

Incentivized Optimal Advert Assignment via Utility Decomposition

FRANK KELLY, University of Cambridge,
PETER KEY, Microsoft Research,
NEIL WALTON, University of Amsterdam.

We develop a framework for the analysis of large-scale Ad-auctions where adverts are assigned over a continuum of search types. Via a decomposition argument, the social welfare can be maximized through separate optimizations conducted by the advertisement platform and advertisers. The framework assumes a separation of time-scales, so that on each search occurrence the platform solves an assignment problem and, on a slower time scale, each advertiser varies her bid to match her demand for click-throughs with supply. By separating the problem in this way, knowledge of global parameters, such as the distribution of search terms, is not required. Exploiting the inherent information asymmetry between the platform and the advertiser, we describe a simple mechanism which implements Vickrey pricing and thus incentivizes truthful bidding. The mechanism has a unique Nash equilibrium, which maximizes social welfare. Finally, we consider models where advertisers adapt their bids smoothly over time, and prove convergence to the Nash equilibrium.

The framework is flexible and tractable: we describe extensions to allow complex page layouts, reserve prices, budget constraints, and the allocation of resources across multiple keywords.

Categories and Subject Descriptors: K.4.4 [Computers and Society]: Electronic Commerce

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Utility optimization, Decomposition

ACM Reference Format:

Frank Kelly, Peter Key, and Neil Walton, 2014. *ACM X*, X, Article X (June 2014), 33 pages.
DOI : <http://dx.doi.org/10.1145/0000000.0000000>

1. INTRODUCTION

Ad-auctions lie at the heart of search markets, and provide a mechanism where advertisers compete for their adverts to be shown to users of the search platform, by bidding on search terms associated with queries. Successful adverts are allocated to different positions on the search page (impressions), and if the user clicks on an ad, a payment is made by the advertiser to the platform under the Pay-Per-Click model.

In current auctions, there is a fundamental information asymmetry between the platform and an advertiser, in that the platform typically knows more than the advertiser about the searcher. Hence the platform can potentially choose prices and an allocation that depends on the platform's additional information. The advertiser has to rely on more coarse-grained information, perhaps just the search terms of a query together with a crude categorization of the user.

Neil Walton thanks The Netherlands Organization for Scientific Research (NWO) for funding through a VENI fellowship.

Authors' addresses: Frank Kelly, Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WB, UK; Peter Key, Microsoft Research Cambridge, 21 Station Road, Cambridge CB1 2FB, UK; Neil Walton, Science Park 904, 1098 XH Amsterdam, The Netherlands.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2014 ACM 0000-0000/2014/06-ARTX \$15.00

DOI : <http://dx.doi.org/10.1145/0000000.0000000>

In this paper we develop a framework to address this information asymmetry directly. We use an optimization decomposition and a separation of time-scales to show that it is possible to have the advertisers solve their individual utility maximization problems and the platform solve an assignment problem in such a way that these distributed optimizations yield a solution which maximizes social welfare. Such a framework is well known in the communication network community, where the phrase “Network Utility Maximization” has been coined, but has only recently found its way into Mechanism Design.

Randomness of search queries is an intrinsic aspect of our model. Much of the existing literature on Sponsored Search has needed to restrict attention to an isolated instance of an auction (a single query) to make progress ([Varian 2007; Edelman et al. 2007]). Athey and Nekepolov [2012] introduce a specific distributional assumption across the rank scores (weighted bids) to deal with randomness and provide a static analysis, while others [Pin and Key 2011] assume that randomness arises both from the number of bids in the auction and from randomness of the bids themselves.

Denote by p_{il}^τ the click-through rate (CTR) to advertiser i if the query is of type τ and her advert is shown in position l ; we assume that the platform can predict or estimate p_{il}^τ , but advertiser i only receives aggregated feedback, namely the click-through rate y_i to its advert, averaged over time and over the distribution of queries. We model advertisers as utility maximizers, who submit a bid λ_i and want to maximize their net benefit, namely $U_i(y_i) - \pi_i y_i$, where U_i is their private, concave, utility function and π_i the price they actually pay.

It is well known in the economic literature that market clearing prices that equate to marginal utility will maximize aggregate utility. However, this does not guarantee prices can be formulated and implemented on the time-scales relevant to sponsored search: adverts must be assigned per impression and charged per click, while advertisers maximize their long-run benefit. Further, the search space is vast, and the supply of search queries is unknown to all parties. But the contribution of our paper is to show that in this sponsored search setting we can separate the welfare maximization problem into relevant time-scales which can be solved, by using an optimization decomposition approach.

In Section 3 we apply this decomposition argument which is based on techniques from convex optimization and duality and show that if advertisers choose their bids λ_i to match their marginal utility, and the platform then solves an assignment problem to maximize revenue for each search instance τ using the submitted bids and the CTR estimates $\{p_{il}^\tau\}$, then the resulting solution maximizes aggregate utility. The advertisers are optimizing over a slower timescale than the platform, and the platform uses the submitted bids to solve an on-line maximum-weighted assignment problem, a form of generalized first price auction.

In Section 4 we make the connection with mechanism design and strategic advertisers. In particular, we find a form of a rebate which incentivizes advertisers to truthfully declare bids λ_i that equate to their marginal utility. This produces a unique Nash equilibrium which implements our decomposition. The rebate takes a form familiar from optimal tolls for traffic flow [Beckmann et al. 1956] or characterizations of truthful mechanisms [Vickrey 1961] for single parameter agent types [Myerson 1981; Babaioff et al. 2010], even though our advertisers are not described by a single parameter but by a utility function. We show that the rebate can be computed by a simple randomized mechanism that requires a single additional computation, namely the solution of an assignment problem, for each click-through. Hence assignment and pricing per search query involves straightforward polynomial-time computations, solving (at most two) assignment problems. For a simple example the prices paid for click-throughs are those arising from an ascending price auction. In Section 5 we consider dynamics and

convergence under adaptive bid updates by advertisers, and show that under smooth updating of bids, bid trajectories converge to the unique Nash equilibrium, and hence converge to the solution that maximizes aggregate utility.

In Section 6 we describe several extensions, intended to illustrate the flexibility and tractability of our framework.

Related Work

In this paper we consider a problem where the aggregate utility of an auction system is optimized subject to the capacity constraints of that system. The fundamental result of Eisenberg and Gale relates the equilibrium price of goods for buyers with linear utilities to a convex optimization problem [Fisher 1892; Eisenberg and Gale 1959]. Further, aggregate utility optimization has long been an objective in the design of effective market mechanisms, Vickrey [1961]. However, only in the recent literature have computationally efficient methods been considered for market and auction design, see Birnbaum et al. [2011], Jain and Vazirani [2007], Vazirani [2010]. In the context of electronic commerce and specifically Sponsored Search auctions, these computational considerations are of critical importance given the increased diversity and competition associated with online advertising.

We consider a decomposition approach to the task of optimizing advert allocation over the vast distribution of searches that can be conducted. This enables us to analyse the trade-off in objectives between the platform and advertiser. In recent work investigating trade-offs in sponsored search, [Roberts et al. 2013] focus on ranking algorithms, trading off revenue against welfare, while [Bachrach et al. 2014] also include the user as an additional stakeholder. In the decomposition applied in our work, a large optimization taken over the entire advert space is decomposed into numerous subproblems which can be implemented by each advertiser, and on each keyword search. The decompositions of interest are familiar, and have been important in the context of network design [Rockafellar 1984; Kelly 1997]. In particular, these decompositions suggest that current communication protocols converge to solutions of global optimization problems [Srikant 2004; Kelly and Yudovina 2014]. However, in this setting, agents are assumed to be price taking. Strategic formulations have since been developed: Johari and Tsitsiklis [2009] show a price of anarchy of 75% at a Nash equilibrium, and Yang and Hajek [2007] develop a single parameter VCG-mechanism to yield efficient resource allocation. The work of Srikant and Tan [2010] considers a queueing approach to optimizing average reward in an on-line advert campaign, using optimization decomposition ideas and an approach related to scheduling in wireless networks.

In our work, we first prove our decomposition result and then consider an adversarial setting with strategic buyers. We consider strategic buyers who look to optimize their long-run reward over a diverse set of auctions. This approach contrasts the one-shot and sequential approaches to sponsored search auctions, see [Varian 2007; Edelman et al. 2007; Syrgkanis and Tardos 2012]. Indeed the diverse stochastic variability found in sponsored search can make such approaches unrealistic [Pin and Key 2011]. Nonetheless, there are interpretations of the prices set in our framework in terms of one-shot auctions. The textbook by Vohra [2011] treats mechanism design using a linear programming (optimization) framework, making analogous connections with network flow problems.

Finally, we consider the dynamics of our auction mechanism when advertisers are allowed to slowly change their bids. We prove convergence to the unique Nash equilibrium, by exhibiting a Lyapunov function for the system. The work of Yang and Hajek [2006] is close in spirit to our approach in this Section.

2. THE ASSIGNMENT MODEL

We begin with notation that reflects a Sponsored Search setting, where a limited set of adverts are shown in response to users submitting search queries. We let $i \in \mathcal{I}$ index the large, finite set of advertisers. Each advertiser has an advert which they wish to be shown on the pages of a variety of search results. An advert, when shown, is placed in a slot $l \in \mathcal{L}$. Let $\tau \in \mathcal{T}$ index the *type* of a search conducted by a user. The set \mathcal{T} is an infinitely large set. It incorporates information such as the keywords of the search, but also other factors such as the geographic region where the search is conducted, the time of day, the gender of the searcher etc. All of these factors are combined to form a probability of click-through p_{il}^τ which is estimated by the search provider.

Over time, a large number of searches from the set \mathcal{T} are made. We assume these occur with distribution \mathbb{P}_τ . Thus we view the click-through probability $p_{il} : \mathcal{T} \rightarrow [0, 1]$ as a random variable defined on the type space \mathcal{T} , with distribution \mathbb{P}_τ . For example, the random variables $p = (p_{il} : i \in \mathcal{I}, l \in \mathcal{L})$ might admit a joint probability density function, $f(p)$. So, for $c = (c_{il} : i \in \mathcal{I}, l \in \mathcal{L}) \in [0, 1]^{\mathcal{I} \times \mathcal{L}}$,

$$\mathbb{P}_\tau(p \leq c) = \int_{[0,1]^{\mathcal{I} \times \mathcal{L}}} \mathbf{I}[p \leq c] f(p) dp.$$

Here \mathbf{I} is the indicator function and vector inequalities, e.g. $p \leq c$, are taken componentwise, $p_{il} \leq c_{il} \forall i \in \mathcal{I}, l \in \mathcal{L}$.

We exploit the inherent randomness in p_{il} for the optimal placement of adverts. We shall assume that the platform has access to information about the query, knows τ , and so can successfully predict the click-through probability p_{il}^τ , whilst the advertiser does not have access to such fine-grained search information. Later, in Sections 4 and 5, we shall see that the platform can use this information asymmetry to guide the auction towards an optimal outcome.

Next we describe a mechanism by which the platform may wish to assign adverts. Suppose advertiser i submits a bid λ_i , which reflects what the advertiser is willing to pay for a click-through. Let $\lambda = (\lambda_i, i \in \mathcal{I})$. Given the information (τ, λ) , the following optimization maximizes the expected revenue from a single search.

ASSIGNMENT (τ, λ)

$$\text{Maximize} \quad \sum_{i \in \mathcal{I}} \lambda_i \sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau \quad (1a)$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}} x_{il}^\tau \leq 1, \quad l \in \mathcal{L}, \quad (1b)$$

$$\sum_{l \in \mathcal{L}} x_{il}^\tau \leq 1, \quad i \in \mathcal{I}, \quad (1c)$$

$$\text{over} \quad x_{il}^\tau \geq 0, \quad i \in \mathcal{I}, l \in \mathcal{L}. \quad (1d)$$

The above optimization is an assignment problem, where the constraint (1b) prevents a slot containing more than one advert, and the constraint (1c) prevents any single advert being shown more than once on a search page. The solution is a maximum weighted matching of advertisers \mathcal{I} with slots \mathcal{L} . This is highly appealing from a computational perspective, firstly, because assignment problems can be solved efficiently [Kuhn 1955; Bertsekas 1988] and, secondly, because there is no need to pre-compute the assignment. The assignment problem can be solved on each occurrence of a search of type $\tau \in \mathcal{T}$. We do not need to estimate the distribution of searches \mathbb{P}_τ but we require an estimate of the click-through probability p_{il}^τ , as is the case for the Generalized Second Price (GSP) auctions currently used in sponsored search.

We apply the convention that if $\lambda_i = 0$ then $x_{il}^\tau = 0$ for $l \in \mathcal{L}$, so that a zero bid does not receive clicks: this can be achieved by adding slots corresponding to adverts not being shown. We make the mild assumption that a solution x^τ to the above problem is unique and is integral with probability one; this would follow, for example, if the distribution of click-thoughts p admits a density.

Let

$$y_i^\tau = \sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau, \quad y_i = \mathbb{E}_\tau y_i^\tau, \quad x_{il} = \mathbb{E}_\tau [x_{il}^\tau]. \quad (2)$$

Note that y_i^τ is the click-through rate for advertiser i from a given search page, and y_i is the click-through rate averaged over \mathcal{T} . (We shall *not* use y_i for the random variable y_i^τ .) We assume that y_i^τ is known to the platform, and y_i is known to advertiser i .

For an optimal solution to the above assignment problem, we shall write $x_{il}^\tau = x_{il}^\tau(\lambda)$ to emphasize the dependence of x_{il}^τ on the vector of bids λ and, similarly, we write $x_{il} = x_{il}(\lambda)$, $y_i^\tau = y_i^\tau(\lambda)$, $y_i = y_i(\lambda)$.

We shall assume the following strict monotonicity property of solutions of $\text{ASSIGNMENT}(\tau, \lambda)$. We assume the function $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ is strictly increasing and continuously differentiable over $\lambda_i \geq 0$, for all values of $\lambda_{-i} = (\lambda_j : j \neq i, j \in \mathcal{I})$ in the positive orthant. Without the regularity condition $y_i(\lambda)$ will be increasing in λ_i but may not be strictly increasing, and this would complicate the statement of several later results.

