Optimising Trade-offs Among Stakeholders in Ad Auctions

YORAM BACHRACH, Microsoft Research
SOFIA CEPPI, Microsoft Research
IAN A. KASH, Microsoft Research
PETER KEY, Microsoft Research
DAVID KUROKAWA, Carnegie Mellon University

We examine trade-offs among stakeholders in ad auctions. Our metrics are the revenue for the utility of the auctioneer, the number of clicks for the utility of the users and the welfare for the utility of the advertisers. We show how to optimize linear combinations of the stakeholder utilities, showing that these can be tackled through a GSP auction with a per-click reserve price. We then examine constrained optimization of stakeholder utilities.

We use simulations and analysis of real-world sponsored search auction data to demonstrate the feasible trade-offs, examining the effect of changing the allowed number of ads on the utilities of the stakeholders. We investigate both short term effects, when the players do not have the time to modify their behavior, and long term equilibrium conditions.

Finally, we examine a combinatorially richer constrained optimization problem, where there are several possible allowed configurations (templates) of ad formats. This model captures richer ad formats, which allow using the available screen real estate in various ways. We show that two natural generalizations of the GSP auction rules to this domain are poorly behaved, resulting in not having a symmetric Nash equilibrium or having one with poor welfare. We also provide positive results for restricted cases.

1. INTRODUCTION

Studies of auctions tend to focus on one particular objective, with the goal of optimizing it. For example, the VCG auction maximizes welfare, while Myerson [1981] studied the design of auctions that maximize revenue in a single parameter setting. However, for designers of practical auctions, such pure objectives are rarely the goal. Instead, the auction designer, even if caring only about long-term revenue, must take into account various aspects of the health of the marketplace.

A notable example is that of search advertising auctions, where the opportunity to advertise alongside search results is auctioned off. There, the search engine wants to maximize revenue, but faces competition from both other search engines and other advertising outlets. As a result, maximizing revenue requires striking a balance between extracting revenue from current advertisers and keeping them or attracting more. A classic result of Bulow and Klemperer [1996] shows that, in some cases, attracting even a single additional bidder can be as valuable as perfectly optimizing revenue.

Given this, the auctioneer should take at least some consideration of the welfare of advertisers.

At the same time, a search engine is a two-sided platform where the good being sold to advertisers is the attention of searchers. Decisions that maximize revenue could result in a poor user experience if more or worse ads are shown. As searchers have other options, the supply of goods to be sold depends on how satisfied they are. Thus, the auctioneer should take their welfare into account as well.

While auction theory provides a rich set of tools for optimizing the welfare of a single group, much less attention has been paid to optimizing tradeoffs among multiple stakeholders. Better understanding this, in the particular context of sponsored search, is the goal of this paper.

1.1. Our Contribution

We begin by analyzing the optimization of a linear combination of objectives relevant to the search engine (revenue), advertisers (welfare), and users (clicks). We provide an auction design that optimizes any objective that is *linear* in the click probability of an advertiser, of which each of these three objectives (and any linear combination of them) is a special case. Like current practice, this auction uses a rank score to order ads and prices can be computed in a truthful or Generalized Second Price (GSP) fashion. Importantly, this auction achieves trade-offs between revenue and other objectives through the use of a per-click reserve.

Next, we approach the same problem through the lens of constrained optimization. Rather than optimize a particular combination, we could constrain some of the objectives to lie above some minimal value while optimizing the remaining ones. This may be a more natural approach in practice, since it is not clear how the various objectives should be weighted, but given a current state of the marketplace it is natural to ask how to make the system one percent better for users while holding revenue neutral. Perhaps unsurprisingly, we are able to show that this problem is equivalent to the unconstrained problem via Lagrangean duality (the analysis is not trivial because the object we seek is an optimal *function*).

Of course, these three objectives are not the only natural ones for an auction designer to consider. For example, using the number of clicks (click yield) treats a deceptive advertisement that tricks people into clicking on it as beneficially relevant for users. An alternative that fits within our constrained optimization framework is to constrain the number of ads shown (impression yield), so that only the "best" are shown to users. We characterize the design of the optimal option with this constraint. Interestingly, the optimal way to ensure this constraint is satisfied through a per-impression reserve, which previous work has identified as a poor tool from a pure revenue maximization standpoint [Roberts et al. 2013; Thompson and Leyton-Brown 2013].

To complement this theoretical exploration, we follow [Roberts et al. 2013] and examine the relative performance of the optimal mechanism to various designs that have previously been considered through extensive experiments on both synthetic and real data. The experiments on synthetic data confirm the theoretical results: a ranking algorithm with a per-impression reserve usually outperforms the other evaluated ranking algorithms in terms of revenue, welfare, and click yield. For the second part of the experiments, we use historical data from Microsoft Bing for a keyword with over 500 bidders and for a keyword with fewer than 10 bidders. In the former case, all the ranking algorithms perform similarly for high values of impression yield while for low values of impression yield the algorithm with a per-impression reserve provides a higher revenue. For the latter case, the algorithm with the per-impression reserve provides an essentially constant revenue while controlling the number of ads shown. This is higher for low numbers of ads and lower for high numbers of ads. Finally, on a

sample from a weeks worth of data across all keywords on Bing, we observe opposite results when trading off between click and impression versus revenue and impression.

Finally, we consider the trade-offs created due to the existence of several different types of ads. For example, instead of displaying several text ads the search engine may at times wish to display one large image-based ad or a block of ads for different products that can be purchased. This setting diverges from classical theory in two ways. First, the set of advertisers in the two camps are not necessarily equivalent (e.g. an advertiser may only desire a placement for image-based ads). Second, the search engine's available slots are no longer pre-set and can change based off the bids. We will then show that the truthful auction in this setting is largely analogous to the standard setting, but attempting to utilize GSP payment rules results in significant complications. Our results here are quite negative, with GSP being poorly behaved except in restricted cases. As both Google and Bing use GSP, identifying a solution to this problem is an important direction for future work.

1.2. Related Work

Our work sits in a long line of papers that study generalized second price auctions for sponsored search [Varian 2007; Edelman et al. 2007]. Ostrovsky and Schwarz [2011] studied the effects of applying Myerson's [1981] optimal per-click reserve on Yahoo! (For practical reasons they actually implemented per-impression reserve.) Lahaie and Pennock [2007] proposed the idea of a squashing parameter to improve revenue, and Lahaie and McAfee [2011] showed that it can increase welfare as well. Thompson and Leyton-Brown [2013] studied a variety of ways of increasing revenue, including through a reserve of the optimal form.

In the study of trade-offs, Likhodedov and Sandholm [2003] showed how to optimize a combination of revenue and welfare. Diakonikolas et al. [2012] explore the trade-off revenue/welfare focusing the attention on computational complexity issues for both exact and approximated deterministic mechanisms. The convex combination of revenue and welfare has also been considered in a study about refinements in the prediction of the relevance of ads done by Sundararajan and Talgam-Cohen [2013]. Liu and Chen [2006] and Liu et al. [2010] compared the designs of revenue-optimal and welfare-optimal auctions in a simple setting. Edelman and Schwarz [2010] argued that setting an optimal reserve price will lead to a significant revenue gain with minimal welfare loss. Athey and Ellison [2011] studied a model of consumer search and showed that reserve prices can increase user welfare. Roberts et al. [2013] studied the revenue optimal auction and showed that empirically it led to good trade-offs between revenue and other objectives, but did not have a theoretical explanation for this. They also showed that symmetric Nash equilibra continue to exist when per-impression reserves, per-click reserves of the optimal form, and other linear alterations to the ranking are performed.

Our motivations for considering trade-offs are driven by issues of endogenous participation. Jehiel and Lamy [2013] examine a general model of this problem in auctions, and their application of their results to sponsored search settings yields a qualitatively similar result to ours: reserves prices should be set, but not at as high a level as in a setting where participation is exogenous.

2. PRELIMINARIES

We now describe the setting and notation we use throughout the paper. Our setting is a standard sponsored search ad auction or a position auction, where there are n bidders or advertisers; bidder i submits a bid b_i for their ad to be displayed, and a displayed ad that is clicked is charged a price p_i.

We use the following notation:

— $b = (b_1, b_2, ..., b_n)$: the vector of bids.
— $t_i \in [c_i, d_i]$: the type of bidder i (their true value of a click), which can take a value in the compact set $[c_i, d_i]$.
— T: the set of valid bidder types/bids vectors (i.e. $b, (t_1, t_2, ..., t_n) \in A$).
— Types are independently distributed with density function f_i for bidder i, giving density $f(t) = \prod_i f_i(t_i)$.
— $\phi_i : \mathbb{R}_{\geq 0} \to \mathbb{R}$, the associated virtual value function of the bidder i, assumed to be regular. I.e. $\phi_i(t_i) = t_i - (1 - F(t_i))/f(t_i)$ is non-decreasing.
— w_i: the ad effect of bidder i.
— $x_i : T \to \mathbb{R}^n$: the allocation rule chosen. This represents the expected slot effect given to bidder i.
— $t_i w_i x_i$ the expected welfare of advertiser i, determined as his value for a click multiplied by his probability of a click (factored into an ad effect and a slot effect).
— $p_i : T \to \mathbb{R}^n$: the payment function of bidder i.

