Efficient Advert Assignment

Frank Kelly[1], Peter Key[2], and Neil Walton[3]

[1]University of Cambridge
[2]Microsoft Research Cambridge
[3]University of Amsterdam

Abstract

We develop a framework for the analysis of large-scale Ad-auctions where adverts are assigned over a continuum of search types. For this pay-per-click market, we provide an efficient and highly decomposed mechanism that maximizes social welfare. In particular, we show that the social welfare optimization can be solved in separate optimizations conducted on the time-scales relevant to the advertisement platform and advertisers. Here, on each search occurrence, the platform solves an assignment problem and, on a slower time scale, each advertiser submits a bid which matches its demand for click-throughs with supply. Importantly knowledge of global parameters, such as the distribution of search terms, is not required when separating the problem in this way. This decomposition is implemented in an adversarial setting. Exploiting the information asymmetry between the platform and advertiser, we describe a simple mechanism which incentivizes truthful bidding and has a unique Nash equilibrium that is socially optimal, and thus implements our decomposition. Further, we consider models where advertisers adapt their bids smoothly over time, and prove convergence to the solution that maximizes aggregate utility. Finally, we describe several extensions which illustrate the flexibility and tractability of our framework.

1 Introduction

Ad-auctions lie at the heart of search markets and generate billions of dollars in revenue for platforms such as Bing and Google. Sponsored search auctions provide a distributed mechanism where advertisers compete for their adverts to be shown to users of the search platform, by bidding on search terms associated with queries. Select adverts are allocated to different positions on the search page, and if the user clicks on an ad, a payment is made by the advertiser to the platform — the so-called *Pay-Per-Click model*.

In current auctions, there is a fundamental information asymmetry between the platform and advertisers, in that the platform typically knows more than an advertiser about the search being conducted, such as information about the searcher. Hence the platform can potentially choose prices and an allocation that depends on the platform's additional information. In contrast, 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.

Variability in the platform's additional information creates variability in the observations available to the advertiser: it is difficult for the advertiser to view consecutive allocations by the platform as repeated instances of the same auction since search terms and thus competitors vary, and even if the keyword and range of competitors stay the same, the platform's additional information varies from search to search. This implies that the observations available to the advertiser are inherently stochastic, and that they need to be filtered in order to extract reliable information. Thus the information asymmetry between platform and advertisers induces a temporal asymmetry: the platform observes each search as it happens, whereas the advertiser has to rely on delayed and aggregated feedback.

In this paper we develop a framework to address this information asymmetry directly. We consider the objective of maximizing the social welfare of advertisers, and show this can be achieved by decomposing the optimization into two separate optimizations. Under the decomposition, first, advertisers bid to solve their individual utility maximization problems and, second, adverts are assigned on each search occurrence. A consequence of the first optimization is that the platform only requires a single parameter bid from the advertiser, rather than, say, the advertiser's entire utility function which would be required to implement a VCG mechanism. The second optimization operates over the platform's entire search space by separately optimizing on each search occurrence. The resulting mechanism does not require off-line computations or estimation of distribution of searches in order to assign and price advertisements.

Optimization frameworks of this form are well-established in the communication network community, where users, the network and its components must be separated. There the phrase "Network Utility Maximization" has been coined, but this framework has only recently found its way into Mechanism Design (see

[Maheswaran and Basar, 2004], [Yang and Hajek, 2007] and [Johari and Tsitsiklis, 2009]). By contrast, 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]). We focus upon streams of search queries: their randomness and the resulting information asymmetry is an intrinsic aspect of our framework.

The generalized second price auction is canonically used by search engines. This mechanism requires an ordered layout of interchangeable adverts. The framework does not readily adapt to more complex page layouts, such as text rich adverts, adverts of variable size or adverts incorporating images – each of which are of current and increasing demand for modern online advertisement platforms. However, the flexibility to compare and price complex assignments is inherent in VCG mechanisms, and through this, we determine efficient pricing implementations for general page layouts.