The strict monotonicity property is satisfied if p admits a density satisfying the following regularity condition. Suppose $f(p)$ is continuous and bounded above by a constant $f_{max} < \infty$ for all $p \in \mathcal{C} = [0, 1]^{\mathcal{I} \times \mathcal{L}}$ and bounded below by a constant $f_{min} > 0$ for all $p \in \mathcal{S} = \{p \in \mathcal{C} : p_{il} \geq p_{ik}, l < k\}$. Observe that in the simplex \mathcal{S} the click-through probability for a given advert decreases as the slot it is shown in increases. We do not require a lower bound on $f(p)$ outside of this simplex. In Lemma C.3 and Proposition 5.2 it is shown that this regularity condition on $f(p)$ implies that $y_i(\lambda)$ possesses our assumed strict monotonicity property.

3. OPTIMIZATION PRELIMINARIES

In this section we present an optimization problem which can be motivated as the maximization of social welfare; we use the problem to develop various decomposition and duality results which we shall need in the next Section.

Suppose that if advertiser i achieves a click-through rate of y_i this has a utility to advertiser i of $U_i(y_i)$. We assume the function $U_i(\cdot)$ is non-negative, increasing, and strictly concave, and that our objective is to place adverts so as to maximize the sum of these utilities. To simplify the statement of results, we shall assume further that $U_i(\cdot)$ is continuously differentiable, with $U_i'(y_i) \rightarrow \infty$ as $y_i \downarrow 0$ and $U_i'(y_i) \rightarrow 0$ as $y_i \uparrow \infty$. The resulting optimization problem is as follows.

SYSTEM($U, \mathcal{I}, \mathbb{P}_\tau$)

$$\text{Maximize} \quad \sum_{i \in \mathcal{I}} U_i(y_i) \quad (3a)$$

$$\text{subject to} \quad y_i = \mathbb{E}_\tau \left[\sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau \right], \quad i \in \mathcal{I}, \quad (3b)$$

$$\sum_{i \in \mathcal{I}} x_{il}^\tau \leq 1, \quad l \in \mathcal{L}, \tau \in \mathcal{T}, \quad (3c)$$

$$\sum_{l \in \mathcal{L}} x_{il}^\tau \leq 1, \quad i \in \mathcal{I}, \tau \in \mathcal{T}, \quad (3d)$$

$$\text{over} \quad x_{il}^\tau \geq 0, y_i \geq 0 \quad i \in \mathcal{I}, l \in \mathcal{L}. \quad (3e)$$

Inequalities (3c) and (3d) are just the scheduling constraints (1b) and (1c) that each slot can show at most one advert and that each slot can show at most one advert, while equality (3b) recaps the definition (2) of y_i , the expected click-through rate.

Note the optimization assumes the system optimizer has knowledge of the utility functions $U_i(\cdot)$, $i \in \mathcal{I}$, the click-through probabilities p_{il}^τ , $i \in \mathcal{I}, l \in \mathcal{L}, \tau \in \mathcal{T}$ and the distribution \mathbb{P}_τ over these probabilities. In the next section, on mechanism design, we consider the game theoretic aspects that arise when, instead of a single system optimizer, the platform and advertisers have differing information and incentives.

Incorporate the constraint (3b) into the objective function (3a) to give the Lagrangian

$$L_{sys}(x, y; \lambda) = \sum_{i \in \mathcal{I}} U_i(y_i) + \sum_{i \in \mathcal{I}} \lambda_i \mathbb{E}_\tau \left[\sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau - y_i \right]. \quad (4)$$

Notice, we intentionally omit the scheduling constraints from our Lagrangian. Thus we seek to maximize the Lagrangian subject to the constraints (3c-3d) as well as (3e). Let \mathcal{A} be the set of variables $x = (x_{il}^\tau : i \in \mathcal{I}, l \in \mathcal{L}, \tau \in \mathcal{T}) \geq 0$ satisfying the assignment constraints (3c-3e). We see that our Lagrangian problem is separable in the following sense

$$\max_{x \in \mathcal{A}, y \geq 0} L_{sys}(x, y; \lambda) = \sum_{i \in \mathcal{I}} \max_{y_i \geq 0} \{U_i(y_i) - \lambda_i y_i\} \quad (5a)$$

$$+ \mathbb{E}_\tau \left[\max_{x^\tau \in \mathcal{A}} \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i p_{il}^\tau x_{il}^\tau \right]. \quad (5b)$$

Define

$$U_i^*(\lambda_i) = \max_{y_i \geq 0} \{U_i(y_i) - \lambda_i y_i\}. \quad (6)$$

The optimization over y_i contained in the definition (6) would arise if advertiser i were presented with a fixed price per click-through of λ_i , and allowed to choose freely her click-through rate: she would then choose y_i such that $U_i'(y_i) = \lambda_i$. By our assumptions on $U_i(\cdot)$, this equation has a unique solution for all $\lambda_i \in (0, \infty)$. Call $D_i(\xi) = \{U_i'\}^{-1}(\xi)$ the *demand* of advertiser i at price ξ . It follows that $U_i^*(\lambda_i)$ can be written in the form

$$U_i^*(\lambda_i) = \int_{\lambda_i}^{\infty} D_i(\xi) d\xi; \quad (7)$$

call this advertiser i 's *consumer surplus* at the price λ_i . From this expression we can deduce that $U_i^*(\lambda_i)$ is positive, decreasing, strictly convex and continuously differentiable.

Observe that the maximization inside the expectation (35b) is simply the problem $\text{ASSIGNMENT}(\tau, \lambda)$, and thus we can write

$$\max_{x \in \mathcal{A}, y \geq 0} L_{sys}(x, y; \lambda) = \sum_{i \in \mathcal{I}} U_i^*(\lambda_i) + \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda).$$

The Lagrangian dual of the SYSTEM problem (3) can thus be written as follows.

DUAL(U^*, y, \mathcal{I})

$$\text{Minimize} \quad \sum_{i \in \mathcal{I}} (U_i^*(\lambda_i) + \lambda_i y_i(\lambda)) \quad (8a)$$

$$\text{over} \quad \lambda_i \geq 0, \quad i \in \mathcal{I}. \quad (8b)$$

Owing to the size of the type space \mathcal{T} , the optimization (3) has a potentially uncountable number of constraints. This presents certain technical difficulties, for instance those associated with proving strong duality. These issues are dealt with in the appendix, where the proof of the following two propositions are presented.

We first show that the SYSTEM problem decomposes into optimizations relevant to the advertisers and to the platform, for each search occurrence, τ .

PROPOSITION 3.1 (DECOMPOSITION). *Feasible variables \tilde{y} , \tilde{x}^τ , $\tau \in \mathcal{T}$, (3b-3e), are optimal for $\text{SYSTEM}(U, \mathcal{I}, \mathbb{P}_\tau)$ if and only if there exist $\tilde{\lambda}_i$, $i \in \mathcal{I}$, such that*

- A. $\tilde{\lambda}_i$ minimizes $U_i^*(\lambda_i) + \lambda_i \tilde{y}_i$ over $\lambda_i \geq 0$, for each $i \in \mathcal{I}$,
- B. \tilde{x}^τ solves $\text{ASSIGNMENT}(\tau, \tilde{\lambda})$, with probability one under the distribution \mathbb{E}_τ over $\tau \in \mathcal{T}$.

In this Proposition, the optimization in Condition A does not naturally correspond to the bidding behavior of strategic advertisers, at least in its present form. Hence we need to examine the implications of Condition A in order to construct prices that do give strategic advertisers the incentive to solve the SYSTEM problem, which we do in the next section, Section 4. The optimal bids $\tilde{\lambda}$ can be further understood through the following dual characterization.

PROPOSITION 3.2 (DUAL OPTIMALITY).

a) *The dual of the SYSTEM problem is DUAL. The objective of the problem (8) is continuously differentiable for $\lambda > 0$ and is minimized by any $\tilde{\lambda} = (\tilde{\lambda}_i : i \in \mathcal{I})$ satisfying, for each $i \in \mathcal{I}$,*

$$\frac{dU_i^*}{d\lambda_i}(\tilde{\lambda}_i) + y_i(\tilde{\lambda}) = 0. \quad (9)$$

b) *If $\tilde{\lambda}$ is an optimal solution to the DUAL problem (8) then $x^\tau(\tilde{\lambda})$, $y(\tilde{\lambda})$ are optimal for the SYSTEM problem (3).*

Example 3.3. For concreteness, we consider a brief example. Three advertisers, A , B and C compete for two advertisement slots shown in response to a specific search query, “Palo Alto Pizza”. Two of the advertisers, A and B , are takeaway pizza companies, one located in north Palo Alto and the other in south Palo Alto. Thus the click through rate of these advertisers is sensitive to the location of the search. The platform is aware of the location of the search whilst the advertisers are not. Thus the platform can exploit this asymmetry. The third advertiser who, say, sells supermarket products is not sensitive to the location but is sensitive to the advertisement slot position and their ad will only be clicked on if it appears in the top slot. The platform observes that,

given τ , the adverts receive click through probabilities

$$p_{Al}^\tau = \tau, \quad p_{Bl}^\tau = \frac{1}{2} - \tau, \quad p_{Cl}^\tau = \frac{1}{3} \mathbf{I}[l = 1], \quad \text{for } l = 1, 2, \quad (10)$$

and where τ accounts for random distance of the search from the advertiser. We assume τ is uniform random variable on $[0, 1/2]$, although this is not known by the platform or the advertisers. We suppose advertisers have the same logarithmic utilities $U_i(y) = \log y$ for $i = A, B, C$. Note that without the decomposition result Proposition 3.1), determining an optimal solution to the SYSTEM optimization is a non-obvious problem.

We suppose that $\lambda = (\lambda_A, \lambda_B, \lambda_C)$ are the prices submitted by the advertisers, and that these prices truthfully declare their marginal utilities $\lambda_i = U'_i(y_i)$. Notice that an advert is only assigned a slot in the ASSIGNMENT problem if it is not the lowest bid. For instance, for advertiser A, we can calculate the click-through rates

$$y_A(\lambda) = \mathbb{E}_\tau \left[\tau (1 - \mathbf{I}[\tau \lambda_A \leq (1/2 - \tau)\lambda, \tau \lambda_A \leq 1/3]) \right] = \frac{1}{4} - \min \left\{ \frac{\lambda_A}{2(\lambda_A + \lambda_B)}, \frac{\lambda_C}{3\lambda_A} \right\}^2. \quad (11)$$

For a logarithmic utility, the demand function is given by $D_A(\lambda_A) = \lambda_A^{-1}$. Thus, along with similar conditions for advertisers B and C, the optimal condition for the dual problem, (9), can be derived

$$\frac{1}{\lambda_A^*} = \frac{1}{4} - \min \left\{ \frac{\lambda_A^*}{2(\lambda_A^* + \lambda_B^*)}, \frac{\lambda_C^*}{3\lambda_A^*} \right\}^2. \quad (12)$$

There exist parameters satisfying these above conditions, and the CTR $y(\lambda^*)$ is then optimal for the SYSTEM problem.

Notice, to assign adverts, the platform required correct click through probabilities (10), search information τ with ‘‘Palo Alto Pizza’’ and prices λ , but did not require information about advertisers utilities or the distribution of searches \mathbb{P}_τ . Further, the advertiser could determine λ_A^* from its own advert average performance (11) and its utility function U_A , but did not require explicit knowledge of other advertisers utilities or the precise search type conducted τ or the distribution of searches \mathbb{P}_τ .

Finally, we remark that a much wider range of advertisers and search types can be considered. For instance, within our framework it is reasonable to assume that advertiser C will bid for search queries containing the word ‘‘Pizza’’ and thus will compete with a much wider class of advertisers.

4. MECHANISM DESIGN

In the last section we demonstrated how the global problem can be decomposed into two types of sub-problem: one, where the platform finds an optimal assignment given click-through probabilities; and the other, where the dual variables λ are each set to solve a certain single parameter dual problem. In this section we suppose the advertisers act strategically, anticipating the result of the search provider’s assignment and attempting to maximize their expected reward.

Henceforth we interpret λ_i as the *bid* submitted by advertiser i and, as a function of these bids, we formulate prices that incentivize the advertisers to choose bids that result in an assignment that optimizes the SYSTEM objective (3).

Consider a mechanism where, given the vector of bids $\lambda = (\lambda_i : i \in \mathcal{I})$, each advertiser, i , receives a click-through rate $y_i(\lambda)$, and from this derives a benefit $U_i(y_i(\lambda))$ and is charged a price $\pi_i(\lambda)$ per click. The reward to advertiser i arising from a vector of

bids $\lambda = (\lambda_i : i \in \mathcal{I})$ is then

$$r_i(\lambda) = U_i(y_i(\lambda)) - \pi_i(\lambda)y_i(\lambda). \quad (13)$$

A *Nash equilibrium* is a vector of bids $\lambda^* = (\lambda_i^* : i \in \mathcal{I})$ such that, for $i \in \mathcal{I}$ and all λ_i

$$r_i(\lambda^*) \geq r_i(\lambda_i, \lambda_{-i}^*). \quad (14)$$

Here $(\lambda_i, \lambda_{-i}^*)$ is obtained from the vector λ^* by replacing the i th component by λ_i .

The main result of this section is the following.

THEOREM 4.1. *If prices are charged according to the price function*

$$\pi_i(\lambda) = \frac{1}{y_i(\lambda)} \int_0^{\lambda_i} (y_i(\lambda) - y_i(\mu_i, \lambda_{-i})) d\mu_i \quad (15)$$

then there exists a unique Nash equilibrium, and it is given by the vector of optimal prices defined in the decomposition, Proposition 3.1.

Remark 4.2. The result states that, given adverts are assigned according to the assignment problem (1), the game theoretic equilibrium reached by advertisers attempting to maximize their respective rewards r_i solves the problem $\text{SYSTEM}(U, \mathcal{I}, \mathbb{P}_\tau)$. Since $y_i(\mu_i, \lambda_{-i})$ is a strictly increasing function of the bid μ_i , the price $\pi_i(\lambda)$ must be strictly lower than the bid λ_i . Setting a price lower than the submitted bid is a prevalent feature of online auctions used by search engines.