As we will be looking at linear objectives and constraints, we introduce:

— $OBJ = \alpha$ revenue$+\beta$ welfare$+\gamma$ click yield: the objective we will frequently maximize in our auctions. We assume $\alpha, \beta, \gamma \geq 0$ and $\alpha + \beta > 0$ (the case of $\alpha = \beta = 0$ is largely trivial and due to its cumbersome edge-case nature will not be considered).
— $\psi_i(z) = \alpha \phi_i(z) + \beta z + \gamma$: here the α, β, γ are equivalent to those used in OBJ.

To define an auction mechanism the designer must specify both an allocation rule (or ranking algorithm) and a payment function. We use ranking algorithms based on the parameters described above that are *monotone*, i.e., the expected valuation of each advertiser does not decrease when their type increases, and payments that provide the "right" incentives to advertisers to report bids that correspond to their types. We can obtain such payments in two ways: a) since the ranking algorithm is monotone, we can compute the payments as prescribed by Myerson [1981], yielding a truthful mechanism; b) We can use GSP payments. In the latter case, we need a solution concept that characterizes a *single* expected outcome, and apply the commonly used one: *symmetric* or *locally envy free* Nash equilibrium (SNE) [Edelman et al. 2007; Varian 2007]. An SNE is a refinement of a Nash equilibrium requiring that for every pair of advertisers i and j, the following constraint is satisfied:

$$x_j(b) \left(t_i - \frac{w_j}{w_i} p_j(b) \right) \leq x_i(b)(t_i - p_i(b))$$

Intuitively, this says that i prefers their slot and payment to j's slot and payment. Previous work has shown that with GSP payments the "lowest" SNE corresponds to the truthful outcome, in the sense that each bidder ends up with the same slot and payment, for a broad class of allocation rules [Aggarwal et al. 2006; Edelman et al. 2007; Varian 2007; Roberts et al. 2013].

3. UNCONSTRAINED OPTIMIZATION

We first begin with auctions that solve an (essentially) unconstrained optimization problem[1]. The natural problem of this form is to assign advertisers to slots to maximize some combination of objectives such as revenue, welfare, and click yield. A key observation, previously used in a different setting by Likhodedov and Sandholm [2003], is that all of these objectives are linear in the allocation. Thus they can naturally be

---

[1]Strictly speaking there is a constraint that the allocation rule be monotone, but for the cases we examine the optimal solution has this property so we can ignore this constraint and perform unconstrained optimization.

folded together to yield an optimal auction by using a "rank score" that takes the appropriate combination of them. The following theorem captures this for our objective function of the form $OBJ = \alpha$ revenue $+ \beta$ welfare $+ \gamma$ click yield (where $\alpha, \beta, \gamma \geq 0$). This allows trading off measures of the utility of the search platform, the advertisers, and the users. Of course, any other linear term could easily be included.

THEOREM 3.1. *The auction that maximizes OBJ maximizes the equivalent objective*

$$\int_T \sum_{i=1}^n w_i \psi_i(t_i) x_i(t) f(t) dt.$$

Thus, the truthful auction that maximizes for any instance of types/bids t, the expression $\sum_{i=1}^n w_i \psi_i(t_i) x_i(t)$ is optimal.

PROOF. We first note that α, β, γ are all non-negative and that each of the function components are monotone, resulting in monotonicity.

Let us now consider each of the three terms separately.

(1) revenue: following Myerson [1981], we have that

$$\int_T \sum_{i=1}^n w_i p_i(t) f(t) dt = -\sum_{i=1}^n w_i U_i(x, p, c_i) + \int_T \sum_{i=1}^n w_i \phi_i(t_i) x_i(t) f(t) dt$$

where U_i is the expected utility of bidder i when bidding c_i in the auction described by x and p as in [Myerson 1981].

(2) welfare:

$$\int_T \sum_{i=1}^n w_i t_i x_i(t) f(t) dt$$

(3) click yield:

$$\int_T \sum_{i=1}^n w_i x_i(t) f(t) dt$$

Combining the above, we get that:

$$OBJ = -\alpha \sum_{i=1}^n w_i U_i(x, p, c_i) + \int_T \sum_{i=1}^n w_i \left(\alpha \phi_i(t_i) + \beta t_i + \gamma \right) x_i(t) f(t) dt.$$

Again following [Myerson 1981], the first term can be set to 0 by setting the payment p accordingly, and as feasible direct revelation mechanisms must have that $U_i(p, x, c_i) \geq 0$, we have that this is optimal. Therefore, to optimize OBJ, we need only optimize the second term. $\square$

Note that similarly to the revenue-only case, Theorem 3.1 can clearly correspond to a ranking based on $w_i \psi_i(t_i)$. In the case that the distributions across types are identical, the ϕ_i are equivalent to some ϕ, and in turn the ψ_i are equivalent to some ψ. A single bidder-independent reserve price r can then be implemented where $r = \psi^{-1}(0)$.

3.1. Implications for GSP

The analysis in Theorem 3.1 assumes that advertising is sold using a truthful auction. However, the major sponsored search platforms, such as Google and Bing's ad platforms, implement auctions that are *not* truthful. Instead of a truthful payment rule, these platforms use the GSP payment rule (that is, a bidder pays the value of the minimum bid required to retain the slot allocated to them). For many purposes

this distinction is not important, as there exists an equilibrium of the GSP auction that implements the optimal truthful outcome, in the sense that the allocations and payments for all advertisers are the same [Edelman et al. 2007; Varian 2007; Roberts et al. 2013]. However, this analysis has been shown to work for rankings which are *linear* in bids (see in particular [Roberts et al. 2013]). This is important because, even in the regular case, the virtual valuation functions ϕ_i need not be linear, and it is not clear that this good behavior extends to this setting. Nevertheless, Theorem 3.1 can at the very least be applied to optimizing linear approximation of the ϕ_i. Thompson and Leyton-Brown [2013] observed that ϕ_i is in fact linear for uniform value distributions. In fact, many Beta distributions are also close to linear for much of their range.

4. CONSTRAINED OPTIMIZATION

Section 3 allowed linear trade-offs between the utility measures of the different interested parties, but did not allow an objective function that requires a minimal threshold utility for them. We now consider objective functions that allow setting such a minimal threshold. One example is designing a revenue optimal auction under the constraint that the social welfare exceeds a minimal threshold value of θ. More generally, we allow several constraints of the form α_k revenue $+ \beta_k$ welfare $+ \gamma_k$ click yield $\geq \theta_k$.

Note that the equivalent objective function derived in Theorem 3.1 is a linear functional on $x = (x_1, x_2, ..., x_n)$, which we can write in succinct notation as

$$G(x) = \int_T \sum_{i=1}^n w_i \psi_i(t_i) x_i(t) f(t) dt.$$

By analogy, it follows from Theorem 3.1 and its proof that each constraint can be expressed in the form $a_k(x) \geq \theta_k$ where a_k is a linear functional and hence we can represent a set of constraints in the succinct form $A(x) + Z = \theta$, where $\theta = (\theta_1, \theta_2, ..., \theta_r)$, $Z = \{z \in /\mathbb{R}^{\rho(A)} : z \leq 0\}$ is the non-positive cone in the case of inequality constraints and $Z = 0$ for equality constraints. Here A is a vector of linear functionals.

Our optimizations over x are in function space, which we take to be the L^1 Banach space w.r.t. Lebesgue measure. We write $\mathcal{X}$ for the feasible region for the unconstrained problem (which requires x to be positive amongst other things). Hence the constrained problem can be written as

$\mathbf{P_Z}(\theta):$

Maximize $\qquad\qquad OBJ = \alpha$ revenue $+ \beta$ welfare $+ \gamma$ click yield

subject to $\qquad\qquad A(x) + z = \theta$

$\qquad\qquad\qquad x \in \mathcal{X}$

over $x \in L^1, z \in Z$.

We now introduce some technical conditions, used to facilitate proofs: we assume that the allocation function $x : T \to \mathbb{R}^n$ is a Lebesgue integrable function $\in L^1$, and we require the constraint feasibility region $\mathcal{X}$ for the allocation to be convex with respect to the underlying function space. Note that "natural" requirements for sponsored search auctions translate to convex regions if we allow for randomized allocations.

The Lagrangian for the constrained problem is

$$L(x, \lambda) = G(x) - \lambda^T (\theta - A(x) - z)$$

and we demonstrate that the constrained problem is strong Lagrangian, meaning strong duality applies, and hence

$$\inf_\lambda \sup_{x \in X} L(x, \lambda) = L(x^*, \lambda^*) = \sup_{x \in \mathcal{X}} L(x, \lambda^*) = \sup_{x : x \in \mathcal{X} \& A(x) + z = \theta} G(x).$$

The proof largely mirrors the standard proof for convex optimization over Euclidean spaces, but applied to function space. There are two main parts of the proof: first, we show that if θ is in the relative interior of the constrained optimization then the problem is strong Lagrangian. That is, defining $\Gamma(\theta) = \sup_{x:x\in\mathcal{X}\,\&\,A(x)+z=\theta} G(x)$, we show the following.