The mechanism established by our analysis is simple, flexible and implementable. Given the advertisers' bids and the estimated click-through rates of the search, we assign adverts to maximize the (expected) bid of an advert receiving a click. This is a classical assignment problem and is solved with low computational overhead on each search instance. If an advert receives a click-through, then the pricing of that advert requires one further solution to this assignment problem. We model advertisers as utility maximizers, who 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 where π_i and y_i are, respectively, the (average) price they pay per click and the click-through rate they achieve.

We prove that the Nash equilibrium achieved by advertisers maximizes the social welfare of the advertisers. 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 these prices can be formulated and implemented on the time-scales relevant to sponsored search: adverts are 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. Thus search-engine-wide optimization is highly non-trivial task. Our results show that, despite these informational and computational constraints, optimal social welfare is achievable over a search engine's infinitely large search space, and by a low complexity mechanism which assigns and prices adverts on the time-scales required for sponsored search.

1.1 Related Work

In this paper we consider a problem where the aggregate utility (or social welfare) 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, see [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., 2010], [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 range of searches that can be conducted. In this way, the sponsored search problem is separated into problems with relevant constraints acting on relevant timescales. 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 where they suggest that current communication protocols converge to solutions of global optimization problems, [Srikant, 2004, Kelly and Yudovina, 2014]. The work of [Tan and Srikant, 2012] 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.

Strategic formulations of these decompositions have since been developed: [Johari and Tsitsiklis, 2004] show a price of anarchy of 75% at a Nash equilibrium. A single parameter VCG-mechanism to yield efficient allocation is considered by [Maheswaran and Basar, 2004] and was subsequently generalized by [Yang and Hajek, 2007] and [Johari and Tsitsiklis, 2009]. Here a parametrized surrogate utility is employed in the VCG mechanism, where the parameter is selected strategically. The information passed from the player to the mechanism is thus reduced and the parametrized mechanism requires a computational overhead that is ostensibly equal to its parametrized VCG counterpart. One part of our decomposition can be viewed as deriving a single parameter VCG mechanism with linear utilities; that is, despite allowing more general user utility functions, the derived mechanism is *as if* each user had a linear utility function. Prior works have used strictly concave surrogate functions while in our approach linearization is possible owing to the large search/constraint space employed. The crucial advantage of our linear VCG allocation is that – in addition to decomposing the objectives of advertisers and the platform – further decomposition over the search space is possible, leading to practical assignment and pricing. In particular, the mechanism can be implemented on each search instance: allocation and pricing both involve the solution of standard polynomial time algorithms per-search and per-click, respectively. In essence, we find a simple implementable auction mechanism that yields an efficient allocation of adverts across the entire search space. Further, in special cases this decomposed implementation reduces to a number of known auctions, namely, second price auctions and the generalized English auction of [Edelman et al., 2007].

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, see [Pin and Key, 2011]. [Athey and Nekipelov, 2012] introduce a specific distributional assumption across the rank scores (weighted bids) to deal with randomness and provide a static analysis.

As recently argued by [Varian and Harris, 2014], VCG mechanisms are of practical interest because they are flexible and extendible. For this reason, Facebook implements VCG rather than the generalized second price auction currently used by Bing and Google. These considerations are particularly relevant for contextual advertisement, unordered page layouts, image-text adverts and image-video adverts. Such extensions are important and are of growing interest to online advertisement platforms; see [Goel and Khani, 2014] for some recent considerations. We discuss how these generalizations can be implemented within our framework, whilst maintaining a welfare optimal allocation of advertisements.

1.2 Outline

In Section 2 we develop our model of a sponsored search platform. The assignment used by the platform is presented, and we also give the pricing formulation that is later analyzed for strategic advertisers. The mechanism is presented, but the theory required to understand the rationale behind this choice is developed in subsequent sections.

In Section 3 we apply a decomposition argument to an aggregate utility optimization over the uncountably infinite set of constraints in our model of sponsored search. The argument is based on techniques from convex optimization and duality and proofs, complicated by the infinite setting, are mostly relegated to the appendix. The upshot of these preliminaries is that the advertisers equating bids with their marginal utility and the platform solving a maximum-weighted assignment problem for each search instance will maximize aggregate utility. Crucially, the time-scale and information asymmetry in this decomposition are those relevant to sponsored search. The advertisers are optimizing over a slower time-scale 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.