Remark 4.3. The price function (15) can be readily implemented by the platform at a computational cost of at most one additional instance of the assignment problem, as we now show. Suppose the platform solves $\text{ASSIGNMENT}(\tau, \lambda)$, and observes a click-through on (i, l) — that is the solution has $x_{il}^\tau = 1$, and the user clicks on the advert in position l , which is for advertiser i . If this happens the platform chooses μ_i uniformly and randomly on the interval $(0, \lambda_i)$ and additionally solves $\text{ASSIGNMENT}(\tau, (\mu_i, \lambda_{-i}))$. Let $y_i^\tau(\mu_i, \lambda_{-i}) = \sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau$ under a solution to this problem (the solution will be unique with probability one). The platform then charges advertiser i an amount

$$\lambda_i \left(1 - \frac{y_i^\tau(\mu_i, \lambda_{-i})}{y_i^\tau(\lambda)} \right) \quad (16)$$

for the click-through. This charge does not depend on the distribution \mathbb{P}_τ , and will lie between 0 and λ_i . Taking expectations over τ and μ_i shows that the expected rate of payment by advertiser i is

$$\mathbb{E}_{\tau, \mu_i} \left[\sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau \lambda_i \left(1 - \frac{y_i^\tau(\mu_i, \lambda_{-i})}{y_i^\tau(\lambda)} \right) \right] = \lambda_i (y_i(\lambda) - \mathbb{E}_{\mu_i} [y_i(\mu_i, \lambda_{-i})]) = \pi_i(\lambda)y_i(\lambda).$$

Thus the mechanism implements the reward function (13) where $\pi_i(\lambda)$ is given by expression (15), as desired.

Observe that the additional instance of the assignment problem does not determine the assignment, and thus will not slow down the page impression: rather, it is used to calculate the charge (16) for a click-through. Indeed, one could imagine a charge λ_i on the click-through, followed by a later rebate of a proportion $y_i^\tau(\mu_i, \lambda_{-i})/y_i^\tau(\lambda)$ of the charge. The rebate depends on the uniform random variable μ_i . Of course one could reduce the variance of the rebate on a particular click-through by averaging the calculation over a number of independent replications of the uniform random variable μ_i , but this would seem unnecessary, since there will remain a dependence on τ , which is perceived by the advertiser as a random variable. For a recent discussion of integral

estimation by random sampling in the context of truthful mechanisms, see [Babaioff et al. 2010].

There are other ways the price function (15) can be implemented. Note that

$$\int_0^{\lambda_i} y_i(\mu_i, \lambda_{-i}) d\mu_i = \sum_j \lambda_j y_j(\lambda) - \sum_{j \neq i} \lambda_j y_j(0, \lambda_{-i})$$

since both expressions share the same derivative with respect to λ_i , from Proposition C.4, and both expressions take the value 0 when $\lambda_i = 0$. Thus the price function (15) can be also implemented by a charge λ_i on the click-through followed by a later rebate of a proportion

$$\frac{1}{y_i^\tau(\lambda)} \left(\sum_j \lambda_j y_j^\tau(\lambda) - \sum_{j \neq i} \lambda_j y_j^\tau(0, \lambda_{-i}) \right)$$

of the charge. The rebate calculation again requires the solution of one additional instance of the assignment problem, in a form familiar as the VCG mechanism when the utility function for advertiser j , $j \in \mathcal{I}$, is simply the linear function $\lambda_j y_j$.

To establish Theorem 4.1 we will require Proposition 3.2 from the previous section and an additional result, Proposition 4.4, which indicates how maximal rewards achieved by each advertiser relate to the solution of the dual problem, Proposition 3.2.

PROPOSITION 4.4 (MECHANISM DUAL). *For each choice of $\lambda_{-i} = (\lambda_j : j \neq i, j \in \mathcal{I})$, the following equality holds*

$$\max_{\lambda_i \geq 0} r_i(\lambda) = \min_{\lambda_i \geq 0} \left\{ U_i^*(\lambda_i) + \int_0^{\lambda_i} y_i(\mu_i, \lambda_{-i}) d\mu_i \right\}. \quad (17)$$

Moreover, the optimizing λ_i for both expressions is the same, is unique and finite, and satisfies

$$\frac{d}{d\lambda_i} U_i^*(\lambda_i) + y_i(\lambda) = 0. \quad (18)$$

PROOF. We calculate the dual of the reward function (13). Let $P_i(y_i)$ be the function whose Legendre-Fenchel transform is

$$P_i^*(\lambda_i) = \int_0^{\lambda_i} y_i(\mu_i, \lambda_{-i}) d\mu_i.$$

We know from Fenchel's Duality Theorem, [Borwein and Lewis 2006, Theorem 3.3.5], that

$$\max_{y_i \geq 0} \{U_i(y_i) - P_i(y_i)\} = \min_{\lambda_i \geq 0} \{U_i^*(\lambda_i) + P_i^*(\lambda_i)\}. \quad (19)$$

So what remains is to calculate the function P_i from the dual of the function P_i^* above. By the Fenchel–Moreau Theorem [Borwein and Lewis 2006], we know this to be

$$P_i(y_i) = \min_{\lambda_i \geq 0} \left\{ \lambda_i y_i - \int_0^{\lambda_i} y_i(\mu_i, \lambda_{-i}) d\mu_i \right\}. \quad (20)$$

The optimum in this expression occurs when $y_i(\lambda) = y_i$. Substituting this back, since $\lambda_i \mapsto y_i(\lambda)$ is strictly increasing, we have that

$$P_i(y_i) = \int_0^\infty (y_i - y_i(\mu_i, \lambda_{-i})) \mathbf{I}[y_i(\mu_i, \lambda_{-i}) \leq y_i] d\mu_i. \quad (21)$$

In other words, as expected with the Legendre-Fenchel transform, the area under the curve $y_i(\lambda_i, \lambda_{-i})$ is converted to the area to the left of the curve $y_i(\lambda_i, \lambda_{-i})$. Further, notice, if $y_i > \max_{\lambda_i} y_i(\lambda_i, \lambda_{-i})$ then $P_i(y_i) = \infty$, and thus the finite range of the function $y_i \mapsto P_i(y_i)$ is exactly the same as that of $\lambda_i \mapsto P_i(y_i(\lambda))$. Noting (21) and this last observation, the equality (19) now reads

$$\begin{aligned} \min_{\lambda_i \geq 0} \left\{ U_i^*(\lambda_i) + \int_0^{\lambda_i} y(\mu_i, \lambda_{-i}) d\mu_i \right\} &= \max_{y_i \geq 0} \{U_i(y_i) - P_i(y_i)\} \\ &= \max_{\lambda_i \geq 0} \{U_i(y_i(\lambda)) - P_i(y_i(\lambda))\} \\ &= \max_{\lambda_i \geq 0} \{U_i(y_i(\lambda)) - \pi_i(\lambda)y_i(\lambda)\}. \end{aligned}$$

In the final equality we note from the definition (15) of $\pi_i(\lambda)$ that

$$P_i(y_i(\lambda)) = \pi_i(\lambda)y_i(\lambda). \quad (22)$$

This gives the equality (17).

For both of the optimizations in (17), the optimum is determined by the derivative of the objective with respect to $\lambda_i \in (0, \infty)$. In particular, we note that

$$U_i^{*'}(\lambda_i) + y_i(\lambda) < 0 \quad (23)$$

$$\iff U_i'(y(\lambda)) - \lambda_i > 0 \quad (24)$$

$$\iff \frac{\partial y_i(\lambda)}{\partial \lambda_i} (U_i'(y(\lambda)) - \lambda_i) > 0. \quad (25)$$

In the first equivalence, we use the fact that the inverse of the strictly decreasing function U_i' is $U_i^{*'}$. In the second equivalence, we note that $y_i(\lambda)$ is strictly increasing as function of λ_i . See Lemma C.3 for a proof that $y_i(\lambda)$ is strictly increasing. Since (23) gives the derivative of the objective on the right-hand side of (17) and (25) gives the derivative of the objective on the left-hand side of (17), an optimal λ_i simultaneously optimizes both expression in (17). Notice the optimal λ_i must be unique since the right-hand side of (17) is strictly convex. Finally, we note that the value of λ_i optimizing both expressions must be finite. Notice the inequality (24), which is satisfied at $\lambda_i = 0$, cannot be sustained for all λ_i since $U_i'(y(\lambda))$ is a decreasing function of λ_i . Thus the value of λ_i optimizing (17) must be finite. This completes the proof. \square

PROOF OF THEOREM 4.1. By Proposition 4.4 $\lambda = (\lambda_i : i \in \mathcal{I})$ is a Nash equilibrium if and only if condition (18) is satisfied for each $i \in \mathcal{I}$. But by Proposition 3.2b), these conditions hold if and only if the λ solves the dual to the SYSTEM problem. So, the set of Nash equilibria are the optimal prices defined for the decomposition, Proposition 3.1. By Proposition 3.2b), the assignment achieved by Nash equilibrium bids maximizes the utilitarian objective $\text{SYSTEM}(U, \mathcal{I}, \mathbb{P}_\tau)$. Finally, by Strong Duality (Theorem B.2), there exists λ^* which optimizes the dual problem (8), and thus must also be a Nash equilibrium. \square

Remark 4.5. The optimality condition (9) or (18) states that each advertiser's demand, $D_i(\lambda_i)$, and supply, $y_i(\lambda_i)$, should equate, and is a consequence of the envelope theorem. A more familiar context for this form of result is Vickrey pricing [Vickrey 1961] and Myerson's Lemma (or the Payoff Equivalence Theorem), see [Myerson 1981] and [Milgrom 2004, Theorem 3.3], which are also consequences of the envelope theorem. But observe that we are using general utilities, which despite the single input parameter λ_i , takes us out of a single parameter type space to which Myerson generally applies.

We can also interpret the system optimization as an infinitely large bipartite congestion game. The utility function of an advertiser gauges his sensitivity to different levels of congestion in a network. As with an equilibrium for traffic in a network [Wardrop 1952], if we charge the advertisers their bids, this can lead to an inefficient allocation of resources, a user equilibrium rather than a system optimum. However, by charging prices (15), we are effectively internalizing the externalities so that the user equilibrium maximizes social welfare [Beckmann et al. 1956].

Finally, we give two settings where closed forms are available.

Example 4.6. If there is a single slot then the slot will be assigned to the bidder i with the highest value of $\lambda_i p_{i1}^\tau$ and if this results in a click-through then the charge will be

$$\max_{j \neq i} \frac{\lambda_j p_{j1}^\tau}{p_{i1}^\tau},$$

a second price auction on the products $\lambda_j p_{j1}^\tau$.

Suppose next there are L slots, more than L advertisers bidding, and suppose the click-through probabilities are the same for each advertiser, $p_{il}^\tau = p_l^\tau, i \in \mathcal{I}$, where $p_1^\tau > p_2^\tau > \dots > p_L^\tau$, and the bids are $\lambda_1 > \lambda_2 > \dots$. Then advertisers $1, 2, \dots, L$ are allocated slots $1, 2, \dots, L$ respectively. In this example it is possible to calculate the expectation of the expression (16) explicitly, and thus to determine the charge without the need for randomization over μ_i : a click-through on slot l is charged the amount

$$\pi_l^\tau = \lambda_{l+1} - \frac{1}{p_l^\tau} \sum_{m=l+1}^L p_m^\tau (\lambda_m - \lambda_{m+1}), \quad l = 1, 2, \dots, L.$$

This implies

$$\pi_l^\tau = \lambda_{l+1} - \frac{p_{l+1}^\tau}{p_l^\tau} (\lambda_{l+1} - \pi_{l+1}^\tau), \quad l = 1, 2, \dots, L,$$

where $p_{L+1}^\tau = 0$, recovering the generalized English (or ascending price) auction of [Edelman et al. 2007]. The revenue received is

$$\sum_{m=1}^L \pi_m^\tau p_m^\tau = \sum_{m=1}^L m \lambda_{m+1} (p_m^\tau - p_{m+1}^\tau),$$

where $p_{L+1}^\tau = 0$; note the dependence on λ_{L+1} , the largest unsuccessful bid.

We note that as described (with click-through probabilities the same for all advertisers) our assumption on $y(\lambda)$, that $y_i(\lambda)$ is a strictly increasing function of λ_i , is not satisfied. But if with arbitrarily small probability $\epsilon > 0$ there exists another bidder with a random click-through probability with support $(0, 1)$, then our assumption on $y(\lambda)$ will be satisfied.

Example 4.7. If there is a single slot and advertisers CTRs are independent and identically distributed, with density function f and distribution function F , then it is a relatively straight forward calculation that $y_i(\lambda)$ and $\pi_i(\lambda)$ are

$$y_i(\lambda) = \int_0^1 \prod_{j \neq i} F\left(p \frac{\lambda_i}{\lambda_j}\right) \times pf(p) dp$$

$$\pi_i(\lambda) y_i(\lambda) = \int_0^1 \left[\lambda_i \prod_{j \neq i} F\left(p \frac{\lambda_i}{\lambda_j}\right) - \int_0^{\lambda_i} \prod_{j \neq i} F\left(p \frac{\mu}{\lambda_j}\right) d\mu \right] pf(p) dp.$$

5. DYNAMICS AND CONVERGENCE

In this section we consider whether advertisers will adapt their prices in order to converge towards an assignment of adverts that is optimal when averaged across the entire type space. The follow example illustrates a difficulty that may prevent convergence of prices.

Example 5.1. Suppose there is a single search type τ , and two identical advertisers compete over a single slot. Assuming their utilities identical and strictly concave, the solution to the system problem results in both advertisers equally sharing the slot, and equal prices: $\lambda_1 = \lambda_2$. This is consistent with Theorem 3.1, with both advertisers submitting equal bids: $\lambda_1 = \lambda_2$. But as the advertisers update their prices one cannot expect $\lambda_1 = \lambda_2$ to hold within the continuum of possible prices. Most simple price update rules will not lead to equal bids, thus the solutions to the assignment problem will fluctuate between assigning the slot to advertiser 1 and 2 depending which is greater of λ_1, λ_2 .

Essentially the difficulty occurs because the type space in this example is discrete. The search engine does not have enough additional information from the search type τ to fine tune its discrimination between the two advertisers. However, under our assumption that the distribution \mathbb{P}_τ over \mathcal{T} admits a continuous bounded probability density function, the allocation of clicks $y_i(\lambda)$ is a continuous function of prices and we shall see that, under models of advertiser response, we are able to deduce convergence towards a system optimum.

Recall that $(x_{il}^\tau(\lambda) : i \in \mathcal{I}, l \in \mathcal{L})$ defines an optimal solution to the assignment problem $\text{ASSIGNMENT}(\tau, \lambda)$.