THEOREM 4.1. *If $\theta \in \mathrm{relint}\,(\mathrm{dom}(\Gamma))$, then the optimization problem is strong Lagrangian.*

The second part is to give the equivalent version of Slater's constraint qualification — which gives sufficient conditions for the problem to be strong Lagrangian.

COROLLARY 4.2. *When $Z = \{z \in /\mathbb{R}^{\rho(A)} :\,\le 0\}$ is the non-positive cone, the problem $P_Z(\theta)$, is strong Lagrangian if there is some $\tilde{x} \in \mathcal{X}$ such that $A(\tilde{x}) > \theta$. If $Z = 0$, the problem $P_Z(\theta) = P(\theta)$ is strong Lagrangian provided there is some $\tilde{x} \in \mathrm{relint}\,(\mathcal{X})$ such that $A(\tilde{x}) = \theta$.*

Appendix A goes through the proofs in detail.

The next theorem is then a direct consequence of the problem being strong Lagrangian, and shows that auctions which satisfy linear constraints are essentially those which have the aforementioned linear objectives. We give the theorem for inequalities (the theorem for equalities is similarly proved).

THEOREM 4.3. *Suppose x^* maximizes α_0 revenue$+\beta_0$ welfare$+\gamma_0$ click yield subject to the following constraints:*

$$\alpha_1 \text{ revenue} + \beta_1 \text{ welfare} + \gamma_1 \text{ click yield} \ge \theta_1$$
$$\alpha_2 \text{ revenue} + \beta_2 \text{ welfare} + \gamma_2 \text{ click yield} \ge \theta_2$$
$$\vdots$$
$$\alpha_r \text{ revenue} + \beta_r \text{ welfare} + \gamma_r \text{ click yield} \ge \theta_r.$$

Then there exists some α^*, β^*, and γ^* such that x^* maximizes α^* revenue$+\beta^*$ welfare$+\gamma^*$ click yield.

PROOF. Since a solution x^* exists, the feasible region is non-empty. If a feasible solution exists with all the inequalities strict, then we know the problem is strong Lagrangian, and hence there are Lagrange multipliers $\lambda^* = [\lambda_1^*, ..., \lambda_r^*]$ such that the Lagrangian $L(x^*, \lambda^*)$ solves the problem — and the latter can be written as OBJ with $\alpha^* = \alpha_0 + \sum_i \alpha_i \lambda_i^*$, $\beta^* = \beta_0 + \sum_i \beta_i \lambda_i^*$, $\gamma^* = \gamma_0 + \sum_i \gamma_i \lambda_i^*$. Note that complementary slackness ensures that $\lambda.(\theta - A(x^*)) = 0$. If no feasible solution exists except on the boundary of the constraint set, then essentially there is one solution, whose values determine $\alpha^*, \beta^*, \gamma^*$. □

Each of the stakeholders would like to dominate the objective function (by setting their term in OBJ to 1 and the other terms to zero). We call an auction approximately optimal with level δ for one of the stakeholders if it achieves at least a δ fraction of the utility the optimal auction achieves for that stakeholder alone. One consequence of the theorem above is that the Pareto optimal surface is concave, so that there always exists a deterministic auction that achieves any convex combination of optimality levels for all the stakeholders.

THEOREM 4.4. *The Pareto surface is concave. That is, the surface of points (revenue, welfare, click yield) that are feasible by some direct revelation auction such that there is*

no other auction that simultaneously achieves weakly better revenue, welfare and click yield, and also achieves strictly better of one of the tree, is concave.

PROOF. Fix some constraint θ. Then we know that for every $OBJ(\alpha, \beta, \gamma)$ there is a non-vertical supporting hyperplane (the Lagrangian) at the point θ to the hypograph $\Gamma(\theta)$ which bounds the (constrained) objective function from above. Taking the point-wise infimum of these affine hyperplanes as α, β, and γ vary results in a concave surface. □

COROLLARY 4.5. *Let $\alpha, \beta, \gamma \geq 0$ satisfy $\alpha + \beta + \gamma = 1$. Then there exists a deterministic auction that achieves at least an α fraction of the max revenue, a β fraction of the max welfare, and a γ fraction of the max click yield.*

Another class of constrained auctions of interest are those where we limit the number of ads shown. That is, given J search queries over various search terms we may wish to show no more than k ads in expectation. This is another natural way to capture the utility of searchers, who presumably would generally prefer to see fewer, better-targeted ads. We now investigate such ad-limited auctions, making the simplifying assumption that all J auctions are single item auctions. In contrast to previous work which has suggested that a per-impression reserve is a poor tool for increasing revenue or trading-off between revenue and other objectives [Roberts et al. 2013; Thompson and Leyton-Brown 2013], this analysis shows that the use of such a reserve is the *optimal* way of controlling the number of ads shown.

THEOREM 4.6. *Suppose we have J single slot search terms where search term j appears with probability q_j and has a slot effect of s_j for its lone slot. Moreover, suppose we wish to show no more than $\theta \leq 1$ ads per search (in expectation) and when auctioning off search term j, we wish to maximize α_j revenue $+ \beta_j$ welfare $+ \gamma_j$ click yield. Then the optimal solution introduces a per-impression reserve price in all J auctions. Specifically, for each search term j we will separately wish to solve the following:*

$$\max_{x_j} \int_{T_j} \left(\sum_{i=1}^{n_j} (w_{j,i} \psi_{j,i}(t_{j,i}) - \lambda/s_j) \, x_{j,i}(t_j) \right) f_j(t_j) dt_j$$

where quantities indexed by j refer to the jth search term's auction (e.g. $\psi_{j,i}$ is the ψ function for search term j's bidder i), λ is some common value to all search terms, and $\alpha_j, \beta_j, \gamma_j \geq 0$ with $\alpha_j + \beta_j + \gamma_j > 0$ for each j.

PROOF. We first note that the expression is monotone, as the sum of monotone components. Now via a similar analysis to the theorem 3.1 it is not difficult to see that our problem is equivalent to:

$$\max_{x_1, x_2, \dots, x_J} \sum_{j=1}^{J} q_j \int_{T_j} \sum_{i=1}^{n_j} w_{j,i} \psi_{j,i}(t_{j,i}) x_{j,i}(t_j) f_j(t_j) dt_j$$

$$s.t. \sum_{j=1}^{J} q_j \int_{T_j} \sum_{i=1}^{n_j} \frac{x_{j,i}(t_j)}{s_j} f_j(t_j) dt_j = \theta.$$

Using a Lagrange multiplier λ gives a dual problem for fixed λ, $L(\lambda, x^*)$

$$\max_{x_1, x_2, \dots, x_J} \lambda \theta + \sum_{j=1}^{J} q_j \int_{T_j} \left(\sum_{i=1}^{n_j} (w_{j,i} \psi_{j,i}(t_{j,i}) - \lambda/s_j) \, x_{j,i}(t_j) \right) f_j(t_j) dt_j.$$

For a fixed value of λ not only does the $\lambda\theta$ in the above maximization become irrelevant, but also the terms in the sum indexed by j are independent of each other. Thus, the maximum of the sum can be separated into the sum of the maximums. The resulting solution (for fixed λ) is feasible; taking the minimum over λ of $L(\lambda, x^*)$ also produces a solution which is feasible, and hence by Lagrangian sufficiency $L(\lambda^*, x^*)$ solves the constrained problem. (c.f. Lemma A.2). □

5. EXPERIMENTS

Roberts et al. [2013] empirically considered a number of ways of making trade-offs between revenue and clicks / welfare and found that a form of per-click reserve they proposed (independently proposed by Thompson and Leyton-Brown [2013]) lead to the best results. In Theorem 3.1, we derived the optimal way of making such trade-offs, by optimizing a weighted combination of virtual valuations, clicks, and welfare. The proposed form can be thought of as a linear approximation to this, providing a theoretical explanation for this observation. Another observation from this prior work is that, despite their popularity, per-impression reserve prices tended to enable poor tradeoffs. However, Theorem 4.6 suggests they are the right tool for controlling the number of ads shown, as they correspond to the shadow price of a constraint on the number of ads. In this section, we revisit the experiments from [Roberts et al. 2013] to validate this empirically.

Our experiments look at four metrics: the number of ads shown (impression yield), the revenue, the (advertiser) welfare, and the number of clicks generated (click yield). In the short term, the auctioneer cares mostly about the revenue metric. Welfare reflects the total value created, and is a focus of the auctioneer in the long term, as it determines the ability of the platform to attract participants (who may have alternative platforms available). Click yield can be thought as a proxy for the value created for the users who are clicking on (presumably) useful ads. Impression yield can be thought of as another proxy for the value created for users, who presumably do not wish to see useless ads.

We consider several ranking algorithms in our analysis: the method Roberts et al. [2013] and Thompson and Leyton-Brown [2013] used to approximate the revenue optimal auction (denoted by $(b_i - r)w_i$, where r is a per-click reserve), the standard ranking algorithm coupled with a per-click reserve r (denoted by $b_i w_i / r$) and with a per-impression reserve ρ (denoted by $b_i w_i / \rho$), and a two parameter ranking algorithm that combines the first and the third (denoted by $(b_i - r)w_i / \rho$). Since this ranking algorithm is characterized by two parameters, the operating points form a region represented in the following figures by a shaded area.