Section 3 provides the necessary groundwork for Section 4 where we make the connection with mechanism design and strategic advertisers. In particular, we find the form of a rebate which incentivizes advertisers to truthfully declare bids that equate to their marginal utilities. This produces a unique Nash equilibrium which implements our decomposition. By the separability of our optimization, the rebate can be computed by a simple mechanism whose pricing requires a single additional computation, namely the solution of an assignment problem, for each click-through. Hence assignment and pricing occur per search query involve straightforward polynomial-time computations.

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. These include more complex page-layouts, controlling the number of display slots (for instance, through reserve prices), multivariate utilities, and bidding for budget constrained advertisers.

2 The Assignment and Pricing 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 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 $l \in \mathcal{L}$. The set of slots is ordered, with the first (lowest ordered) slot representing the top slot. 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 of the search, the time of day, the gender and age of the searcher etc. All of these factors are combined to form a probability of click-through p^τ_{il} which is estimated by the platform.

Over time, a large number of searches from the set $\mathcal{T}$ are made. We assume these occur with distribution $\mathbb{P}_\tau$. Thus we view the click-through probability $p_{il} : \mathcal{T} \to [0, 1]$ as a random variable defined on the type space $\mathcal{T}$ and with distribution $\mathbb{P}_\tau$. 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 $u = (u_{il} : i \in \mathcal{I}, l \in \mathcal{L}) \in [0, 1]^{\mathcal{I}\times\mathcal{L}}$,

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

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

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 captured in τ, and so can successfully predict the click-through probability p^τ_{il}, 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 system towards an optimal outcome.

2.1 Assignment Model

Next we describe a mechanism by which the platform assigns 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 sum of bids on click-throughs from a single search.

ASSIGNMENT(τ, λ)

$$\begin{aligned}
&\text{Maximize} && \sum_{i\in\mathcal{I}} \lambda_i \sum_{l\in\mathcal{L}} p^\tau_{il} x^\tau_{il} && \text{(2.1a)}\\
&\text{subject to} && \sum_{i\in\mathcal{I}} x^\tau_{il} \leq 1, \quad l \in \mathcal{L}, && \text{(2.1b)}\\
& && \sum_{l\in\mathcal{L}} x^\tau_{il} \leq 1, \quad i \in \mathcal{I}, && \text{(2.1c)}\\
&\text{over} && x^\tau_{il} \geq 0, \quad i \in \mathcal{I}, l \in \mathcal{L}. && \text{(2.1d)}
\end{aligned}$$

The above optimization is an assignment problem, where the constraint (2.1b) prevents a slot containing more than one advert, and the constraint (2.1c) 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 (see [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 $\mathbb{P}_\tau$ but we require an estimate of the click-through probability p^τ_{il}, as is the case for the generalized second price auctions currently used in sponsored search.

We apply the convention that if $\lambda_i = 0$ then $x^\tau_{il} = 0$ 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 with probability one; this follows, for example, when the distribution of click-through probabilities p admits a density.

We let

$$y^\tau_i = \sum_{l\in\mathcal{L}} p^\tau_{il} x^\tau_{il}, \qquad y_i = \mathbb{E}_\tau y^\tau_i, \qquad x_{il} = \mathbb{E}_\tau \left[x^\tau_{il}\right]. \tag{2.2}$$

Note that y^τ_i 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^τ_i.) To model the asymmetry between the platform and advertiser, we assume that y^τ_i is known to the platform, and y_i can be observed by advertiser i. For an optimal solution to the above assignment problem, write $x^\tau_{il} = x^\tau_{il}(\lambda)$ to emphasize the dependence of

x^τ_{il} on the vector of bids λ and, similarly, write $x_{il} = x_{il}(\lambda)$, $y^\tau_i = y^\tau_i(\lambda)$, $y_i = y_i(\lambda)$. Let (μ_i, λ_{-i}) be the vector obtained from λ by replacing the ith component by μ_i.