PROPOSITION 5.2. *Under the assumption that the distribution \mathbb{P}_τ over \mathcal{T} admits a continuous bounded probability density function then both*

$$x_{il}(\lambda) := \mathbb{E}_\tau x_{il}^\tau(\lambda), \quad y_i(\lambda) := \mathbb{E}_\tau \sum_l p_{il}^\tau x_{il}^\tau(\lambda)$$

are differentiable functions of λ for $\lambda > 0$. Moreover, both functions $x(\lambda)$ and $y(\lambda)$ are Lipschitz continuous on any set where λ is bounded away from zero.

The proof, given in Appendix C, is somewhat technical. However, from it we can conclude that if click-through probabilities vary sufficiently with τ , then allocation rates $x_{il}(\lambda)$ and click-through rates $y_i(\lambda)$ vary smoothly with λ .

Consider the objective function for the dual of the system problem as derived in Proposition 3.2,

$$\mathcal{V}(\lambda) = \sum_{i \in \mathcal{I}} (U_i^*(\lambda_i) + \lambda_i y_i(\lambda)). \quad (26)$$

This expression is the sum of the consumer surpluses and the revenue achieved by the platform at prices λ and, when λ is optimal, it is equal to the maximal total welfare as

defined by the SYSTEM problem (3). We note that

$$\frac{\partial \mathcal{V}}{\partial \lambda_i} = -D_i(\lambda_i) + y_i(\lambda).$$

This holds by the Envelope Theorem and is argued in Lemma C.4.

We next model advertisers' responses to their observation of click-through rates. We suppose advertiser i changes its price $\lambda_i(t)$ smoothly (i.e. continuously and differentially) as a consequence of its observation of its current click-through rate $y_i(t)$ so that

$$\frac{d}{dt} \lambda_i(t) \geq 0 \text{ according as } \lambda_i(t) \leq U'_i(y_i(t)).$$

This is a natural dynamical system representation of advertiser i varying λ_i smoothly in order to track the optimum of her return $r_i(\lambda)$ under prices (15), from (24 - 25).

THEOREM 5.3 (CONVERGENCE OF DYNAMICS). *Starting from any point $\lambda(0)$ in the interior of the positive orthant, the trajectory $(\lambda(t) : t \geq 0)$ of the above dynamical system converges to a solution of the dual of the SYSTEM optimization, (8). Thus $y(\lambda(t))$, the assignment achieved by the prices $\lambda(t)$, converges to a solution of the SYSTEM optimization.*

PROOF. We prove that the objective of the dual problem $\mathcal{V}(\lambda)$, defined above, is a Lyapunov function for the dynamical system. Note that $\mathcal{V}(\lambda)$ is continuously differentiable for λ strictly positive. Since $D_i(\lambda_i) \rightarrow \infty$ as $\lambda_i \rightarrow 0$ and $y_i(\lambda)$ is bounded, there exist $\delta > 0$ such that for all $y_i \leq \delta$

$$\frac{d}{dt} \lambda_i(t) < 0.$$

We deduce that the paths of our dynamical system $(\lambda(t) : t \geq 0)$ are strictly positive and $\mathcal{V}(\lambda)$ is continuously differentiable on these paths. Further, the level sets $\{\lambda : \mathcal{V}(\lambda) \leq \kappa\}$ are compact: this is an immediate consequence of the facts that the functions $U_i^*(\lambda_i)$ are positive and decreasing, and, as proven in Lemma C.4, that

$$\lim_{\|\lambda\| \rightarrow \infty} \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda) = \infty.$$

Observe that, from the definition of the demand function $D_i(\cdot)$,

$$y_i \leq D_i(\lambda_i) \text{ according as } \lambda_i \leq U'_i(y_i).$$

Differentiating $\mathcal{V}(\lambda(t))$ yields

$$\frac{d}{dt} \mathcal{V}(\lambda(t)) = \sum_{i \in \mathcal{I}} \frac{\partial \mathcal{V}}{\partial \lambda_i} \frac{d}{dt} \lambda_i(t) = - \sum_{i \in \mathcal{I}} (D_i(\lambda_i(t)) - y_i(\lambda(t))) \frac{d}{dt} \lambda_i(t) \leq 0,$$

where the inequality is strict unless $D_i(\lambda_i(t)) = y_i(\lambda_i(t))$ for $i \in \mathcal{I}$. Now recall $\frac{dU_i^*}{d\lambda_i} = -D_i(\lambda_i)$. By Lyapunov's Stability Theorem, see [Khalil 2002, Theorem 4.1], the process $(\lambda(t) : t \geq 0)$ converges to the set of points λ^* satisfying, for $i \in \mathcal{I}$,

$$\frac{dU_i^*(\lambda_i^*)}{d\lambda_i} + y_i(\lambda^*) = 0.$$

Thus, by Proposition 3.2a), the price process $\lambda(t)$ converges to an optimal solution to the dual problem (8). Further, by Proposition 3.2c), we know that $y(\lambda^*)$ is optimal for the SYSTEM problem and, by Proposition 5.2, $y(\lambda)$ is a continuous function. Thus the process of click-through rates $y(\lambda(t))$ converges to an optimal solution for the SYSTEM problem. \square

In the above result we model advertisers that smoothly change their bids over time. However, we remark that other convergence mechanisms could be considered. For instance, since our dual optimization problem is convex, we can minimize the dual through a coordinate descent algorithm, where each component λ_i is sequentially minimized. Such an algorithm could correspond to advertisers sequentially maximizing over λ_i their reward function $r_i(\lambda_i, \lambda_{-i})$ as described by the Nash equilibrium (14).

6. EXTENSIONS

In this section we describe several straightforward extensions, intended to illustrate the flexibility and tractability of our framework.

6.1. Advertiser weightings

Suppose that advertiser i receives some information on the type of a click-through, and judges some types of click-through as more valuable than others. For example, an advertiser may prefer click-throughs that come from one geographical area rather than another or from one slot position rather than another if such click-throughs are more likely to convert into sales.

In particular, let's suppose that advertiser i assigns a weight w_{il}^τ to click-throughs of type τ from slot l , and has utility $U_i(y_i)$ where now

$$y_i = \mathbb{E}_\tau \left[\sum_{l \in \mathcal{L}} p_{il}^\tau w_{il}^\tau x_{il}^\tau \right].$$

We expect w_{il}^τ to be constant over regions of the type space \mathcal{T} : even though advertiser i receives some information on the type τ , the platform knows more. Further the platform uses this additional information to solve the revised assignment problem $\text{ASSIGNMENT}(\tau, \lambda, w)$ defined as problem (1) with the revised objective:

$$\text{Maximize} \quad \sum_{i \in \mathcal{I}} \lambda_i \sum_{l \in \mathcal{L}} p_{il}^\tau w_{il}^\tau x_{il}^\tau.$$

Assume that $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ is strictly increasing and continuously differentiable over $\lambda_i \geq 0$, for all values of $\lambda_{-i} = (\lambda_j : j \neq i, j \in \mathcal{I})$ in the interior of the positive orthant: then we are again able to prove Theorems 4.1 and 5.3, and the proofs are similar.

We give a further illustration of this extension. Suppose that the assignment problem (1) is run on a stream of searches that contain either or both of the keywords A and B, and the type τ contains information on this which is passed to the advertisers on a click-through; and suppose that some advertisers are interested in searches for keyword A, some in searches for keyword B, some in searches for either keyword, and some in searches for both keywords. Then various preferences of advertiser i can be expressed by allowing w_{il}^τ to depend on whether τ lies in $\{\text{A but not B}\}$, $\{\text{B but not A}\}$, $\{\text{A or B}\}$ or $\{\text{A and B}\}$, a partition into four of the type space \mathcal{T} .

6.2. More complex page layouts

The platform may wish to allow adverts of different sizes: for example, an advertiser may wish to offer an advert that occupies two adjacent slots. Or the platform may have a more complex set of possible page layouts than simply an ordered list of slots $1, 2, \dots, L$. Let $\sigma \in \mathcal{S}$ describe a possible layout of the adverts for advertisers $i \in \mathcal{I}$, and let $p_{i\sigma}^\tau$ be the probability of a click-through to advertiser i under layout σ . Then the

generalization of the assignment problem (1) becomes

$$\begin{array}{ll} \text{Maximize} & \sum_{i \in \mathcal{I}} \lambda_i p_{i\sigma}^\tau \\ \text{over} & \sigma \in \mathcal{S}. \end{array}$$

Indeed, this formulation allows the click-through probabilities for an advert to depend not just on the advertiser and the slot position, but also on which other adverts are shown on the page, provided only the probabilities $p_{i\sigma}^\tau$ can be estimated.

The complexity of this optimization problem depends on the design of the page layout through the structure of the set \mathcal{S} , and may depend on any structural information on the probabilities $p_{i\sigma}^\tau$, but for a variety of cases it will remain an assignment problem with an efficient solution. If $y_i(\lambda)$ is again defined as the expected click-through rate for advertiser i from a bid vector λ then Theorems 4.1 and 5.3 hold, with identical proofs.

6.3. Controlling the number of slots

The platform may wish to limit the number of slots filled, if it judges the available adverts as not sufficiently interesting to searchers. (Ultimately showing the wrong or poor quality ads can cause searchers to move platform and so hurt long-term platform revenue.)

Suppose the platform judges there is a benefit (positive or negative) q_{il}^τ to a searcher for an impression of the advert from advertiser i in slot l for a search of type τ , whether or not the user clicks through, so that the system objective function (3a) becomes

$$\sum_{i \in \mathcal{I}} U_i(y_i) + \mathbb{E}_\tau \left[\sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} q_{il}^\tau x_{il}^\tau \right].$$

Then the assignment objective function (1a) becomes

$$\sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} (\lambda_i p_{il}^\tau + q_{il}^\tau) x_{il}^\tau$$

and our results hold with minor amendments. In particular, equation (15) for the price function and equation (26) for the Lyapunov function are unaltered, although of course the functions $y_i(\lambda)$ will now be defined in terms of solutions to the new assignment problem.

An important special case is when $q_{il}^\tau \equiv -R$, where R is a *reserve price*. We next give an alternative interpretation of this case. Let the platform include in the assignment problem a collection of fictitious advertisers whose adverts are realised as empty slots, and for whom $\lambda_k = R, p_{kl}^\tau = 1$. Then a (non-fictitious) advert will be shown in a slot only if its expected contribution to the objective function of the assignment problem meets at least the reserve R .

Of course a reserve R may also have a favourable effect on the revenue received by the platform [Ostrovsky and Schwarz 2011; Bachrach et al. 2014]. As an illustration, consider the generalized English auction of Example 4.6. A reserve of R will reduce the number of slots filled if $R > \lambda_L p_L^\tau$, and may well increase the revenue. Nevertheless our framework is one of utility maximization: we assume the platform is trying to assure its long-term revenue by producing as much benefit as possible for its users, its advertisers and itself. There are, of course, several ways in which the platform could increase its own revenue within the utility maximization framework: in the absence of competition from other platforms, it could for example charge an advertiser a fixed fee, less than the advertiser's consumer surplus, to participate.

As yet a further example of the flexibility of the framework, instead of a fixed reserve price we could allow an organic search result k to compete for a slot, with a positive benefit q_{kl}^T , but with $\lambda_k = 0$.

6.4. Multivariate utility functions

In this subsection we suppose the platform divides the stream of search queries into several distinct streams, and runs separate auctions for each stream. For our exposition we suppose the distinction between streams is defined in terms of keywords, but it could involve additional or other characteristics of search queries observable to advertisers.

An advertiser may well be interested in several quite different keywords: for example, a manufacturer may be able to shift production from haute couture to casual clothes, products that are advertised to quite different audiences. This example suggests we need a utility function more general than considered so far.

Suppose that advertiser i 's utility $U_i(\cdot)$ is a strictly concave, continuously differentiable function of the vector $y_i = (y_{ik}, k \in \mathcal{K}_i)$, where \mathcal{K}_i is the set of keywords of interest to advertiser i and y_{ik} is the click-through rate to advertiser i from searches on the keyword k . Assume that the partial derivative $\partial U_i / \partial y_{ik}$ decreases from ∞ to 0 as y_{ik} increases from 0 to ∞ .

Let λ_{ik} be the bid of advertiser i for keyword k , and let $\lambda_i = (\lambda_{ik} : k \in \mathcal{K}_i)$ and $\lambda = (\lambda_{ik} : i \in \mathcal{I}, k \in \mathcal{K}_i)$. Let $\mathcal{K} = \cup_{i \in \mathcal{I}} \mathcal{K}_i$, the set of keywords, set $\lambda_{ik} = 0$ for $k \notin \mathcal{K}_i$. Let

$$U_i^*(\lambda_i) = \max_{y_i \geq 0} \left(U_i(y_i) - \sum_{k \in \mathcal{K}_i} \lambda_{ik} y_{ik} \right),$$

the Legendre-Fenchel transform of $U_i(y_i)$, interpretable as the consumer surplus of advertiser i at prices λ_i . Our conditions on U_i and its partial derivatives ensure there is a unique maximum, interior to the positive orthant, for any price vector λ_i in the positive orthant. Let $(D_{ik}(\lambda_i), k \in \mathcal{K}_i)$ be the argument y_i that attains this maximum: it is the demand vector of advertiser i at prices λ_i . Further

$$\frac{\partial}{\partial \lambda_{ik}} U_i^*(\lambda_i) = D_{ik}(\lambda_i). \quad (27)$$

We assume the platform runs separate assignment problems for each keyword, so that the auction for keyword $k, k \in \mathcal{K}$, depends on λ only through $\lambda_{*k} = (\lambda_{ik} : i \in \mathcal{I})$; and we assume the platform charges according to the price function (15), that is

$$\pi_{ik}(\lambda_{*k}) = \frac{1}{y_{ik}(\lambda_{*k})} \int_0^{\lambda_{ik}} \left(y_{ik}(\lambda_{*k}) - y_{ik}(\mu_{ik}; \lambda_{jk}, j \neq i) \right) d\mu_{ik},$$

where $(\mu_{ik}; \lambda_{jk}, j \neq i)$ is the vector obtained from the vector λ_{*k} by replacing the i th component, λ_{ik} , by μ_{ik} .