We begin with a simple example characterized by eight advertisers bidding for three slots. Advertisers have i.i.d. types (t_i, w_i) where t_i and w_i are independent and uniformly distributed on $[0, 1]$. Figure 1 shows how each of the metrics changes with the allowed number of ads shown, using different auction designs.

The figure illustrates Theorem 4.6: given an expected number of ads shown, the optimal solution is provided by a ranking algorithm with a reserve score. In particular, the welfare and the click yield are always maximized by the standard ranking algorithm with a reserve score. The revenue is always maximized by the two-parameter algorithm.

We also evaluate the ranking algorithms in a more realistic setting considered by Lahaie and Pennock [2007]), which they selected by fitting Yahoo! data from a particular query. In this setting, bidder valuations are correlated with relevance, and have a lognormal distribution (which does not yield regular virtual valuations). The results for this setting are shown in Figure 2, which indicates that the observations made for the simpler setting hold in this case as well.
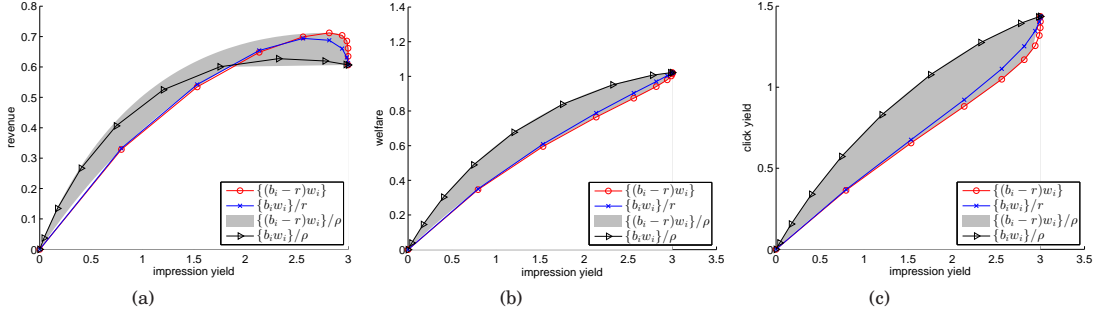
Fig. 1. Feasible impression yield – revenue (a), impression yield – welfare (c), and impression yield – click yield (c) operating points in a simple setting.
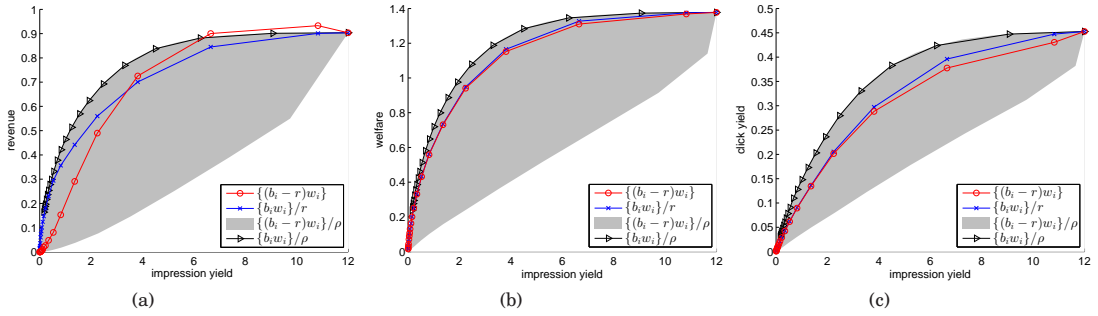

Fig. 2. Feasible impression yield – revenue (a), impression yield – welfare (c), and impression yield – click yield (c) operating points in Lahaie and Pennock's setting.

The above results are based on the assumption that bidders are in equilibrium. In reality, if parameters are changed, advertisers may take some time to reach equilibrium. Therefore, we want to investigate what happens in the short term, when the ranking method is changed but advertisers do not react.

As we are not interested in the equilibrium in this case, we can simply examine the performance of different ranking algorithms on historical data. This dataset has many realistic features which were not captured by the previously analysis, such as changing bidders, matching of bids to multiple queries, and stochastic quality scores.

The data we used for our simulations (used also in [Roberts et al. 2013]) is historical data from Microsoft Bing for a keyword with over 500 bidders, which we selected as representative of a "thick" market (Figure 3), and for a keyword with fewer than 10 bidders, as representative of a "thin" market (Figure 4). The data was normalized, but the exact values are not relevant for our purpose. Both figures show the effect of the number of ads displayed on the revenue when changing from the standard ranking algorithm with a reserve price to the ranking algorithm proposed by Roberts et al. [2013] and the standard ranking algorithm with a reserve score.

In the thick market, no ranking algorithm always outperforms all the others. However, though the three algorithms show a similar trend, the standard algorithm with reserve score provides a higher revenue when only few ads are displayed. In the thin market, the standard ranking algorithm with reserve score exhibits a different trend than the other algorithms (which are similar to each other in their behavior). In particular, the revenue provided by the former algorithm remains constant for almost all the number of ads shown. The non-smoothed trend all the ranking algorithms show

is explained by the advertisers' bids. In the dataset we use, some bids are more frequent than others (see [Roberts et al. 2013]) and this creates thresholds for the reserve price/score: when the reserve increases and moves from below a threshold to above it, the impression yield suddenly reduces. Interestingly, in both markets the standard ranking algorithm with reserve score performs better than the others for the same range of ads shown.
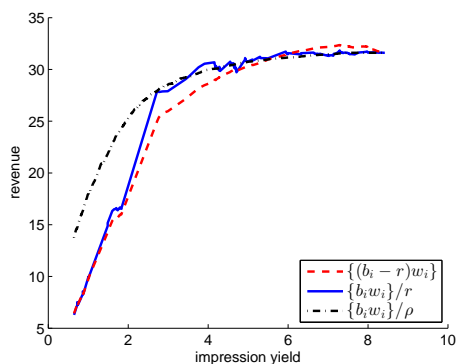

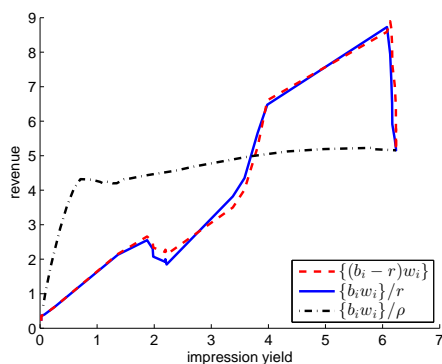Fig. 3. Revenue comparison in thick market auction replays.

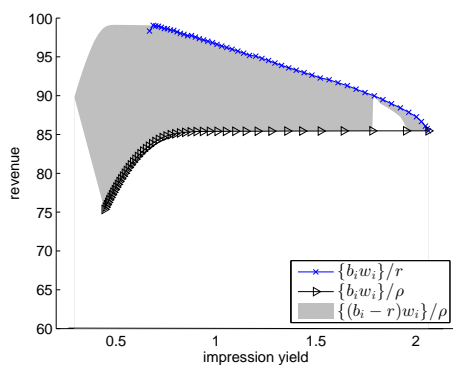Fig. 4. Revenue comparison in thin market auction replays.

Fig. 5. Revenue trade-offs on real data

Fig. 6. Click trade-offs on real data

Finally, on a sample from a week's worth of data across all keywords on Bing, Figures 5 and 6 show exactly opposite results when trading off between clicks and impressions versus revenue and impressions. For clicks, consistent with the theory, the optimal parameter to use is a per-impression reserve ρ. For revenue, as raising the per-click reserve always increased revenue for most of the range we explored, using that as the sole parameter is optimal for most of the range. Only once increasing it further begins to reduce revenue does it become optimal to add in a per-impression reserve. Again, this is consistent with the theory, despite this data not reflecting equilibrium behaviour.

6. GSP AUCTIONS WITH CLASSES AND TEMPLATES

We now turn to examine a more elaborate constrained optimization problem. Our motivation comes from changes in the world of sponsored search auctions, where recently several platforms have allowed enriching ads with features such as larger formats, reviews, maps, or phone numbers. A richer ad augmented with images or video is more likely to be clicked, but uses up more screen space. As the same space on the screen can be used in different ways or formats, the search engine faces the problem of determining which of many possible arrangements of advertising, or *templates* to use. Deng et al. [2010] make an initial step to solve this problem by proposing and studying the multi-slots pricing scenario.

As a guiding example for our work, consider the setting studied by Goel and Khani [2014], where a search result page can either display one image ad or k text ads. To analyse trade-offs in this type of setting, we require the following two extensions to the classical framework.

— There may be several *classes* of ads/bidders that are disparate in nature: the bidders of a class bid only for ads of their own class. Importantly, we assume throughout this section that the sets of bidders of two classes are *disjoint*. In our example, one class consists of the advertisers interested only in text ads, while the other class consists of those interested only in image-rich ads.
— There may be several mutually-exclusive *templates*, each of which describe an entire slot layout. For example, one template allows a single image-rich ad (and no text ads) and another template allows k text ads (but no image-rich ads).

