{ij}| + \sum_{i=1}^{n-1}|q_{ij}(\alpha_j)| \leq 2 \tag{85}$$

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

$$\nabla^2\mathcal{L}_j(\alpha_j) = 2\cdot\sum_{i=1}^{n-1}\nabla q_{ij}(\alpha_j)\nabla q_{ij}(\alpha_j)^\top - \sum_{i=1}^{n-1}(\delta_{ij} - q_{ij}(\alpha_j))\cdot H(\alpha_j) \tag{86}$$

We know from Lemma 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,

$$\|H(\alpha_j)\|_\star \overset{\text{Lemma 12}}{\leq} 2n(L + n\mu_{\max}^2) \tag{87}$$

$$\|q_{ij}(\alpha_j)q_{ij}(\alpha_j)^\top\|_\star \leq \|q_{ij}(\alpha_j)\|_2^2 \overset{(84)}{\leq} n^2\mu_{\max}^2. \tag{88}$$

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

$$\|A + B\|_\star \leq \|A\|_\star + \|B\|_\star \quad (\text{Sub-additivity of } \|\cdot\|_\star, 89)$$

This gives us the following

$$\begin{aligned}
&\|\nabla^2\mathcal{L}_j(\alpha_j)\|_\star \\
&\overset{(88)}{\leq} 2\sum_{i\in[n-1]} \|\nabla q_{ij}(\alpha_j)\nabla q_{ij}(\alpha_j)^\top\|_\star + (\delta_{ij} - q_{ij}(\alpha_j))\|H(\alpha_j)\|_\star \\
&\overset{(88)}{\leq} 2\sum_{i\in[n-1]} \|\nabla q_{ij}(\alpha_j)\nabla q_{ij}(\alpha_j)^\top\|_\star + (\delta_{ij} - q_{ij}(\alpha_j))\|H(\alpha_j)\|_\star \\
&\overset{(87),(88)}{\leq} 2\sum_{i\in[n-1]} n^2\mu_{\max}^2 + 4n(L + n\mu_{\max}^2)\sum_{i\in[n-1]}(\delta_{ij} - q_{ij}(\alpha_j)) \\
&\overset{(85)}{\leq} 2n^3\mu_{\max}^2 + 4n(L + n\mu_{\max}^2)
\end{aligned}$$

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

D.12. Proof of Lemma 10

Proof of Lemma 10. Without loss of generality consider a reordering of $(x_1, x_2, \dots, x_n)$, s.t., for some $p \leq n-1$,

$$x_i \geq 0 \;\forall\, i \leq p \tag{90}$$

$$x_i < 0 \;\forall\, i > p \tag{91}$$

Case A: $\sum_{i\in[p]} x_i < 1/2$**:**

We can replace x by $-x$, since this does not change the norm $\left\|\sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha_j)\right\|_2$. 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. $(\alpha_{1j}, \alpha_{2j}, \dots, \alpha_{nj})$. Therefore we have for all $i \in [n-1]$

$$\begin{aligned}
\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{ij}} + \sum_{k\in[n-1]\setminus\{i\}} \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} &\overset{(36)}{=} -\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{nj}} \\
&\overset{(37)}{=} \left|\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{nj}}\right| \\
&\overset{\text{Lemma 8}}{\geq} \eta\mu_{\max}.
\end{aligned} \tag{92}$$

Calculating the weighted sum of Equation (93) over $i \in [p]$ with weights x_i we get

$$\begin{aligned}
\sum_{i\in[p]} x_i\left(\sum_{k\in[n-1]} \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}}\right) &\overset{(93)}{>} \sum_{i\in[p]} x_i\eta\mu_{\min} \\
&> \frac{\eta\mu_{\min}}{2}.
\end{aligned}$$

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

$$\sum_{i\in[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \geq \frac{1}{n-1}\sum_{k\in[n-1]}\left(\sum_{i\in[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}}\right) \tag{93}$$

$$\geq \frac{\eta\mu_{\min}}{2(n-1)}. \tag{94}$$

From Equation (37) for all $i \in [p]$ and $k > p$, $\frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} < 0$. Therefore, $k \leq p$ in Equation (94). From this we get

$$\begin{aligned}
\sum_{i\in[n-1]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} &= \sum_{i\in[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} + \sum_{i\in[n-1]\setminus[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \\
&\overset{(94)}{\geq} \frac{\eta\mu_{\min}}{2(n-1)} + \sum_{i\in[n-1]\setminus[p]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}} \\
&\overset{\text{Lemma 8}}{\geq} \frac{\eta\mu_{\min}}{2(n-1)} + \eta\mu_{\min}\sum_{i\in[n-1]\setminus[p]}(-x_i) \\
&\overset{(91)}{\geq} \frac{\eta\mu_{\min}}{2(n-1)}.
\end{aligned} \tag{95}$$