Then the question for advertiser i is how to balance her bids $(\lambda_{ik}, k \in \mathcal{K}_i)$ over the keywords \mathcal{K}_i that are of interest to her. The reward to advertiser i arising from a vector of bids $\lambda = (\lambda_i : i \in \mathcal{I}) = (\lambda_{ik} : i \in \mathcal{I}, k \in \mathcal{K}_i)$ is then

$$r_i(\lambda) = U_i(y_i(\lambda)) - \sum_{k \in \mathcal{K}_i} \pi_{ik}(\lambda_{*k}) y_{ik}(\lambda_{*k}), \quad (28)$$

and the condition for a Nash equilibrium is again (14) where now λ_i is a vector.

Assume, as usual, that $\lambda_{ik} \mapsto y_{ik}(\lambda_i, \lambda_{-i})$ is strictly increasing and continuously differentiable. Then from the form (28)

$$\frac{\partial}{\partial \lambda_{ik}} r_i(\lambda) = \left(\frac{\partial U_i}{\partial y_{ik}} - \lambda_{ik} \right) \frac{\partial y_{ik}}{\partial \lambda_{ik}}. \quad (29)$$

Thus there is a unique Nash equilibrium, identified by equating the bid λ_{ik} with advertiser i 's marginal utility for click-throughs on keyword k for each $i \in \mathcal{I}, k \in \mathcal{K}_i$. These conditions also identify the unique system optimum.

Next suppose that for each $k \in \mathcal{K}_i$ advertiser i changes her bid $\lambda_{ik}(t)$ smoothly (i.e. continuously and differentiably) as a consequence of her observation of her current click-through rate $y_{ik}(\lambda_i(t))$ so that

$$\frac{d}{dt} \lambda_{ik}(t) \geq 0 \text{ according as } y_{ik}(\lambda_i(t)) \leq D_{ik}(\lambda_i(t)). \quad (30)$$

This is a dynamical system representation of advertiser i varying λ_{ik} smoothly in order to increase or decrease her bid for keyword k according to whether the currently observed click-through rate $y_{ik}(t)$ seems too low or too high for her current bid. Then trajectories converge to the solution of the system problem, by essentially the same Lyapunov argument as used to prove Theorem 5.3, as we now sketch.

Let

$$\mathcal{V}(\lambda) = \sum_{i \in \mathcal{I}} U_i^*(\lambda_i) + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_i} \lambda_{ik} y_{ik}(\lambda).$$

Differentiating $\mathcal{V}(\lambda(t))$ yields, from (27), Proposition C.4 and (30),

$$\frac{d}{dt} \mathcal{V}(\lambda(t)) = \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_i} \frac{\partial \mathcal{V}}{\partial \lambda_{ik}} \frac{d}{dt} \lambda_{ik}(t) = - \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_i} (D_{ik}(\lambda_i(t)) - y_{ik}(\lambda_i(t))) \frac{d}{dt} \lambda_{ik}(t) \leq 0$$

where the inequality is strict unless $D_{ik}(\lambda_i(t)) = y_{ik}(\lambda_i(t))$ for $i \in \mathcal{I}, k \in \mathcal{K}_i$. But this holds if and only if y solves the system problem.

In view of the derivative (29) a possibly more natural dynamical system representation of advertiser i 's response would be that she changes her bid $\lambda_{ik}(t)$ smoothly as a consequence of her observation of her current click-through rate $y_{ik}(t)$ so that

$$\frac{d}{dt} \lambda_{ik}(t) \geq 0 \text{ according as } \lambda_{ik}(t) \leq \frac{\partial U_i}{\partial y_{ik}}(y_i(t)).$$

This could be viewed as a myopic attempt to improve the return (28) in the immediate future. In a single dimension, where the set \mathcal{K}_i is a singleton, this is equivalent to (30) by the concavity of U_i , but in higher dimensions this is not the same condition and global convergence of trajectories under this response is not assured. For small perturbations from the Nash equilibrium the conditions are equivalent.

6.5. Budget constraints

In this subsection we consider advertisers who have budget constraints on what they can spend across different types of search, for example, in an advertising campaign. We begin with an example.

Example 6.1. A simple approach to a budget constraint would be to use a utility function which directly captures the constraint within the framework of Section 6.4. We illustrate this as follows.

Suppose

$$U_i(y_i) = \frac{b_i}{q} \log \sum_{k \in \mathcal{K}_i} (w_{ik} y_{ik})^q$$

for $0 < q < 1$. Then

$$\frac{\partial U_i}{\partial y_{ik}} = \frac{b_i w_{ik}^q y_{ik}^{q-1}}{\sum_{j \in \mathcal{K}_i} (w_{ij} y_{ij})^q}$$

and so at the unique Nash equilibrium found in Section 6.4, where $\partial U_i / \partial y_{ik} = \lambda_{ik}$, the budget constraint

$$\sum_{k \in \mathcal{K}_i} \lambda_{ik} y_{ik} = b_i$$

is automatically satisfied, provided the constraint is on the rate of bidding rather than expenditure – not taking into account rebates.

We require $q < 1$ to ensure the strict concavity of $U_i(\cdot)$. As $q \rightarrow 1$, maximizing $U_i(y_i)$ subject to the budget constraint becomes equivalent to maximizing $\sum_{k \in \mathcal{K}_i} w_{ik} y_{ik}$ subject to the same budget constraint, and we recover the early model for the equilibrium price of goods for buyers with linear utilities [Fisher 1892; Eisenberg and Gale 1959]. The allocations y_{ik} may not be unique when $q = 1$, which complicates discussions of convergence.

Next we consider more general utilities, and suppose that advertiser $i \in \mathcal{I}_B \subset \mathcal{I}$ has a scalar constraint b_i on her rate of bidding. Consider the following optimization problem:

$$\begin{aligned} & \text{Maximize} && U_i(y_i) - \sum_{k \in \mathcal{K}_i} \lambda_{ik} y_{ik} \\ & \text{subject to} && \sum_{k \in \mathcal{K}_i} \lambda_{ik} y_{ik} \leq b_i, \\ & \text{over} && y_{ik} \geq 0, \quad k \in \mathcal{K}_i. \end{aligned}$$

The Lagrangian for this problem is

$$\begin{aligned} L(y_i, z_i; \lambda_i, \mu_i) &= U_i(y_i) - \lambda_i \cdot y_i + \mu_i (b_i - \lambda_i \cdot y_i - z_i) \\ &= U_i(y_i) - (1 + \mu_i) \lambda_i \cdot y_i + \mu_i b_i - \mu_i z_i \end{aligned}$$

where μ_i, b_i, z_i are scalars and λ_i, y_i are vectors. This is straightforwardly maximized over $z_i \geq 0$ by $\mu_i z_i = 0$ provided $\mu_i \geq 0$, and by

$$y_{ik} = D_{ik}((1 + \mu_i) \lambda_i). \quad (31)$$

Note that if μ_i is positive then the budget constraint is tight and advertiser i 's demand is reduced, via the form (31), to meet the constraint. Such a choice of μ_i is possible, and is unique, since $\sum_{k \in \mathcal{K}_i} \lambda_{ik} D_{ik}((1 + \mu_i) \lambda_i)$ is a strictly decreasing and continuous function of μ_i approaching zero as $\mu_i \rightarrow \infty$.

Let

$$U_i^*(\lambda_i, \mu_i) = \max_{y_i, z_i \geq 0} L(y_i, z_i; \lambda_i, \mu_i).$$

Then, by the envelope theorem,

$$\frac{\partial}{\partial \lambda_{ik}} U_i^*(\lambda_i, \mu_i) = -(1 + \mu_i) D_{ik}((1 + \mu_i) \lambda_i), \quad (32a)$$

$$\frac{\partial}{\partial \mu_i} U_i^*(\lambda_i, \mu_i) = b_i - \sum_{k \in \mathcal{K}_i} \lambda_{ik} D_{ik}((1 + \mu_i) \lambda_i). \quad (32b)$$

Next we consider a dynamical system where advertiser i smoothly varies her bids $\lambda_i(t)$ and a further parameter $\mu_i(t)$ as a function of her realised rates $y_i(t)$, in an attempt to solve equation (31) and to satisfy her budget constraint. Suppose for $i \in \mathcal{I}_B$

$$\frac{d}{dt} \lambda_{ik}(t) \geq 0 \text{ according as } y_{ik}(\lambda_i(t)) \leq D_{ik}((1 + \mu_i(t))\lambda_i(t)) \quad (33a)$$

$$\frac{d}{dt} \mu_i(t) \geq 0 \text{ according as } \sum_{k \in \mathcal{K}_i} \lambda_{ik}(t) (y_{ik}(\lambda_i(t)) - D_{ik}((1 + \mu_i(t))\lambda_i(t))) \leq 0, \quad (33b)$$

while for $i \in \mathcal{I} \setminus \mathcal{I}_B$ relation (30) holds. Let $\mu = (\mu_i, i \in \mathcal{I}_B)$ and define

$$\begin{aligned} \mathcal{V}(\lambda, \mu) = & \sum_{i \in \mathcal{I}_B} \left(U_i^*(\lambda_i, \mu_i) + (1 + \mu_i) \sum_{k \in \mathcal{K}_i} \lambda_{ik} y_{ik}(\lambda) - \mu_i b_i \right) \\ & + \sum_{i \in \mathcal{I} \setminus \mathcal{I}_B} \left(U_i^*(\lambda_i) + \sum_{k \in \mathcal{K}_i} \lambda_{ik} y_{ik}(\lambda) \right). \end{aligned}$$

Differentiating $\mathcal{V}(\lambda(t), \mu(t))$ yields, from Proposition C.4 and from either (32) and (33) or (27) and (30),

$$\frac{d}{dt} \mathcal{V}(\lambda(t), \mu(t)) = \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_i} \frac{\partial \mathcal{V}}{\partial \lambda_{ik}} \frac{d}{dt} \lambda_{ik}(t) + \sum_{i \in \mathcal{I}_B} \frac{\partial \mathcal{V}}{\partial \mu_i} \frac{d}{dt} \mu_i(t) \leq 0,$$

where the inequality is strict except at the unique minimum of $\mathcal{V}(\lambda, \mu)$.

In this subsection we have assumed budgets constrain the advertisers' rates of bidding rather than expenditure. We might assume that the existence of rebates broadly encourages truthful behaviour, and that rebates in one time period allow a larger budget in the next time period.

7. CONCLUDING REMARKS

The framework we describe attempts to capture the system architecture of Ad-auctions. The assignment problem must be solved rapidly, for each search; while an advertiser is primarily interested in aggregates over longer periods of time. Thus we model in detail each random instance of the assignment problem, while we describe an advertiser's behaviour in terms of averages evolving in time. The platform knows more about search types and thus more about click-through probabilities, while an advertiser knows more about the value to her of additional click-throughs and is incentivised to communicate this information via her bids. On a slow time-scale the platform may decide which search types to pool in distinct auctions, across which the advertisers will have different preferences they are able to communicate.

Our formal framework is one of utility maximization, as in the seminal work of Vickrey (1961), and the form of charging we describe to induce truthful declarations from advertisers is familiar from that paper. Within our framework it is possible to study the platform's revenue as a function of, for example, reserve price; and one special case of our framework recovers an equilibrium of the generalized second price auction with reserve familiar from studies of revenue maximization [Ostrovsky and Schwarz 2011].

We have used sponsored search auctions as the motivation, and our model reflects current practice in sponsored search, where platforms such as BingAds or Google Adwords use a variant of the generalized second price auction to solve the assignment problem for every search query, while advertisers alter bids on timescales measured in hours or days. However, our results apply much more widely, to display ads and other online settings where time-scale asymmetry coexists with information asymmetry.

Under the assumption of strategic advertisers, we showed that a Nash equilibrium exists for the advertisers which produces the system optimum, provided prices are set appropriately. We gave a simple way to implement such prices: namely, by giving advertisers a rebate, constructed by solving a second assignment problem using uniform sampling between zero and the submitted bid for advertisers rather than the bid itself. Note that the solution to this second assignment problem is not used for the allocation but only for pricing, and hence could be adapted for use in current Ad-auctions.

REFERENCES

- ATHEY, S. AND NEKIPELOV, D. 2012. A structural model of sponsored search advertising auctions. Working paper.
- BABAIOFF, M., KLEINBERG, R. D., AND SLIVKINS, A. 2010. Truthful mechanisms with implicit payment computation. In *ACM Conference on Electronic Commerce*. 43–52.
- BACHRACH, Y., CEPPI, S., KASH, I. A., KEY, P., AND KUROKAWA, D. 2014. Optimising trade-offs among stakeholders in ad auctions. In *Proc. of the Fifteenth ACM Conf. on Electronic Commerce*. EC '14. ACM.
- BECKMANN, M., MCGUIRE, C., AND WINSTEN, C. B. 1956. *Studies in the Economics of Transportation*. Yale Univ. Press, New Haven.
- BERTSEKAS, D. P. 1988. The auction algorithm: A distributed relaxation method for the assignment problem. *Annals of operations research* 14, 1, 105–123.
- BIRNBAUM, B., DEVANUR, N. R., AND XIAO, L. 2011. New convex programs and distributed algorithms for fisher markets with linear and spending constraint utilities.
- BORWEIN, J. AND LEWIS, A. 2006. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. CMS books in mathematics. Springer-Verlag New York.
- 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.
- EISENBERG, E. AND GALE, D. 1959. Consensus of subjective probabilities: the pari-mutuel method. *The Annals of Mathematical Statistics* 30, 165–168.
- FISHER, I. 1892. *Mathematical investigations in the theory of value and prices*.
- JAIN, K. AND VAZIRANI, V. V. 2007. Eisenberg-gale markets: algorithms and structural properties. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. STOC '07. ACM, New York, NY, USA, 364–373.
- JOHARI, R. AND TSITSIKLIS, J. N. 2009. Efficiency of scalar-parameterized mechanisms. *Oper. Res.* 57, 823–839.
- KELLY, F. 1997. Charging and rate control for elastic traffic. *European Transactions on Telecommunications* 8, 33–37.
- KELLY, F. AND YUDOVINA, E. 2014. *Stochastic Networks*. Cambridge University Press.
- KHALIL, H. 2002. *Nonlinear Systems*. Prentice Hall PTR.
- KUHN, H. W. 1955. The hungarian method for the assignment problem. *Naval research logistics quarterly* 2, 1-2, 83–97.
- MILGROM, P. 2004. *Putting Auction Theory to Work*. Churchill Lectures in Economics. Cambridge University Press.
- 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 *ACM Conference on Electronic Commerce*. 59–60.
- PIN, F. AND KEY, P. 2011. Stochastic variability in sponsored search auctions. In *Proc. 12th ACM conference on Electronic Commerce*. EC'11. San Jose, California, 61–70.