With these extensions to the classical framework we will require the following updates to our notation and terminology.

— $x_{i,j} : A \to \mathbb{R}^n$: the allocation function of bidder i on template j. Intuitively, this represents the slot effect given to bidder i if the search engine decides to use template j. We therefore have the following constraint. Suppose template j has slot effects $s_{j,1}, s_{j,2}, ..., s_{j,n}$ (we can assume there is a slot for every player by setting $s_{j,i} = 0$ as needed). For every $\ell \in \{1, 2, ..., n\}$, let $\tilde{x}_{i,j,\ell} : A \to [0,1]$ be functions such that for every b, we have $\sum_i \tilde{x}_{i,j,\ell}(b), \sum_\ell \tilde{x}_{i,j,\ell}(b) \leq 1$. Intuitively $\tilde{x}_{i,j,\ell}$ represents the probability that bidder i receives slot ℓ (in template j). Then the $x_{i,j}$ are valid allocation functions only if there exists some $\tilde{x}_{i,j,\ell}$ such that for every i, j, and b, we have $x_{i,j}(b) = \sum_\ell s_\ell \tilde{x}_{i,j,\ell}(b)$. If they satisfy this constraint and do not assign slots from one class to a bidder of another, then the $x_{i,j}$ are valid.
— $x_i : A \to \mathbb{R}^n$: the actual allocation rule chosen. For every set of bids b, there must exist some template j such that for all i we have $x_i = x_{i,j}$. Note that the order of quantifiers allows the chosen template to depend on the bids.

Our first observation is that, for a truthful auction, this complexity does not introduce any significant issues. We are able to select the optimal template for our objectives, and as the domain remains a single-parameter one we are able to charge appropriate payments to induce truthful bidding.

THEOREM 6.1. *A truthful auction that maximizes OBJ is described by the allocation rule $x_{i,j}$:*

$$\max_j \max_x \sum_{i=1}^n w_i \psi(b_i) x_{i,j}(b).$$

PROOF. This allocation rule selects the template that has the highest value of OBJ when allocated optimally (correctly optimizing revenue because we are still in a single

parameter domain), so it only remains to check monotonicity to verify truthfulness. Suppose for purposes of contradiction it is not. Then there exists a bidder i such that when he raises his bid from b_i to b_i' the slot-effect assigned to him is decreased in value. Explicitly, we have that if s_j is the allocated slot effect for j when i reports a bid of b_i, and s_j' when b_i' is reported, then $s_i' < s_i$. Now as the s_j are chosen optimally for b, we have that:

$$\sum_{j=1}^{n} w_j \psi(b_j) s_j \geq \sum_{j=1}^{n} w_j \psi(b_j) s_j'$$

$$\Rightarrow \psi(b_i) \geq \frac{\sum_{j=1, j \neq i} w_j \psi(b_j)(s_j' - s_j)}{w_i(s_i - s_i')}.$$

Similarly, as the s_j' are chosen optimally for (b_i', b_{-i}), we have that:

$$w_i \psi(b_i') s_i' + \sum_{j=1, j \neq i}^{n} w_j \psi(b_j) s_j' \geq w_i \psi(b_i') s_i + \sum_{j=1, j \neq i}^{n} w_j \psi(b_j) s_j$$

$$\Rightarrow \psi(b_i') \leq \frac{\sum_{j=1, j \neq i} w_j \psi(b_j)(s_j' - s_j)}{w_i(s_i - s_i')}.$$

Combining these two results, we find that $\psi(b_i) \geq \psi(b_i')$, and as ψ is an increasing function, this is a contradiction. $\square$

6.1. Extensions of GSP

We have shown that our earlier results generalize for *truthful* auctions. Unfortunately, as previously discussed, search engines such as Bing and Google use GSP instead. Previously, this was not a significant issue (except possibly for exact revenue maximization) due to the equivalence results of Varian [2007], Edelman et al. [2007], and more generally Roberts et al. [2013]. However, due to the combinatorial nature of the image/text auction, and templates more broadly, the distinctions between truthful and non-truthful auctions become critical. Indeed, it is not even clear what the "right" way is to generalize GSP to our setting, and as we shall see, simple candidates have problems even when solely optimizing for advertiser welfare.

An underlying principle in standard GSP auctions is that "bidders pay the minimum required to keep their slot". We show that several natural extensions of GSP that maintain this principle are provably *not equivalent* to the truthful mechanism and have poor equilibrium properties.

Before we begin, an extension of the notion of the SNE is in order.

Definition 6.2. A *symmetric Nash equilibrium* is a set of bids b that satisfy the conditions:

— for all i and b_i' we have: $x_i(b_i', b_{-i})(t_i - p_i(b_i', b_{-i})) \leq x_i(b)(t_i - p_i(b))$
— for all i and j where i and j are bidding on the same class of ads, we have: $x_j(b)\left(t_i - \frac{w_j}{w_i} p_j(b)\right) \leq x_i(b)(t_i - p_i(b))$

The first of the two conditions is what constitutes a Nash equilibrium and the second defines the symmetric/envy-free nature of the equilibrium. Note that the second does not encompass the first, unlike in the classical single-template scenario.

6.1.1. Template-Considerate GSP. We begin with perhaps the most straightforward extension to multi-template settings.

Definition 6.3. The *template-considerate* GSP payment rule is such that the payment is the value of the minimum bid to ensure the bidder keeps his current slot. That is, the payment is the maximum of the minimum bid needed to be the next highest bidder's bid in the same class, and the minimum bid required to ensure the winning template remains optimal.

Unfortunately, this payment rule suffers from a few critical issues. First, we show that even in the very restricted image/text setting, an SNE may fail to select the same outcome as the truthful mechanism. Thus, the auction fails to optimize the desired objective.

THEOREM 6.4. *Suppose we use the template-considerate GSP payment rule. Then we may not implement the truthful outcome.*

PROOF. Suppose we are maximizing welfare and we have the following setup with all the $w_i = 1$.

— Bidders:
  — Four text ad bidders with true values $100, 50, 25$, and 10.
  — Two image ad bidders with true values 120 and 110.
— Templates:
  — Template 1: three text ads with slot effects $1, 1 - \epsilon$, and $1 - 2\epsilon$.
  — Template 2: a single image ad with slot effect 1.

In this given setup one problematic SNE is when the two image ad bidders bid their true value, while the text ad bidders bid zero. On the other hand, the truthful outcome clearly selects the first template. □

Note that this example relies on losing bidders whose class is not shown bidding 0. If instead we add the requirement that such bidders bid their true value, it is possible to construct examples where SNE do not exist.

Second, In more general settings, an SNE may not even exist (even without such an added requirement). Essentially, the issue is that the constraints on bids imposed by the requirement for a lack of envy on a given template may force the bids to be inconsistent with that template being chosen. This example does rely on bidders being *conservative* [Paes Leme and Tardos 2009] and not bidding above their true value (which is a dominated strategy).

THEOREM 6.5. *Suppose we use the template-considerate GSP payment rule and bidders do not bid higher than their true value. Then there may not exist any SNE.*

PROOF. Suppose we are maximizing welfare and we have the following setup where $\epsilon \approx 0$ with all the $w_i = 1$.

— Classes A and B (that are entirely symmetrical).
— Bidders:
  — Four class A bidders with true values $350, 300, 200$, and 100.
  — Four class B bidders with true values $350, 300, 200$, and 100.
— Templates:
  — Template 1: three class A slots of effects $1, 1 - \epsilon$, and $1 - 2\epsilon$, and three class B slots of effects ϵ, ϵ^2, and ϵ^3
  — Template 2: three class B slots of effects $1, 1 - \epsilon$, and $1 - 2\epsilon$, and three class A slots of effects ϵ, ϵ^2, and ϵ^3

Suppose that the first template is the winning template in an SNE. We claim the four class A bidders can bid at most (approximately) $350, 200, 200$, and 100 respectively.

All but the second bidder are bidding their true value so only the second bidder requires explanation. They cannot bid more than (approximately) 200 as otherwise the first bidder will envy the second bidder's slot for its price. This is due to the near identical value of the slots. Thus, the objective value of the first template is at most (approximately) $350 + 200 + 200 = 750$. Alternatively, we claim that the four class B bidders must bid at least (approximately) $300, 300, 200$, and 100 respectively. The first bidder must bid at least the second bidder's bid and the last bidder must bid their true value as they will not receive a slot. The second bidder must bid approximately their true value as the first slot is vastly superior to the second and so they must bid near their value to ensure that they are not envious of the first bidder. Similarly, due to the vast superiority of the second slot to the third the third bidder must bid approximately their true value. Thus, the objective value of the second template is at least (approximately) $300 + 300 + 200 = 800$. As this is larger than the first template's worth this is a contradiction. The second template cannot be the winning template in an SNE via similar analysis. □

While our example uses ties in valuations and slot effects for ease of exposition, the values above can be slightly perturbed so that they are not tied and the analysis is not fundamentally affected. Such issues are therefore not zero probability events.

Third, even if an SNE exists, the objective value may be arbitrarily worse in comparison to the truthful auction.