We shall assume the following *monotonicity property* of solutions of ASSIGNMENT(τ, λ). We assume the function $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ is positive for $\lambda_i > 0$ and is strictly increasing and continuous when $\lambda_{-i} > 0$. Without the monotonicity property $y_i(\lambda)$ will be increasing in λ_i but may not be strictly increasing or continuous, and this would complicate the statement of several later results.

The monotonicity property will generally follow from sufficient variability of click-through rates. For instance, a sufficient condition is that the random variables p admit a continuous density $f(p)$ positive on the simplex of click-through probabilities $\tilde{\mathcal{P}} = \{p \in [0,1]^{|\mathcal{I}|\times|\mathcal{L}|} : p_{il} \geq p_{ik}, l < k\}$. Observe that on $\tilde{\mathcal{P}}$, the click-through probability for a given advert increases as the slot it is shown in decreases. The following result establishes the monotonicity property under the above sufficient condition.

Proposition 1. *If the distribution $\mathbb{P}_\tau$ admits a continuous probability density function positive on $\tilde{\mathcal{P}}$ then the monotonicity property is satisfied, namely, $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ is positive for $\lambda_i > 0$ and is strictly increasing and continuous when $\lambda_{-i} > 0$.*

The proof of Proposition 1 is given in Appendix A.

2.2 Pricing Model

Once adverts are allocated, prices must be determined for any resulting click-throughs. We consider a mechanism where the (average) price charged per click to advertiser i is given by

$$\pi_i(\lambda) = \frac{1}{y_i(\lambda)} \int_0^{\lambda_i} \Big(y_i(\lambda) - y_i(\mu_i, \lambda_{-i}) \Big) d\mu_i. \tag{2.3}$$

Here, as before, $\lambda = (\lambda_i : i \in \mathcal{I})$ is the vector of advertisers' bids and $y_i(\lambda)$ is the resulting click-through rate for advertiser i. We discuss later the reasoning behind this pricing mechanism in the proper context of a highly decomposed sponsored search problem. However, for now we will note that it is an implementable form of (parameterized) VCG pricing.

The price function (2.3) 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 ASSIGNMENT(τ, λ), and observes a click-through on (i, l) — that is the solution has $x^\tau_{il} = 1$, and the user clicks on the advert in position l, which is for advertiser i. To price this advert, the platform chooses μ_i uniformly and randomly on the interval $(0, \lambda_i)$ and additionally solves ASSIGNMENT$(\tau, (\mu_i, \lambda_{-i}))$. Let $y^\tau_i(\mu_i, \lambda_{-i}) = \sum_{l\in\mathcal{L}} p^\tau_{il} x^\tau_{il}$ under a solution to this problem. The platform then charges advertiser i an amount

$$\lambda_i \left(1 - \frac{y^\tau_i(\mu_i, \lambda_{-i})}{y^\tau_i(\lambda)} \right) \tag{2.4}$$

for the click-through. This charge does not depend on the distribution $\mathbb{P}_\tau$, 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

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

Thus the mechanism charges an average price per click to advertiser i of $\pi_i(\lambda)$, given by expression (2.3).

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 (2.4) for a click-through. Indeed, one could imagine a charge λ_i on the click-through, followed by a later rebate of a proportion $y^\tau_i(\mu_i, \lambda_{-i})/y^\tau_i(\lambda)$ 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 not observed by the advertiser.

There are other ways the price function (2.3) 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 (see Proposition 5 of Appendix A) and both expressions take the value 0 when $\lambda_i = 0$. Thus the price function (2.3) can be also implemented by a charge λ_i on the click-through followed by a later rebate

$$\frac{1}{y^\tau_i(\lambda)} \Big(\sum_j \lambda_j y^\tau_j(\lambda) - \sum_{j\neq i} \lambda_j y^\tau_j(0, \lambda_{-i}) \Big). \tag{2.5}$$

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 $\lambda_j y_j$.

Finally, we give some settings where closed forms are available for (2.3).