- 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. ACM, 751–766.
- ROCKAFELLAR, R. 1997. *Convex Analysis*. Convex Analysis. Princeton University Press.
- ROCKAFELLAR, R. T. 1984. *Network Flows and Monotropic Optimization*. Wiley.
- SRIKANT, R. 2004. *The Mathematics of Internet Congestion Control*. Birkhauser.
- SRIKANT, R. AND TAN, B. 2010. Online advertisement, optimization and stochastic networks. *CoRR abs/1009.0870*.
- SYRGKANIS, V. AND TARDOS, E. 2012. Bayesian sequential auctions. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM, 929–944.
- VARIAN, H. R. 2007. Position auctions. *International Journal of Industrial Organization* 25, 6, 1163–1178.
- VAZIRANI, V. V. 2010. Spending constraint utilities with applications to the adwords market. *Math. Oper. Res.* 35, 458–478.
- VICKREY, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance* 16, 1, 8–37.
- VOHRA, R. 2011. *Mechanism Design: A Linear Programming Approach*. Econometric Society Monographs. Cambridge University Press.
- WARDROP, J. G. 1952. Road paper. some theoretical aspects of road traffic research. In *ICE Proceedings: Engineering Divisions*. Vol. 1. Thomas Telford, 325–362.
- WILLIAMS, D. 1991. *Probability with Martingales*. Cambridge University Press.
- YANG, S. AND HAJEK, B. 2006. Revenue and stability of a mechanism for efficient allocation of a divisible good.
- YANG, S. AND HAJEK, B. 2007. VCG-Kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals. *IEEE journal on selected areas in communications* 25, 1237 – 1243.

A. PROOF OF PROPOSITIONS 3.1 AND 3.2

PROOF OF PROPOSITION 3.1. A Lagrangian of the system problem can be written as follows

$$L_{sys}(x, y; \lambda) = \sum_{i \in \mathcal{I}} U_i(y_i) + \sum_{i \in \mathcal{I}} \lambda_i \mathbb{E}_\tau \left[\sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau - y_i \right]. \quad (34)$$

Notice, we intentionally decide not to include the scheduling constraints in our Lagrangian. Thus we must maximize subject to these constraints when optimizing our Lagrangian (3c-3d). Let \mathcal{A} be the set variables $x = (x_{il}^\tau : i \in \mathcal{I}, l \in \mathcal{L}, \tau \in \mathcal{T}) \geq 0$ satisfying the assignment constraints, (3c-3d) for all $i \in \mathcal{I}, l \in \mathcal{L}$ and $\tau \in \mathcal{T}$. We see that our Lagrangian problem is separable in the following sense

$$\max_{\substack{x \in \mathcal{A} \\ y, z \in \mathbb{R}_+^{\mathcal{I}}}} L_{sys}(x, y; \lambda) = \sum_{i \in \mathcal{I}} \max_{y_i \geq 0} \{U_i(y_i) - \lambda_i y_i\} \quad (35a)$$

$$+ \mathbb{E}_\tau \left[\max_{x^\tau \in \mathcal{X}} \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i p_{il}^\tau x_{il}^\tau \right] \quad (35b)$$

Here \mathcal{X} denotes the set of $x \in \mathbb{R}_+^{\mathcal{I} \times \mathcal{L}}$ such that for each $i \in \mathcal{I}$ and $l \in \mathcal{L}$

$$\sum_{l' \in \mathcal{L}} x_{il'} \leq 1, \text{ and } \sum_{i' \in \mathcal{I}} x_{i'l} \leq 1. \quad (36)$$

As we will now discuss, we can see that solutions \tilde{x}, \tilde{y} and $\tilde{\lambda}$ satisfying the Conditions A and B of our Theorem are optimal for Lagrangian (35a) and (35b) when $\lambda = \tilde{\lambda}$.

Firstly, if $\tilde{\lambda}_i$ a solution to for the optimization

$$\text{minimizes } U_i^*(\lambda_i) + \lambda_i \tilde{y}_i \text{ over } \lambda_i \geq 0, \quad (37)$$

then the solution is achieved when $D_i(\tilde{\lambda}_i) = \tilde{y}_i$ or equivalently when $U_i'(\tilde{y}_i) = \lambda_i$. Thus it is clear that \tilde{y}_i solves the optimization

$$\max_{y_i \geq 0} \{U_i(y_i) - \tilde{\lambda}_i y_i\}. \quad (38)$$

Thus, if Condition A, is satisfied, then \tilde{y}_i optimizes (35a) when we choose $\lambda_i = \tilde{\lambda}_i$.

Secondly if \tilde{x}^τ solves ASSIGNMENT($\tau, \tilde{\lambda}$) for each τ , since each maximization expressed inside the expectation (35b) is an assignment problem, (35b) is maximized when we take $\lambda = \tilde{\lambda}$.

These two conditions, Condition A and B, show that the Lagrangian, (34), is maximized by \tilde{x} and \tilde{y} with Lagrange multipliers $\tilde{\lambda}$. In addition \tilde{x} and \tilde{y} are feasible for the system optimization (3). So, we have a feasible optimal solution for this Lagrangian problem. As we demonstrate in Proposition B.1 in the appendix, the Lagrangian sufficiency still holds for the system problem (3) – even though it has an infinite number of constraints. Thus, we have shown a solution to conditions A and B is optimal for the system problem.

Conversely, we know that strong duality holds for the system optimization (3) – once again, this is despite the infinite number of constraints for this optimization. See Theorem B.2 in the appendix for a proof. In otherwords, there exists a vector $\tilde{\lambda}$ such that an optimal solution to the system problem is also an optimal solution to the Lagrangian problem when we chose lagrange multipliers $\tilde{\lambda}$. Thus, an optimal solution to the SYSTEM($U, \mathcal{I}, \mathbb{P}_\tau$) problem must optimize (35a) and (35b), and as discussed these

solutions correspond to Conditions A and B. In otherwords, an optimal solution to the system problem satisfies Conditions A and B with this choice of $\tilde{\lambda}$. \square

PROOF OF PROPOSITION 3.2. a) As found in Theorem 3.1, the Lagrangian of the system problem can be written as follows

$$L_{sys}(x, y; \lambda) = \sum_{i \in \mathcal{I}} U_i(y_i) + \sum_{i \in \mathcal{I}} \lambda_i \mathbb{E}_\tau \left[\sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau - y_i \right]. \quad (39)$$

As we recall from (35), this Lagrangian is separable and is maximized as

$$\begin{aligned} \max_{\substack{x \in \mathcal{A} \\ y, z \in \mathbb{R}_+^{\mathcal{I}}}} L_{sys}(x, y; \lambda) &= \sum_{i \in \mathcal{I}} \max_{y_i \geq 0} \{U_i(y_i) - \lambda_i y_i\} \\ &\quad + \mathbb{E}_\tau \left[\max_{x^\tau \in \mathcal{X}} \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i p_{il}^\tau x_{il}^\tau \right] \\ &= \sum_{i \in \mathcal{I}} U_i^*(\lambda_i) + \mathbb{E}_\tau \left[\sum_{i \in \mathcal{I}} \lambda_i \sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau(\lambda) \right] = \sum_{i \in \mathcal{I}} U_i^*(\lambda_i) + \lambda_i y_i(\lambda). \end{aligned}$$

In the second equality above, we rearrange the assignment optimization in terms of the click-through rate of each advertiser, $y_i(\lambda)$.

Thus the dual of this optimization problem is as required:

$$\text{Minimize } \sum_{i \in \mathcal{I}} [U_i^*(\lambda_i) + \lambda_i y_i(\lambda)] \quad \text{over } \lambda_i \geq 0, \quad i \in \mathcal{I}.$$

Now analyze the objective of this dual problem. We first show that optimization (8) is minimized when $0 < \lambda_i < \infty$ for each $i \in \mathcal{I}$. We consider the function

$$\sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda). \quad (40)$$

With technical lemma, Lemma C.4, we see that we this function is continuously differentiable with i th partial derivative given by the continuous function $y_i(\lambda)$. In addition, by definition $D_i(\lambda) = (U_i^*)'(\lambda) = (U_i')^{-1}(\lambda)$. Thus the objective of (8) is continuously differentiable for $\lambda > 0$.

Further it is positive for $\lambda \neq 0$ and increases by a constant factor when we multiply λ_i by a constant. Thus,

$$\lim_{\|\lambda\| \rightarrow \infty} \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda) = \infty. \quad (41)$$

Thus since $U_i^*(\lambda_i)$ is a positive function, we see that the dual minimization (8) must be achieved by a finite solution λ^* . Since in expression (7) the surplus demand satisfies $D_i(0) = \infty$, the minimum of the dual problem (8) must be achieved by $\lambda_i^* > 0$ for each $i \in \mathcal{I}$. Now, as objective of (8) is continuously differentiable for λ strictly positive, it is minimized iff for each $i \in \mathcal{I}$

$$\frac{dU_i^*}{d\lambda_i}(\lambda_i^*) + y_i(\lambda^*) = 0. \quad (42)$$

b) For the Lagrangian for the system problem, (39), Strong Duality holds by Theorem B.2. So, there exist Lagrange multipliers λ^* , such that

$$\sum_{i \in \mathcal{I}} [U_i^*(\lambda_i^*) + \lambda_i^* y_i(\lambda^*)] = \max_{\substack{x \in \mathcal{A} \\ y, z \in \mathbb{R}_+^{\mathcal{I}}}} L_{sys}(x, y, z; \lambda^*) = \max_{\substack{x \in \mathcal{A} \\ y, z \in \mathbb{R}_+^{\mathcal{I}}}} \sum_{i \in \mathcal{I}} U_i(y_i)$$

where there are feasible vectors x^*, y^*, z^* achieving the optimum of both maximizations above. By weak duality it is clear that λ^* must be optimal for the dual problem (8). Further, since x^* optimizes the Lagrangian L_{sys} with Lagrange multipliers λ^* , it solves the assignment problem, $x^{*\tau} = x^\tau(\lambda^*)$. \square

B. LAGRANGIAN OPTIMIZATION

In this paper, we consider optimization problems that have a potentially infinite number of constraints, in particular, for the system wide optimization (3). Thus it is not immediately clear that the Lagrangian approach – ordinarily applied with a finite number of constraints – immediately applies to our setting. We demonstrate that certain principle results, namely weak duality and the Lagrangian Sufficiency Theorem, apply to our setting.

We consider an optimization of the form

$$\text{Maximize} \quad g(y) \quad (43a)$$

$$\text{subject to} \quad y_i \leq \mathbb{E}_\mu[x_i], \quad i = 1, \dots, n, \quad (43b)$$

$$f_j(x(\tau)) \leq c_j, \quad \tau \in \mathcal{T}, \quad j = 1, \dots, m, \quad (43c)$$

$$\text{over} \quad y \in \mathbb{R}^n, \quad x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n). \quad (43d)$$

In the above optimization, we consider probability space $(\mathcal{T}, \mathbb{P}_\mu)$ and measurable random variable $x : \mathcal{T} \rightarrow \mathbb{R}^n$. We let $\mathcal{B}(\mathcal{T}, \mathbb{R}^n)$ index the set of Borel measurable functions from \mathcal{T} to \mathbb{R}^n . We assume that $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a concave function and that $f_j : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function, for each $j = 1, \dots, d$. We assume the solution to this optimization is bounded above.

Although there are an infinite number of constraints in this optimization, we can define a Lagrangian for this optimization as follows

$$L(x, y, z; \lambda) = g(y) + \sum_{i=1}^n \lambda_i \mathbb{E}_\mu[x_i - y_i - z_i]$$

Here the Lagrange multipliers λ_i , $i = 1, \dots, n$, can be assumed to be positive, slack variables z_i are added for each constraint (43b) and the optimization of the Lagrangian is taken over y_i real, z_i positive and real, and x_i a Borel measurable random variable for $i \in \mathcal{I}$. We let \mathcal{F} be the set of (x, y) feasible for the optimization (43).

Weak duality and Lagrangian Sufficiency both hold for this Lagrangian problem.

PROPOSITION B.1 (WEAK DUALITY).

a) [Weak Duality] For g^* the optimal value of the optimization (43),

$$\sup_{\substack{y \in \mathbb{R}^n, \\ x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)}} L(x, y, z; \lambda) \geq g^*.$$

b) [Lagrangian Sufficiency] If there exists $x^* \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)$ and $y^* \in \mathbb{R}^n$ that is both feasible for the optimization (43) and maximizes the Lagrangian $L(x, y, z; \lambda)$ with $z_i^* := y_i^* - \mathbb{E}_\mu x_i^*$ then x^*, y^*, z^* is optimal for (43).

PROOF. a) Because \mathcal{F} is a subset of $\mathcal{B}(\mathcal{T}, \mathbb{R}^n) \times \mathbb{R}^n$, we have

$$\sup_{\substack{y \in \mathbb{R}^n, z \in \mathbb{R}_+^n, \\ x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)}} L(x, y, z; \lambda) \geq \sup_{\substack{(x, y) \in \mathcal{F} \\ z \in \mathbb{R}_+^n}} L(x, y, z; \lambda) = \sup_{\substack{(x, y) \in \mathcal{F} \\ z \in \mathbb{R}_+^n}} L(x, y, z; \lambda) = g^*.$$

This proves weak duality.

b) Now applying this inequality, if a feasible solution optimizes the Lagrangian

$$g(y^*) = L(x^*, y^*, z^*; \lambda, \alpha) = \sup_{\substack{y \in \mathbb{R}^n, z \in \mathbb{R}_+^n, \\ x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)}} L(x, y, z; \lambda) \geq g^*. \quad (44)$$

Thus, (x^*, y^*) is optimal for (43). \square

For $z \in \mathbb{R}^{\mathcal{I}}$, we use $\mathcal{F}(z)$ to denote the set of (x, y) satisfying constraints (43c-43d) and satisfying constraints

$$z_i + y_i \leq \mathbb{E}_\mu[x_i], \quad i = 1, \dots, n. \quad (45)$$

Note, $\mathcal{F} = \mathcal{F}(0)$. We now show that there exists a Lagrange multiplier λ^* where the optimized Lagrangian function also optimizes (43).