THEOREM 6.6. *Suppose bidders do not bid higher than their true values. Then there exists an SNE of an auction governed by the template-considerate GSP payment rule that may be arbitrarily worse than the worst NE of the corresponding truthful auction. That is, if OBJ_{GSP} is the worst OBJ value in an SNE for the template-considerate GSP auction, and $OBJ_{truthful}$ is the worst OBJ value in a NE for the corresponding truthful auction, then $OBJ_{GSP}/OBJ_{truthful}$ can be arbitrarily close to zero (with a sufficiently large number of bidders and slots).*

We defer the proof to Appendix B.

Aside from these crucial existence and optimality issues, we note that the payment rule can be counter-intuitive to the bidders. For example, it is possible that by increasing one's bid, a bidder does not affect his own slot allocation, but decreases the payment required of their competitors (who are being priced by the minimum bid they need to make to maintain the template).

6.1.2. Template-Indifferent GSP. With the failure of the template-considerate GSP rule, an alternate extension we can consider is to simply ignore the existence of other templates for pricing purposes.

Definition 6.7. The *template-indifferent* GSP payment rule is one where after the template is selected, GSP is performed as if the winning template were the only one.

An immediately obvious issue with this payment rule is that the highest bidder of every class has incentive to raise his bid arbitrarily high. This can be rectified by assuming for the purposes of template selection the highest bidder's bid is equivalent to the second highest bid, however this does undermine the optimality of the allocation. Even with this fix, the three problems identified with template-considerate GSP remain. As essentially the same examples work, we omit them. We have also explored other extensions to GSP, but have been unable to find one with desirable equilibrium properties in the general case.

6.2. Positive Results For Restricted Cases

The lack of a natural generalization of GSP with desirable properties is disheartening. Fortunately, with restrictions on the set of templates we can get somewhat more favourable results. The most trivial restriction is to a single template, where the classical results apply. Only slightly less trivial is the case where one class has the same set of slots in every template. Then, at least for this class, we are effectively again in the single-template setting.

More interestingly, we can return to our guiding example of the image/text setting explored by Goel and Khani [2014]. In this setting, VCG is known to have the undesirable property that adding bidders can reduce revenue. They seek to rectify this by designing a truthful *revenue monotone* mechanism which they call MITA, consisting of a monotone allocation rule and the payments to make it truthful. Interestingly, while MITA does not optimize OBJ, we are able to show that the MITA allocation with template-considerate GSP payments does have an SNE that implements the truthful MITA outcome.

Definition 6.8. Let $\psi_i(t_i)$ be the objective value of the text ad with the ith highest value (using a linear approximation to ϕ if it is non-linear), $\psi_I(t_I)$ be the value of the best image ad, and $s_{I,1}$ and $s_{T,1} \ldots s_{T,k}$ be the slot effects. Define $C = \{1 \leq j \leq k \mid \psi_j(t_j) \sum_{i=1}^{j} s_{T,i} \geq \psi_I(t_I) s_{I,1}\}$. The MITA allocation is the first $\max C$ text ads if C is nonempty and the image ad otherwise[2].

THEOREM 6.9. *The MITA allocation with template-considerate GSP payments has an SNE that implements the truthful outcome.*

PROOF. First, suppose the image ad is winning in the truthful outcome. We construct the SNE by having all ads bid their true value. This clearly implements the truthful MITA outcome because the payment rule and bids are exactly the same as in truthful MITA. No text ad wishes to deviate, because their payment for a bid high enough to be shown is the same as it would be in the truthful auction.

Now consider the case where the text ads win in the truthful outcome. Again, the image ad bidders bid their true value. Let j^* be the number of ads shown in the truthful MITA outcome. The text ads make the bids they would make in the lowest SNE of the auction where the text ad template is the only template but there is a per-impression reserve of $\psi_I(t_I) s_{I,1} / \left(\sum_{i=1}^{j^*} s_{T,i} \right)$ (such an SNE exists by a result from [Roberts et al. 2013]). The truthful outcome of this single-template auction is the same as the truthful MITA auction, so since the SNE implements the former it implements the latter. □

While this provides a setting where positive results are possible, we note that it is somewhat special as one class has only a single slot. While limited generalisations appear possible when each class has multiple slots, they show a greatly reduced number of ads (essentially only those that would qualify to be shown if they were the first slot).

Another special case where we can provide positive results is one where each class corresponds to a block of ads, e.g. there could be a block of text ads somewhere on the page, a block of image ads, and a block of product ads. The assumption we make about these blocks is similar in spirit to the assumption in the traditional GSP model that the probability of an ad being clicked can be factored into a slot effect and an advertiser effect. We assume that the slot effect can be further factored into a term that depends on the ordering within the block and a term that depends on where the

---

[2]Goel and Khani give their allocation rule for the special case where slots are identical, advertisers have identical qualities, and the objective is to maximize welfare. We have generalized their allocation rule to our setting.

block is shown. Equivalently, we can phrase this in terms of the ratio of slot effects in different templates, as in the following definition.

Definition 6.10. A *class selection* set of templates is one where whenever a class appears in multiple templates, the ratio of slot effects which are not zero are constant. More formally, if a class appears in a template with slot effects $s_1, s_2, ..., s_a$, and it appears in another template with slot effects $s'_1, s'_2, ..., s'_a$ then s_i/s'_i is constant.

Consider any monotone allocation rule that ranks bidders using a rank score and chooses the template based solely off the second highest ranked bidder of each class (a *second-highest* allocation rule). Such rules could be relatively efficient in some settings, but in general cannot provide any bounded approximation to OBJ because the value of the first bidder is ignored and can be arbitrarily high. Nevertheless, we show that this allocation rule has an SNE with GSP payments.

THEOREM 6.11. *With a class selection template set with a second-highest allocation rule and template-indifferent GSP payment rule, an SNE is guaranteed to exist.*

PROOF. For every class take a template in which that class appears and then determine an SNE for that class in that template. Furthermore, alter the first bidder's bid to be equivalent (or slightly higher) than the second bidder's bid. Note that (for this class) this is not only still an SNE in the template, but by the nature of class selection templates the bids are then in SNE for all templates. Now upon determining the bids for all classes by doing this for every class, take the optimal template T via the second-highest allocation rule. We claim that this is an SNE. This is because no single player can unilaterally deviate to raise the second highest bid of their class and the bidders are already in an SNE for the winning template T. □

Unfortunately, this positive result appears to rely both on restricting the setting (to class selection templates) as well as the allocation rule (to an inherently inefficient one). The following theorem shows that with the standard allocation rule (optimizing OBJ) there may not exist an SNE for conservative bidders. As the example is somewhat lengthy, we defer the proof to Appendix C.

THEOREM 6.12. *Suppose we use the template-indifferent GSP payment rule with the standard allocation rule and furthermore, do not allow bidders to bid higher than their true value as before. Then there may not exist any SNE.*

The example chosen in the proof is not carefully constructed. The sole essential ingredient is that the winning class A bidders' true values are extremely close in value. This ensures that for every bidder, the range of bids allowed in an SNE is small. In fact, one can show via a similar proof that any class selection set such that every class appears in every template is enough to ensure there exists no SNE — if the types/valuations of the bidders are chosen carefully.

7. DISCUSSION

Our analysis examined trade-offs among different stakeholders in ad auctions. We have shown how to make these trade-offs optimally and how to handle constraints on them. We have also used simulations and real-world auction data to show the effectiveness of this approach, examining the impact of the allowed number of ads on the welfare of the interested parties. In doing so, we demonstrated the efficacy of a per-impression reserve for this purpose, a tool which had previously been shown to have poor revenue-increasing properties.

We then examined a richer domain, where ads may be shown in several mutually exclusive templates. Here our results are mostly negative, and we demonstrated the

shortcomings of various natural generalizations of GSP. While we were able to show some positive results, they were for restricted settings and allocation rules that did not achieve optimal trade-offs.

Our results on templates point to several directions for future research. First, are our examples for template auctions under natural GSP generalizations, which show arbitrarily bad or no equilibrium, representative of real world ad-auctions? In particular, a few of our constructions rely on a large number of bidders and slots. Second, can good equilibria be found in ad auctions with templates under less demanding assumptions than we made to achieve our positive results? Finally, could a different generalization or set of assumptions lead to the existence of (approximately) optimal equilibria?

REFERENCES

AGGARWAL, G., GOEL, A., AND MOTWANI, R. 2006. Truthful auctions for pricing search keywords. In *ACM Conference on Electronic Commerce*. 1–7.

ATHEY, S. AND ELLISON, G. 2011. Position auctions with consumer search. *The Quarterly Journal of Economics 126,* 3, 1213–1270.

BULOW, J. AND KLEMPERER, P. 1996. Auctions versus negotiations. *American Econ. Rev. 86,* 1, 180–194.

DENG, X., SUN, Y., YIN, M., AND ZHOU, Y. 2010. Mechanism design for multi-slot ads auction in sponsored search markets. In *FAW*. 11–22.

DIAKONIKOLAS, I., PAPADIMITRIOU, C., PIERRAKOS, G., AND SINGER, Y. 2012. Efficiency-revenue trade-offs in auctions. In *ICALP'12*. 488–499.