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

$$\begin{aligned}\left\|\sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha)\right\|_2^2 &= \sum_{t\in[m]}\sum_{k\in[n-1]}\left(\sum_{i\in[n-1]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kt}}\right)^2 \\ &\overset{(40)}{=} \sum_{k\in[n-1]}\left(\sum_{i\in[n-1]} x_i \frac{\partial q_{ij}(\alpha)}{\partial \alpha_{kj}}\right)^2 \\ &\overset{(95)}{\geq} \left(\frac{\eta\mu_{\min}}{2(n-1)}\right)^2 \qquad (96) \\ \left\|\sum_{i\in[n-1]} x_i \nabla q_{ij}(\alpha)\right\|_2 &\geq \frac{\eta\mu_{\min}}{2(n-1)}\end{aligned}$$

■

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

$$\frac{\partial \mathrm{rev}_{\mathrm{shift},\ j}(\alpha)}{\partial \alpha_{ij}} = \Pr_{\mathcal{U}}[j] \sum_{k \neq i} \int_{\mathrm{supp}(f_{ij})} y f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy - \Pr_{\mathcal{U}}[j] \sum_{k \neq i} \int_{\mathrm{supp}(f_{kj})} y f_{kj}(y) f_{ij}(y + \alpha_{kj} - \alpha_{ij}) \prod_{\ell \neq i,k} F_{\ell j}(y + \alpha_{kj} - \alpha_{\ell j}) dy \tag{45}$$

$$\frac{\partial \mathrm{rev}_{\mathrm{shift},\ j}(\alpha)}{\partial \alpha_{ik}} = 0 \tag{46}$$

Figure 7. Gradient of $\mathrm{rev}_{\mathrm{shift},j}(\cdot)$. Equations from the proof of Lemma 4.

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

$$q_{ij}(tu + \alpha) = \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) F_{nj}(y + tu_i + \alpha_{ij}) \prod_{k \neq i,n} F_{kj}(y + t(u_i - u_k) + \alpha_{ij} - \alpha_{kj}) dy \tag{58}$$

$$\frac{\partial q_{ij}(tu + \alpha)}{\partial t} = u_i \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{nj}(y + tu_i + \alpha_{ij}) \prod_{k \neq i,n} F_{kj}(y + t(u_i - u_k) + \alpha_{ij} - \alpha_{kj}) dy + (u_i - u_k) \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) F_{nj}(y + tu_i + \alpha_{ij}) \sum_{k \neq i,n} f_{kj}(y + t(u_i - u_k) + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq i,k,n} F_{\ell j}(y + t(u_i - u_\ell) + \alpha_{ij} - \alpha_{\ell j}) dy \tag{59}$$

Figure 8. Directional derivative of $q_{ij}(\cdot)$. Equations from the proof of Lemma 9.

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

$$\frac{\partial^2 q_{ij}}{\partial \alpha_{ij} \partial \alpha_{ij}} = \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) \sum_{k \neq i} f'_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy + \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) \sum_{k \neq i} f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \sum_{\ell \neq i,k} f_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) \prod_{h \neq \ell,k,i} F_{hj}(y + \alpha_{ij} - \alpha_{hj}) dy \tag{66}$$

$$\frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{kj}} = \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f'_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy \tag{67}$$

$$\frac{\partial^2 q_{ij}}{\partial \alpha_{kj} \partial \alpha_{ij}} = \frac{\partial^2 q_{ij}}{\partial \alpha_{ij} \partial \alpha_{kj}} = -\int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f'_{kj}(y + \alpha_{ij} - \alpha_{kj}) \prod_{\ell \neq k,i} F_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) dy - \int_{\mathrm{supp}(f_{ij})} f_{ij}(y) f_{kj}(y + \alpha_{ij} - \alpha_{kj}) \sum_{\ell \neq k,i} f_{\ell j}(y + \alpha_{ij} - \alpha_{\ell j}) \prod_{h \neq \ell,k,i} F_{hj}(y + \alpha_{ij} - \alpha_{hj}) dy \tag{68}$$

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

Figure 9. Hessian of $q_{ij}(\cdot)$. Equations from proof of Lemma 11.