THEOREM B.2 (STRONG DUALITY). *There exists a $\lambda^* \in \mathbb{R}_+^n$ such that*

$$\max_{(x,y) \in \mathcal{F}} g(y) = \max_{\substack{y \in \mathbb{R}^n \\ x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)}} g(y) + \sum_{i \in \mathcal{I}} \lambda_i^* \mathbb{E}[x_i - y_i]. \quad (46)$$

In particular, if there exist $(x^, y^*) \in \mathcal{F}$ maximizing (43) then it maximizes (46).*

PROOF. Firstly, since $\mathcal{F} \subset \mathcal{B}(\mathcal{T}, \mathbb{R}^n) \times \mathbb{R}^n$, we proved the weak duality expression

$$\max_{(x,y) \in \mathcal{F}} g(y) = \max_{(x,y) \in \mathcal{F}} g(y) + \sum_{i \in \mathcal{I}} \lambda_i^* \mathbb{E}[x_i - y_i] \leq \max_{\substack{y \in \mathbb{R}^{\mathcal{I}} \\ x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)}} g(y) + \sum_{i \in \mathcal{I}} \lambda_i^* \mathbb{E}[x_i - y_i]. \quad (47)$$

It remains to show the reverse inequality. We consider the following set

$$\mathcal{C} = \{(z, \gamma) \in \mathbb{R}^{\mathcal{I}} \times \mathbb{R} : \text{there exists } (x, y) \in \mathcal{F}(z) \text{ with } g(y) \geq \gamma\}.$$

We claim that \mathcal{C} is convex. Take $(z^0, \gamma^0), (z^1, \gamma^1) \in \mathcal{C}$ and take $(x^0, y^0) \in \mathcal{F}(z^0), (x^1, y^1) \in \mathcal{F}(z^1)$ respectively achieving bounds $g(y^0) \geq \gamma^0$ and $g(y^1) \geq \gamma^1$. For each term $u = x, y, z, \gamma$ just defined, we correspondingly define $u^q = (1 - q)u^0 + qu^1$, for $q \in [0, 1]$.

By concavity of g , convexity of f_j , $j = 1, \dots, m$, and linearity, we have

$$\begin{aligned} g(y^q) &\geq (1 - q)g(y^0) + qg(y^1) \geq \gamma^q, \\ f_j(x^q(\tau)) &\leq (1 - q)f_j(x^0(\tau)) + qf_j(x^1(\tau)) \leq c_j, \\ \mathbb{E}_\mu[x_i^q - y_i^q] &\geq (1 - q)z_i^0 + qz_i^1 = z_i^q, \end{aligned}$$

for $\tau \in \mathcal{T}, j = 1, \dots, m$ and $i = 1, \dots, n$. These above inequalities show that $(z^q, \gamma^q) \in \mathcal{C}$ and thus our claim is holds: \mathcal{C} is convex.

Let $\gamma^* = \max_{(x,y) \in \mathcal{F}} g(y)$. Here we are optimizing over $\mathcal{F}(z)$ with $z = 0$. So, it is clear that $(0, \gamma^*)$ does not belong to the interior of \mathcal{C} . Thus by the Supporting Hyperplane Theorem [Rockafellar 1997], there exists a hyperplane through $(0, \gamma^*)$ supporting \mathcal{C} . In other words, there exists a non-zero vector $(\lambda, \phi) \in \mathbb{R}^{\mathcal{I}} \times \mathbb{R}$ such that

$$\phi \gamma^* \geq \phi \gamma + \lambda^{\top} z,$$

for all $(z, \gamma) \in \mathcal{C}$. Firstly, it is clear that $\phi \geq 0$, otherwise γ^* is not maximal for $(x, y) \in \mathcal{F}$.

We now claim $\phi \neq 0$. We proceed by contradiction. If $\phi = 0$, then $0 \geq \lambda^{\top} z$ for all $(z, \gamma) \in \mathcal{C}$. But notice, for any $x \in \mathcal{B}(\mathcal{T}, \mathbb{R}_+^n)$, we can choose $y_i \in \mathbb{R}$ such that $y_i - \mathbb{E}_\mu[x_i] = \lambda_i$, thus for this choice of (x, y) we have $z = \lambda$. Thus, $\lambda^{\top} z = \lambda^{\top} \lambda > 0$, and so we have a contradiction. It must be that $\phi > 0$.

As $\phi > 0$, we can define $\lambda^* = (\lambda_i/\phi : i \in \mathcal{I})$. Since for each $(x, y) \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n) \times \mathbb{R}^{\mathcal{I}}$, if we set $z_i = \mathbb{E}_\mu [x_i - y_i]$ and $\gamma = g(y)$ then we have $(z, \gamma) \in \mathcal{C}$. With this we have

$$\max_{(x', y') \in \mathcal{F}} g(y') = \gamma^* \geq \gamma + \lambda^{*\top} z = g(y) + \sum_{i \in \mathcal{I}} \lambda_i^* \mathbb{E} [x_i - y_i]$$

Thus, maximizing over $x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)$ $y \in \mathbb{R}^{\mathcal{I}}$, we have

$$\max_{(x, y) \in \mathcal{F}} g(y) \geq \max_{\substack{y \in \mathbb{R}^{\mathcal{I}} \\ x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)}} g(y) + \sum_{i \in \mathcal{I}} \lambda_i^* \mathbb{E} [x_i - y_i]. \quad (48)$$

Together (47) and (48) give the required equality (46). In addition, given (43) has a finite optimum, inequality (48) can only hold when $\lambda^* \geq 0$.

Finally, if $(x^*, y^*) \in \mathcal{F}$ are optimal for (43) then equality (46) implies

$$g(y^*) \geq g(y^*) + \sum_{i \in \mathcal{I}} \lambda_i^* \mathbb{E} [x_i^* - y_i^*]. \quad (49)$$

However, the feasibility of (x^*, y^*) and positivity of λ^* implies

$$\sum_{i \in \mathcal{I}} \lambda_i^* \mathbb{E} [x_i^* - y_i^*] \geq 0. \quad (50)$$

So we see these two inequalities imply complementary slackness $\lambda_i^* \mathbb{E} [x_i^* - y_i^*] = 0$ and that $(x^*, y^*) \in \mathcal{F}$ maximizing (43) also maximizes (46). \square

C. ADDITIONAL LEMMAS AND PROPOSITIONS

This section gives proofs of Proposition 5.2 and Proposition C.4. These propositions characterize the continuity and differentiability of the functions

$$\bar{x}_{il}(\lambda) = \mathbb{E}_\tau x_{il}^{*\top}(\lambda), \quad y_i(\lambda) = \mathbb{E}_\tau \sum_l p_{il}^\top x_{il}^{*\top}(\lambda), \quad \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda).$$

Proposition 5.2 requires two technical lemmas, Lemma C.1 and Lemma C.2, which give the differentiability of a random point belonging to a polytope as we smoothly change the boundary condition. Proposition C.4 employs the Envelope Theorem [Milgrom 2004, Chap. 3], as is commonly applied in auction theory.

LEMMA C.1. *1) If U is a random vector uniformly distributed inside the unit sphere, $S_n = \{u \in \mathbb{R}^n : \|u\| \leq 1\}$, then there exist a constant K_1 such that for any two non-zero vectors $\lambda, \tilde{\lambda} \in \mathbb{R}^n \setminus \{0\}$*

$$\mathbb{P}(\lambda^\top U \geq 0 > \tilde{\lambda}^\top U) \leq \frac{K_1}{\|\lambda\| \wedge \|\tilde{\lambda}\|} \|\lambda - \tilde{\lambda}\|. \quad (51)$$

2) If X is a random variable with density f_X continuous on the interior of its support \mathcal{P} , a polytope $\mathcal{P} \subset [-1, 1]^n$. The function $\mathbb{P}(\mu_1^\top X \geq 0, \dots, \mu_k^\top X \geq 0)$ is Lipschitz continuous provided $\|\mu_1\|, \dots, \|\mu_k\|$ are bounded away from zero.

PROOF. 1) We give a geometric proof of the result. We assume, wlog, that $\|\lambda\| \geq \|\tilde{\lambda}\|$, and we let V_n be the volume of S . For every u satisfying $\lambda^\top u \geq 0 > \tilde{\lambda}^\top u$, there exists a $\theta \in [0, 1]$ such that $\lambda^\top u + \theta(\tilde{\lambda}^\top - \lambda^\top)u = 0$. Let λ_θ be the unit vector proportional to $\lambda + \theta(\tilde{\lambda} - \lambda)$. By continuity there exist a $\tilde{\theta}$ such the $\lambda_{\tilde{\theta}}^\top u = 0$. We note three facts: 1) Each cross section $\{u \in S : \lambda_\theta^\top u = 0\}$ has the same volume V_{n-1} in its \mathbb{R}^{n-1} subspace; 2) the

path $\mathcal{P} = \{\lambda_\theta : \theta \in [0, 1]\}$ is a circular path and thus has length bounded above by the terms

$$2\pi \left\| \frac{\lambda}{\|\lambda\|} - \frac{\tilde{\lambda}}{\|\tilde{\lambda}\|} \right\| \leq \frac{2\pi}{\|\lambda\|} \|\lambda - \tilde{\lambda}\|; \quad (52)$$

and, 3) $\{u \in S : \lambda^\top u \geq 0 > \tilde{\lambda}^\top u\} = \{u \in S : \lambda_\theta^\top u = 0, \theta \in [0, 1]\}$. Thus, we see we can bound the probability $\mathbb{P}(\lambda^\top U \geq 0 > \tilde{\lambda}^\top U)$ by the length of the path \mathcal{P} times the volume of cross sections $\{u \in S : \lambda_\theta^\top u = 0\}$. In other words,

$$\mathbb{P}(\lambda^\top U \geq 0 > \tilde{\lambda}^\top U) \leq \frac{2\pi V_{n-1}}{\|\lambda\|} \|\lambda - \tilde{\lambda}\|, \quad (53)$$

as required.

2) Let's first deal with the case where $k = 1$. Observe

$$\begin{aligned} & \left| \mathbb{P}(\mu^\top X \geq 0) - \mathbb{P}(\tilde{\mu}^\top X \geq 0) \right| \\ &= \left| \mathbb{P}(\mu^\top X \geq 0 > \tilde{\mu}^\top X) - \mathbb{P}(\tilde{\mu}^\top X \geq 0 > \mu^\top X) \right| \\ &\leq \mathbb{P}(\mu^\top X \geq 0 > \tilde{\mu}^\top X) + \mathbb{P}(\tilde{\mu}^\top X \geq 0 > \mu^\top X) \end{aligned} \quad (54)$$

Also since f has a density that is bounded by a constant, we can bound the above probabilities with uniform random variables:

$$\mathbb{P}(\mu^\top X \geq 0 > \tilde{\mu}^\top X) \leq K_2 \mathbb{P}(\mu^\top U \geq 0 > \tilde{\mu}^\top U)$$

for a constant K_2 and for U is a uniform random variable on the unit sphere in \mathbb{R}^n . Now applying part 1) of this Lemma

$$\begin{aligned} \mathbb{P}(\mu^\top X \geq 0 > \tilde{\mu}^\top X) &\leq \frac{K_1 K_2}{\|\mu\| \wedge \|\tilde{\mu}\|} \|\mu - \tilde{\mu}\| \\ &\leq \frac{K_1 K_2}{K_3} \|\mu - \tilde{\mu}\|. \end{aligned}$$

where K_3 is the constant by which μ and $\tilde{\mu}$ are bounded away from zero. Thus, applying this inequality to (54). We have that $\mathbb{P}(\mu^\top X \geq 0)$ is Lipschitz continuous. For the case $k \geq 1$, we know that from the $k = 1$ case that the function

$$\mathbb{P}(\mu_1^\top X \geq 0 | \mu_2^\top X \geq 0, \dots, \mu_k^\top X \geq 0) \quad (55)$$

is Lipschitz as a function of μ_1 . Here, if $\mathbb{P}(\mu_2^\top X \geq 0, \dots, \mu_k^\top X \geq 0) = 0$, we apply the convention that (55) is zero – which is certainly a Lipschitz function. Thus, since

$$\begin{aligned} & \mathbb{P}(\mu_1^\top X \geq 0, \mu_2^\top X \geq 0, \dots, \mu_k^\top X \geq 0) \\ &= \mathbb{P}(\mu_1^\top X \geq 0 | \mu_2^\top X \geq 0, \dots, \mu_k^\top X \geq 0) \\ & \quad \times \mathbb{P}(\mu_2^\top X \geq 0, \dots, \mu_k^\top X \geq 0). \end{aligned}$$

This function is Lipschitz continuous component-wise and thus must also be Lipschitz continuous. \square

LEMMA C.2. *Given $\mu \in (0, \infty)^n$, X is a random variable with density f_X continuous on the interior of its support \mathcal{P} , a polytope $\mathcal{P} \subset [-1, 1]^n$.*

1) *Provided the plane $\mu^\top x = 0$ is not a boundary of this polytopes then $\mathbb{P}(\mu X \geq 0)$ is a*

differentiable function of μ .

2) Given distinct vectors $\mu_1, \dots, \mu_k \in (0, \infty)$ satisfy

$$\mathbb{P}(\mu_1^T X \geq 0, \dots, \mu_k^T X \geq 0) > 0, \quad (56)$$

the function $\mathbb{P}(\mu_1^T X \geq 0, \dots, \mu_k^T X \geq 0)$ is differentiable.

PROOF. 1) If the plane $\mu^T x = 0$ does not intersect \mathcal{P} then the derivative of $\mathbb{P}(\mu^T X \geq 0)$ must be zero. Similarly, if the plane $\mu^T x = 0$ intersects the boundary of \mathcal{P} , the derivative of $\mathbb{P}(\mu^T X \geq 0)$ has magnitude that is proportional to the area of the intersection of $\mu^T x = 0$ in \mathcal{P} . But since $\mu^T x = 0$ is not a boundary surface this area is zero and so the derivative of $\mathbb{P}(\mu^T X \geq 0)$ is zero.