EDELMAN, B., OSTROVSKY, M., AND SCHWARZ, M. 2007. Internet advertising and the Generalized Second-Price auction: Selling billions of dollars worth of keywords. *American Economic Review 97,* 1, 242–259.

EDELMAN, B. AND SCHWARZ, M. 2010. Optimal auction design and equilibrium selection in sponsored search auctions. *American Economic Review 100,* 2.

GOEL, G. AND KHANI, M. R. 2014. Revenue monotone mechanisms for online advertising. WWW '14.

JEHIEL, P. AND LAMY, L. 2013. On discrimination in procurement auctions.

LAHAIE, S. AND MCAFEE, R. P. 2011. Efficient ranking in sponsored search. In *Proc. 7th international conference on Internet and Network Economics*. WINE'11. Singapore, 254–265.

LAHAIE, S. AND PENNOCK, D. M. 2007. Revenue analysis of a family of ranking rules for keyword auctions. In *Proc. 8th ACM conference on Electronic Commerce*. EC'07. San Diego, California, 50–56.

LIKHODEDOV, A. AND SANDHOLM, T. 2003. Auction mechanism for optimally trading off revenue and efficiency. In *Proceedings of the 4th ACM Conference on Electronic Commerce*. EC '03. 212–213.

LIU, D. AND CHEN, J. 2006. Designing online auctions with past performance information. *Decision Support Systems 42,* 3, 1307–1320.

LIU, D., CHEN, J., AND WHINSTON, A. B. 2010. Ex ante information and the design of keyword auctions. *Information Systems Research 21,* 1, 133–153.

MYERSON, R. B. 1981. Optimal auction design. *Mathematics of operations research 6,* 1, 58–73.

OSTROVSKY, M. AND SCHWARZ, M. 2011. Reserve prices in internet advertising auctions: a field experiment. In *Proc. 12th ACM conference on Electronic Commerce*. EC'11. San Jose, California, 59–60.

PAES LEME, R. AND TARDOS, E. 2009. Sponsored search equilibria for conservative bidders. In *Ad Auctions Workshop*.

ROBERTS, B., GUNAWARDENA, D., KASH, I. A., AND KEY, P. 2013. Ranking and tradeoffs in sponsored search auctions. In *Proc. of the Fourteenth ACM Conf. on Electronic Commerce*. EC '13. 751–766.

SUNDARARAJAN, M. AND TALGAM-COHEN, I. 2013. Refine predictions ad infinitum? *CoRR abs/1302.6700*.

THOMPSON, D. R. AND LEYTON-BROWN, K. 2013. Revenue optimization in the generalized second-price auction. In *Proc. of the Fourteenth ACM Conf. on Electronic Commerce*. EC '13. ACM, 837–852.

VARIAN, H. R. 2007. Position auctions. *International Journal of Industrial Organization 25,* 6, 1163–1178.

A. PROOFS OF SECTION 4

Recall that we are assuming:

— The type space of allowable types T is a closed compact set.
— The allocation function $x : T \to \mathbb{R}^n$ is a Lebesgue integrable function $\in L^1$.
— The constraint feasibility region $\mathcal{X}$ for the allocation function x is convex w.r.t. the underlying function space, namely the L^1 Banach space w.r.t. Lebesgue measure

In addition we shall assume that the linear system (w.r.t. the revenue, welfare, and click yield) of constraints is non-degenerate.

En route to proving Theorem 4.1, we first show that two critical properties hold: weak duality and Lagrangian sufficiency. We prove the results for equality constraints $A(x) = \theta$ (the same results hold true for inequality constraints, $A(x) \geq \theta$ with the provision that $\lambda \geq 0$, and can be similarly proved). The Lagrangian in this case is

$$L(x, \lambda) = G(x) - \lambda^T (\theta - A(x)).$$

The primal problem is

$$\Gamma(\theta) = \sup_{x \in \mathcal{X}(\theta)} G(x)$$

where $\mathcal{X}(\theta) = \{x : x \in \mathcal{X} \text{ and } A(x) = \theta\}$ is the feasible set, and the dual is denoted

$$h(\lambda) = \sup_{x \in \mathcal{X}} L(x, \lambda).$$

LEMMA A.1 (WEAK DUALITY). *For all λ, $h(\lambda) \geq \Gamma(\theta)$ and in particular for the optimal value of the dual, $h^* = \inf_\lambda h(\lambda) \geq \Gamma(\theta)$.*

PROOF.
$$h(\lambda) = \sup_{x \in \mathcal{X}} L(x, \lambda) \geq \sup_{x \in \mathcal{X}(\theta)} L(x, \lambda) = \sup_{x \in \mathcal{X}(\theta)} G(x) = \Gamma(\theta) \quad \square$$

LEMMA A.2 (LAGRANGIAN SUFFICIENCY). *If there is an x^* and λ^* such that*

$$L(x^*, \lambda^*) = \sup_{x \in \mathcal{X}} L(x, \lambda^*)$$

and x^* is feasible for the constrained optimization (i.e. $x^* \in \mathcal{X}(\theta)$) then x^* is optimal.

PROOF. For all $x \in \mathcal{X}(\theta)$, we have $G(x) = L(x, \lambda) + \lambda^T (\theta - A(x))$. Since $x^* \in \mathcal{X}(\theta)$,

$$G(x^*) = L(x^*, \lambda^*) = \sup_{x \in \mathcal{X}} L(x, \lambda^*) \geq L(x, \lambda^*) = G(x) \quad \text{for all } x \in \mathcal{X}(\theta)$$

and hence x^* is optimal and $G(x^*) = \Gamma(\theta)$. $\square$

We now use the lemmas to prove Theorem 4.1, restated here.

THEOREM A.3. *If $\theta \in \operatorname{relint}(\operatorname{dom}(\Gamma))$, then the optimization problem is strong Lagrangian.*

PROOF.

(1) G is concave (indeed, it is a linear functional), the constraint set $\mathcal{X}$ is convex, and the constraints involve a vector of linear functionals, hence we can show directly that Γ is concave.
Assuming $\operatorname{dom}(\Gamma) \neq \varnothing$, take $\theta^1, \theta^2 \in \operatorname{relint}(\operatorname{dom}(\Gamma))$. Then there exists x^1, x^2 such that $G(x^i) = \Gamma(\theta^i)$ for $i \in \{1, 2\}$; these are feasible: $x^i \in \mathcal{X}(\theta^i)$. Now for any arbitrary $\lambda \in (0, 1)$ let $\tilde{\theta} = \lambda\theta^1 + (1 - \lambda)\theta^2$ and similarly $\tilde{x} = \lambda x^1 + (1 - \lambda)x^2$. By convexity of $\mathcal{X}$ and the linearity of constraints $A(x) = \theta$, we then have $\tilde{x} \in \mathcal{X}(\tilde{\theta})$. Thus:

$$\begin{aligned}
\Gamma(\tilde{\theta}) &= \sup_{x \in \mathcal{X}(\tilde{\theta})} G(x) \\
&\geq G(\tilde{x}) \\
&= G(\lambda x^1 + (1 - \lambda)x^2) \\
&= \lambda_1 G(x^1) + (1 - \lambda)G(x^2) \\
&= \lambda\Gamma(\theta^1) + (1 - \lambda)\Gamma(\theta^2)
\end{aligned}$$

and so Γ is concave.

(2) Since Γ is concave, for $\theta \in \operatorname{relint}(\operatorname{dom}(\Gamma))$, there is a non-vertical supporting hyperplane (NVSH) to Γ at θ (which means that the hypograph of Γ has a supporting hyperplane), which is equivalent to the problem being strong Lagrangian. Note that $\Gamma : \theta \to /\mathbb{R}$ is defined on Euclidean space. That such a problem is strong Lagrangian for $\Gamma(b)$ if and only if there is a NVSH at b is a well known theorem, and simple to prove directly. For the "if" direction: suppose there exists a NVSH at b, then there is a $\lambda \in /\mathbb{R}^{\rho(A)}$ such that

$$\Gamma(c) \leq \Gamma(b) + \lambda \cdot (c - b) \quad \text{for all } c \in /\mathbb{R}^{\rho(A)}.$$

Hence

$$\begin{aligned}
\Gamma(b) &\geq \sup_{c \in /\mathbb{R}^{\rho(A)}} [\Gamma(c) - \lambda \cdot (c - b)] \\
&= \sup_{c \in /\mathbb{R}^{\rho(A)}} \sup_{x \in \mathcal{X}(c)} [G(x) - \lambda \cdot (A(x) - b)] \\
&= \sup_{x \in \mathcal{X}} L(x, \lambda) \\
&= h(\lambda).
\end{aligned}$$

But from Weak Duality (Lemma A.1) $\Gamma(b) \leq h(\lambda)$, and thus $\Gamma(b) = h(\lambda)$ and the problem is strong Lagrangian. The "only if" is similarly proved by reversing the argument.

$\square$

When the problem is strong Lagrangian, then $h^* = \Gamma(\theta)$, and complementary slackness applies.

We can now prove the equivalent of Slater's constraint qualification to derive sufficient conditions to have $\theta \in \operatorname{relint}(\operatorname{dom}(\Gamma))$. This is Corollary A.4 which we restate and prove.