Now, we suppose $\mu^T x = 0$ intersects the interior of \mathcal{P} . Since there are negligible boundary effects, we can apply standard calculus arguments to differentiate $\mathbb{P}(\mu^T X \geq 0)$. Given, $X = (X_1, \dots, X_I)$ has density $f_X(x_1, \dots, x_I)$ then $\mu X = (\mu_1 X_1, \dots, \mu_I X_I)$ has density

$$f_{\mu X}(y_1, \dots, y_I) = f_X\left(\frac{y_1}{\mu_1}, \dots, \frac{y_I}{\mu_I}\right) \prod_{i=1}^n \mu_i^{-1}. \quad (57)$$

Thus $\mu^T X$ is a convolution and so its density is given by

$$f_{\mu^T X}(z) = \int f_X\left(\frac{y_1}{\mu_1}, \dots, \frac{z - \sum_{i=1}^{I-1} \mu_i y_i}{\mu_I}\right) \prod_{i=1}^n \mu_i^{-1} dy_1 \dots dy_{I-1}.$$

And thus differentiating inside the integral, [Williams 1991, §A16] we have

$$\begin{aligned} \frac{\partial f_{\mu^T X}}{\partial \mu_i}(z) &= \\ \int \frac{\partial}{\partial \mu_i} \left\{ f_X\left(\frac{y_1}{\mu_1}, \dots, \frac{z - \sum_{k=1}^{I-1} \mu_k y_k}{\mu_I}\right) \prod_{i=1}^n \mu_i^{-1} \right\} dy_1 \dots dy_{I-1}. \end{aligned} \quad (58)$$

Finally, since again we can differentiate inside the integral, we see that

$$\frac{\partial}{\partial \mu_i} \mathbb{P}(\mu^T X \geq 0) = \int_0^\infty \frac{\partial f_{\mu^T X}}{\partial \mu_i}(z) dz, \quad (59)$$

where the derivative $\frac{\partial f_{\mu^T X}}{\partial \mu_i}(z)$ is given above.

2) Suppose we wish to differentiate with respect to one of the components of μ_1 . The conditional distribution

$$\mathbb{P}(\cdot | \mu_2^T X \geq 0, \dots, \mu_k^T X \geq 0) \quad (60)$$

has a continuous density defined on polytope $\mathcal{P} = \{x : \mu_2^T x \geq 0, \dots, \mu_k^T x \geq 0\}$. Provided the vectors are distinct then $\{x : \mu_1^T x = 0\}$ is not a boundary to this polytope. Thus we can apply the first part of this lemma and the chain rule to give the result. \square

So provided there is a certain amount of variability in p_{ij}^T , then we can expect the average performance of a advertise to be a continuous function of the declared prices λ . This will be useful for achieving convergence.

PROOF OF PROPOSITION 5.2. We let \mathcal{P} index the of assignments from \mathcal{I} to \mathcal{L} . Notice, provide there is a unique maximal assignment,

$$\begin{aligned} x_{il}^\tau(\lambda) &= \sum_{\pi \in \mathcal{P}: \pi(i)=l} \mathbf{I} \left[\sum_k \lambda_k p_{k\pi(k)}^\tau \geq \sum_k \lambda_k p_{k\tilde{\pi}(k)}^\tau, \forall \tilde{\pi} \neq \pi \right] \\ &= \sum_{\pi \in \mathcal{P}: \pi(i)=l} \prod_{\tilde{\pi} \neq \pi} \mathbf{I} \left[\sum_k \lambda_k p_{k\pi(k)}^\tau \geq \sum_k \lambda_k p_{k\tilde{\pi}(k)}^\tau \right] \end{aligned} \quad (61)$$

Here \mathbf{I} is the indicator function.

Notice, since \mathbb{E}_τ admits a density, $f(p^\tau)$, then with probability one, there is a unique maximizer to the assignment problem ASSIGNMENT(τ, λ): The probability two assignments have the same value is

$$\mathbb{P}_\tau \left(\sum_k \lambda_k p_{k\pi(k)}^\tau = \sum_k \lambda_k p_{k\tilde{\pi}(k)}^\tau \right) \quad (62)$$

$$= \int \mathbf{I} \left[\sum_k \lambda_k p_{k\pi(k)}^\tau = \sum_k \lambda_k p_{k\tilde{\pi}(k)}^\tau \right] f(p) dp = 0. \quad (63)$$

In other words, because we integrate over a set of one dimension less than the dimension of the space of p , the integral is zero. Thus the probability that there are two maximal assignments is zero. So the equality (61) holds almost surely for all $\lambda \neq 0$.

For two assignments π and $\tilde{\pi}$, we define the vector

$$\mu_{\pi\tilde{\pi}} := (\lambda_i \mathbf{I}[\pi(i) = l] - \lambda_i \mathbf{I}[\tilde{\pi}(i) = l]) : i \in \mathcal{I}, l \in \mathcal{L}.$$

Notice for any two distinct permutations the non-zero components of $\mathbf{I}[\tilde{\pi}(i) = l]$ are distinct. So the vectors $\mu_{\pi\tilde{\pi}}$ are distinct over $\tilde{\pi} \neq \pi$. Since the maximal assignment is almost surely unique, we have

$$\begin{aligned} \bar{x}_{il}(\lambda) &= \sum_{\pi \in \mathcal{P}: \pi(i)=l} \mathbb{E} \left[\prod_{\tilde{\pi} \neq \pi} \mathbf{I} \left[\sum_k \lambda_k p_{k\pi(k)}^\tau \geq \sum_k \lambda_k p_{k\tilde{\pi}(k)}^\tau \right] \right] \\ &= \sum_{\pi \in \mathcal{P}: \pi(i)=l} \mathbb{P} \left(\mu_{\pi\tilde{\pi}}^\top p \geq 0, \forall \tilde{\pi} \neq \pi \right). \end{aligned} \quad (64)$$

Thus if the function $\mathbb{P}(\mu_{\pi\tilde{\pi}}^\top p \geq 0, \forall \tilde{\pi} \neq \pi)$ is differentiable and Lipschitz continuous then we have same properties for functions $\bar{x}_{il}(\lambda)$. The Lipschitz continuity and differentiability of $\mathbb{P}(\mu_{\pi\tilde{\pi}}^\top p \geq 0, \forall \tilde{\pi} \neq \pi)$ is proven in Lemmas C.1 and C.2 respectively.

We now have the differentiability and Lipschitz property for \bar{x} . As we shall explain, the differentiability and Lipschitz properties of y are a consequence of that holding for \bar{x} , after a change of measure. The summands used to calculate are

$$\begin{aligned} \mathbb{E}_\tau p_{il} x_{il}^\tau(\lambda) &= \int x_{il}^\tau(\lambda) p_{il} f(p) dp \\ &= \mathbb{E}_\tau [p_{il}] \times \int x_{il}^\tau(\lambda) \tilde{f}_{il}(p) dp \end{aligned}$$

where \tilde{f}_{il} is the continuous bounded density function

$$\tilde{f}_{il}(p) = \begin{cases} \frac{p_{il} f(p)}{\mathbb{E}_\tau p_{il}}, & \text{if } \mathbb{E}_\tau p_{il} > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (65)$$

In other words, $\mathbb{E}_\tau p_{il} x_{il}^\tau(\lambda)$ is exactly $\tilde{\mathbb{E}}_\tau x_{il}^\tau$ but the expectation is calculated with a different continuous density function. Thus, differentiability and Lipschitz continuity follow for the summands of y for exactly the same reason as they did for \bar{x}_{il} . Thus y is differentiable and Lipschitz continuous. \square

LEMMA C.3. *The function $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ is strictly increasing.*

PROOF. Given prices λ let $x_{il}^\tau(\lambda)$ is an optimal solution to the assignment problem (1) and define

$$y_i^\tau(\lambda) := \sum_l p_{il}^\tau x_{il}^\tau.$$

Clearly, $y_i(\lambda) = \mathbb{E}_\tau y_i^\tau(\lambda)$. Thus, if we can prove $y_i^\tau(\lambda)$ is increasing then so is $y_i(\lambda)$. Further note

$$\sum_i \lambda_i y_i^\tau(\lambda) = \sum_{il} \lambda_i p_{il}^\tau x_{il}^\tau(\lambda) \quad (66)$$

which is the optimal objective for the assignment problem (1).

Define λ' with $\lambda'_i < \lambda_i$ and $\lambda'_j = \lambda_j$ for each $j \neq i$. We now proceed by contradiction. Suppose that $y_i(\lambda') > y_i(\lambda)$, then the following equalities and inequalities hold

$$\sum_j \lambda_j y_j^\tau(\lambda) = (\lambda_i - \lambda'_i) y_i^\tau(\lambda) + \sum_j \lambda'_j y_j^\tau(\lambda) \quad (67)$$

$$\leq (\lambda_i - \lambda'_i) y_i^\tau(\lambda) + \sum_j \lambda'_j y_j^\tau(\lambda') \quad (68)$$

$$< (\lambda_i - \lambda'_i) y_i^\tau(\lambda') + \sum_j \lambda'_j y_j^\tau(\lambda') = \sum_j \lambda_j y_j^\tau(\lambda'). \quad (69)$$

Here the first equality holds by the optimality of $y^\tau(\lambda')$ and the second holds by assumption. But notice the resulting equality above contradicts the optimality of $y^\tau(\lambda)$. Thus by contradiction, $y_i^\tau(\lambda)$ is increasing in λ_i and thus taking expectations so is $y_i(\lambda)$.

We now prove that $\lambda_i \mapsto y_i(\lambda)$ is strictly increasing. Let λ' be such that $\lambda'_i > \lambda_i$ and $\lambda'_j = \lambda_j$ for all $j \neq i$. The result proceeds by showing that

$$\mathbb{P}(y_i^\tau(\lambda') > y_i^\tau(\lambda) | E) > 0 \quad (70)$$

where we condition on an event E with non-zero probability. Notice, after taking expectations, this implies that $y_i(\lambda') > y_i(\lambda)$.

Now since $f(p) > f_{\min}$, $p = (p_{il} : i \in \mathcal{I}, l \in \mathcal{L})$ stochastically dominates a uniform random variable on the set of increasing click-through rates, \mathcal{S} . Thus it is sufficient to prove the result for $u = (u_{il} : i \in \mathcal{I}, l \in \mathcal{L})$ uniform on \mathcal{S} . Now, for instance, there is positive probability that advertiser i and j , with $\lambda_j > 0$, compete over the top two slots, $l = 1, 2$. This occurs under the event, when i and j have click-through rate over one half and all other advertisers have expected revenue that half of this, namely, the event

$$E := \left\{ \max_{\substack{k=i,j \\ l \in \mathcal{L}}} \lambda_k u_{kl} \geq \frac{1}{2}, \quad 2 \max_{\substack{k \neq i,j \\ l \in \mathcal{L}}} \{u_{kl}\} \leq \frac{1}{2} \min\{\lambda_i, \lambda_j\} \right\}. \quad (71)$$

Given this event, advertiser i achieves the top position with bid λ'_i and the second position with bid λ_i with on condition

$$\lambda'_i(u_{i1} - u_{i2}) > \lambda_j(u_{j1} - u_{j2}) > \lambda_i(u_{i1} - u_{i2}). \quad (72)$$

Since $u_{i1}, u_{i2}, u_{j1}, u_{j2}$ remain uniformly distributed (on $[1, 1/2]$) it is a straightforward calculation that

$$\mathbb{P}(\lambda'_i(u_{i1} - u_{i2}) > \lambda_j(u_{j1} - u_{j2}) > \lambda_i(u_{i1} - u_{i2}) | E) > 0, \quad (73)$$

which, since $y_i(\lambda)$ is increasing, implies

$$\mathbb{P}(y_i^\tau(\lambda') > y_i^\tau(\lambda) | E) > 0, \quad (74)$$

and thus $y_i(\lambda) < y_i(\lambda')$, as required. \square

PROPOSITION C.4. *The function $\lambda \mapsto \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda)$ is a continuous convex and is differentiable function for $\lambda \neq 0$ with derivatives given by*

$$\frac{d}{d\lambda_i} \left\{ \sum_{i' \in \mathcal{I}} \lambda_{i'} y_{i'}(\lambda) \right\} = y_i(\lambda). \quad (75)$$

Further,

$$\lim_{\|\lambda\| \rightarrow \infty} \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda) = \infty. \quad (76)$$

PROOF. From Proposition 5.2, we know that the function $\lambda \mapsto y_i(\lambda)$ is continuously differentiable for $\lambda \neq 0$. Also, since the function $y_i(\lambda)$ is bounded $\lambda_i y_i(\lambda)$ is continuous at zero. Thus the function $\lambda \mapsto \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda)$ is a continuous and is differentiable for $\lambda \neq 0$.

Next we can see that the function

$$\lambda \mapsto \sum_{i \in \mathcal{I}} \lambda_i y_i(\lambda)$$

is convex. Taking $\lambda^0, \lambda^1 \in \mathbb{R}^{\mathcal{I}}$ and defining $\lambda^p = p\lambda^1 + (1-p)\lambda^0$, the argument for this is as follows

$$\begin{aligned} & \sum_{i \in \mathcal{I}} \lambda_i^p y_i(\lambda^p) \\ &= \mathbb{E}_\tau \left[\max_{x^\tau \in \mathcal{X}} \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i^p p_{il}^\tau x_{il}^\tau \right] \\ &= \mathbb{E}_\tau \left[\max_{x^\tau \in \mathcal{X}} \left\{ p \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i^1 p_{il}^\tau x_{il}^\tau + (1-p) \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i^0 p_{il}^\tau x_{il}^\tau \right\} \right] \\ &\leq p \mathbb{E}_\tau \left[\max_{x^\tau \in \mathcal{X}} \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i^1 p_{il}^\tau x_{il}^\tau \right] \\ &\quad + (1-p) \mathbb{E}_\tau \left[\max_{x^\tau \in \mathcal{X}} \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_i^0 p_{il}^\tau x_{il}^\tau \right] \\ &= p \sum_{i \in \mathcal{I}} \lambda_i^1 y_i(\lambda^1) + (1-p) \sum_{i \in \mathcal{I}} \lambda_i^0 y_i(\lambda^0). \end{aligned}$$

The differentiability result (75) follows according the Envelope Theorem as follows – for further details see [Milgrom 2004, Chap. 3]. Since by definition, $x^\tau(\lambda)$ is optimal