COROLLARY A.4. *When $Z = \{z \in /\mathbb{R}^{\rho(A)} : \leq 0\}$ is the non-positive cone, the problem $X_Z(\theta)$, is strong Lagrangian if there is some $\tilde{x} \in \mathcal{X}$ such that $A(\tilde{x}) > \theta$. If $Z = 0$, the problem $X_Z(\theta) = X(\theta)$ is strong Lagrangian provided there is some $\tilde{x} \in \operatorname{relint}(\mathcal{X})$ such that $A(\tilde{x}) = \theta$.*

PROOF. When the conditions are satisfied, then clearly $\theta \in \operatorname{relint}(\operatorname{dom}(\Gamma))$, and hence the result. Note that we do require the condition when $Z = 0$ because we have folded the constraints into $\mathcal{X}$, which means that is possible that $\theta \in \operatorname{bdd}\phi$ because $x \in \operatorname{bdd}\mathcal{X}$; when on the boundary, the separating hyperplane for Γ at θ may be vertical, in which case problem is not strong Lagrangian. The condition when $Z = 0$ precludes this from happening. $\square$

B. PROOF OF THEOREM 6.6

PROOF. Suppose we have the following single-class setup where all the $w_i = 1$ and we are maximizing welfare.

— Bidders:
  — $2m$ bidders with true values $m/2 - \epsilon, m/2 - 2\epsilon, ..., m/2 - (2m)\epsilon$.
  — m^2 bidders with true values $1 - \epsilon, 1 - 2\epsilon, ..., 1 - m^2\epsilon$.
— Templates:
  — Template 1: $2m + m^2 - 1$ slots of effect $1 - \epsilon, 1 - 2\epsilon, ..., 1 - 2m\epsilon, \epsilon, \epsilon^2, ..., \epsilon^{m^2-1}$.
  — Template 2: $2m + m^2 - 1$ slots of effect $x - \epsilon, x - 2\epsilon, ..., x - (2m + m^2 - 1)\epsilon$ where $x = 2(2m - 1)/(2m + m^2)$ which is equivalent to $x \approx 4/m$ for large m.

We first show that the first template will always win in any NE of the truthful auction for sufficiently large m and small ϵ[3]. Recall that in an NE of the truthful auction, a bidder will always bid higher than the true value of the next highest bidder (as otherwise the next highest bidder would outbid them). As we are also assuming bidders cannot bid higher than their true value this implies that the first bidder (i.e. the bidder with valuation $m/2 - \epsilon$) will have bid $b_1 \in [m/2 - \epsilon, m/2 - 2\epsilon]$. More generally, bidder $i < 2m + m^2$ will bid $b_i \in [t_{i-1}, t_i]$. We therefore find that the calculated objective of the first template subtracted by the second is (where calculations are in the limit for large m and small ϵ):

$$(b_1 + b_2 + ... + b_{2m}) - (b_1 + b_2 + ... + b_{2m+m^2-1})x$$
$$= (b_1 + b_2 + ... + b_{2m-1})(1 - x) + b_{2m}(1 - x) - (b_{2m+1} + b_{2m+2} + ... + b_{2m+m^2-1})x$$
$$= (2m - 1)(m/2)(1 - x) + b_{2m}(1 - x) - (m^2 - 1)x$$
$$> (2m - 1)(m/2)(1 - x) - (m^2 - 1)x$$
$$= m^2$$
$$> 0.$$

As this is greater than zero, the first template will always win for sufficiently large m and small ϵ.

We now show that there exists a problematic SNE when we use the template-considerate GSP payment rule where the second template wins. Suppose we have an SNE on the second template, ignoring the first. Assuming $\epsilon \ll x$, the slots in the second template are nearly identical. Thus, if we ignore the first template, the bidders will only bid enough to ensure they win a spot — that is, enough to outbid the last bidder who has a true value of $1 - m^2\epsilon \approx 1$; with the caveat that the first bidder may bid up to their true valuation. We now claim this is an SNE overall, i.e. when we consider this in the context of *both* templates. A bidder who wishes to change to the first template will raise their bid to at most their true value. Let us consider the objective values of the two templates after this maximum unilateral deviation. The first template will have an objective value of at most $\approx 2(m/2) + (m/2 - 2)$ if we assume that one bidder is already bidding their true value. For large m, this is approximately $3m/2$. On the other hand, the second template will have objective value of at least $\approx (2m + m^2 - 1)x$ if all but the last bidder bid approximately 1. For large m this is approximately $4m$. As $4m > 3m/2$, we find that no unilateral deviation can switch the template and so we have an SNE overall.

Now note that this SNE has a welfare of $\approx ((2m)(m/2) + (m^2))x \in \theta(m)$. On the other hand, all NE of the truthful auction must use the first template, so their welfare is $\approx (2m)(m/2) \in \theta(m^2)$. Thus, this SNE is arbitrarily worse than any NE of the truthful auction. □

C. PROOF OF THEOREM 6.12

PROOF. Suppose we are maximizing welfare, all bidders have ad-effect $w_i = 1$ and we have the following setup.

— Classes A and B.
— Bidders:
    — Four class A bidders with true values 100, $100 - \epsilon$, $100 - 2\epsilon$, and 20.
    — Two class B bidders with true values 150 and 135.

---

[3]Specifically, we mean taking a large m and *then* taking a small ϵ. For our purposes, it will suffice if $\epsilon \ll 1/m^2$.

— Templates ($\Delta \in (0,1)$):
   — Template 1: three class A slots of effects 1, $1/2$, and $1/4$, and a class B slot effect of Δ.
   — Template 2: three class A slots of effects Δ, $\Delta/2$, and $\Delta/4$, and a class B slot effect of 1.
   Note that if $\Delta \approx 0$ we have a situation similar to our guiding example that is the image/text auction.

Let us consider the general properties of any SNE. Regarding class B bidders, it is clear that the second bidder will always bid his true value of 135, and the first bidder will outbid him. Now let us consider the far more complex class A bidders. Let b_i denote the bid of class A's bidder i. Clearly $b_4 = 20$ as he will not win a slot and must therefore bid his true value to ensure he is not envious. We further find that:

— As the third bidder must not envy the second bidder:
$$(100 - 2\epsilon - b_4)/4 \geq (100 - 2\epsilon - b_3)/2$$
$$\Rightarrow b_3 \geq 60 - \epsilon.$$

— As the second bidder must not envy the third bidder:
$$(100 - \epsilon - b_3)/2 \geq (100 - \epsilon - b_4)/4$$
$$\Rightarrow b_3 \leq 60 - \epsilon/2.$$

— As the second bidder must not envy the first bidder:
$$(100 - \epsilon - b_3)/2 \geq 100 - \epsilon - b_2$$
$$\Rightarrow b_2 \geq 50 + b_3/2 - \epsilon/2.$$

— As the first bidder must not envy the second bidder:
$$100 - b_2 \geq (100 - b_3)/2$$
$$\Rightarrow b_2 \leq 50 + b_3/2.$$

Together, these imply that:
$$b_3 \in [60 - \epsilon, 60 - \epsilon/2],$$
$$b_2 \in [50 + b_3/2 - \epsilon/2, 50 + b_3/2]$$
$$\subseteq [50 + (60 - \epsilon)/2 - \epsilon/2, 50 + (60 - \epsilon/2)/2]$$
$$= [80 - \epsilon, 80 - \epsilon/4].$$

We now claim that in any SNE the second template is selected via the standard selection rule. To see this, let us consider the objective value of the first template subtracted by the second (when we assume $b_1 = b_2$).
$$(b_2 + b_2/2 + b_3/4 + \Delta 135) - (\Delta(b_2 + b_2/2 + b_3/4) + 135)$$
$$= (1 - \Delta)(b_2 + b_2/2 + b_3/4 - 135)$$
$$\leq (1 - \Delta)(3(80 - \epsilon/4)/2 + (60 - \epsilon/2)/4 - 135)$$
$$= -(1 - \Delta)(\epsilon/2)$$
$$< 0$$

where the first inequality comes from choosing b_3 and b_2 maximally. We next claim that for ϵ sufficiently small the third bidder can raise his bid b_3 so that it is beneficial to him in any SNE by causing the first template to be selected — which would imply that there exists no SNE. To see this, let us consider the objective value of the first template

subtracted by the second in an SNE when the third bidder unilaterally deviates and raises his bid to 70.

$$(b_2 + b_2/2 + 70/4 + \Delta 135) - (\Delta(b_2 + b_2/2 + 70/4) + 135)$$
$$= (1 - \Delta)(3b_2/2 + 70/4 - 135)$$
$$\geq (1 - \Delta)(3b_2/2 + 70/4 - 135)$$
$$\geq (1 - \Delta)(3(80 - \epsilon)/2 + 70/4 - 135)$$
$$\geq (1 - \Delta)(5/2 - 3\epsilon/2)$$
$$> 0$$

where the last inequality holds for sufficiently small ϵ. As $70 < b_2$ for sufficiently small ϵ, the third bidder will remain in the third position and his payment will be unaffected. Therefore, the third bidder will have incentive to raise his bid and so we cannot have an SNE. $\square$