Example 1. *If there is a single slot then the slot will be assigned to the bidder i with the highest value of $\lambda_i p^\tau_{i1}$, and if this results in a click-through then the charge will be*

$$\max_{j \neq i} \frac{\lambda_j p^\tau_{j1}}{p^\tau_{i1}},$$

a second price auction on the products $\lambda_j p^\tau_{j1}$.

Suppose next there are L slots with more than L advertisers bidding. Further suppose the click-through probabilities are the same for each advertiser, $p^\tau_{il} = p^\tau_l, i \in \mathcal{I}$, where $p^\tau_1 > p^\tau_2 > \ldots > p^\tau_L$, and the bids are $\lambda_1 > \lambda_2 > \ldots > \lambda_{|\mathcal{I}|}$. So advertisers $1, 2, \ldots, L$ are allocated slots $1, 2, \ldots, L$ respectively. In this example it is possible to calculate the expectation of the expression (2.4) *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*

$$\pi^\tau_l = \lambda_{l+1} - \frac{1}{p^\tau_l} \sum_{m=l+1}^{L} p^\tau_m (\lambda_m - \lambda_{m+1}), \quad l = 1, 2, \ldots, L.$$

Expressed as a recursion this implies

$$\pi^\tau_l = \lambda_{l+1} - \frac{p^\tau_{l+1}}{p^\tau_l} (\lambda_{l+1} - \pi^\tau_{l+1}), \quad l = 1, 2, \ldots, L,$$

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

$$\sum_{m=1}^{L} \pi^\tau_m p^\tau_m = \sum_{m=1}^{L} m \lambda_{m+1} (p^\tau_m - p^\tau_{m+1}), \tag{2.6}$$

with $p^\tau_{L+1} = 0$; note the dependence on λ_{L+1}, the largest unsuccessful bid.

In this example (with click-through probabilities the same for all advertisers) the monotonicity property, that $y_i(\lambda)$ is a strictly increasing and continuous function of λ_i, does not hold. But if there exists a single additional advertiser i^* with $\lambda_{i*} > \lambda_1 p_1$ and random click-through probabilities which admit a continuous density positive on the simplex $\{p_{i^*l} \geq p_{i^*k}, l < k\} \subset [0,1]^{|\mathcal{L}|}$, then $y(\lambda)$ will have the monotonicity property. The key point is that a change in the bid λ_i should cause a change in the solution to the assignment problem for at least some search types τ and hence a change in the rate $y_i(\lambda)$.

Example 2. *Click-through probabilities may depend on more than just the slot. The advertiser and search type will generally also have an effect. Our framework anticipates this variability between advertisers competing over slots. For instance, if there is a single slot and advertisers' click-through probabilities are independent, with density function f_i and distribution function F_i for advertiser i, then it is a relatively straightforward calculation that $y_i(\lambda)$ and $\pi_i(\lambda)$ are given by*

$$y_i(\lambda) = \int_0^1 \prod_{j \neq i} F_j\Big(p \frac{\lambda_i}{\lambda_j}\Big) \times p f_i(p) dp$$

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

One can, and indeed we do, mix the two previous examples by allowing click-through probabilities to depend on the advertiser, search type, and position. Explicit closed form expressions for (average) prices are not available in general. However our assignment and pricing models do not require such formulae to be implemented.

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 a click-through rate of y_i has a utility of $U_i(y_i)$ to advertiser i, where $U_i(\cdot)$ is non-negative, increasing, and strictly concave. Our objective is to place adverts so as to maximize the sum of these utilities, in other words to maximize social welfare. To simplify the statement of results, we shall assume further that $U_i(\cdot)$ is continuously differentiable, with $U_i'(y_i) \to \infty$ as $y_i \downarrow 0$ and $U_i'(y_i) \to 0$ as $y_i \uparrow \infty$. The resulting optimization for the auction system is as follows.

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

$$\text{Maximize} \quad \sum_{i\in\mathcal{I}} U_i(y_i) \tag{3.1a}$$

$$\text{subject to} \quad y_i = \mathbb{E}_\tau\Big[\sum_{l\in\mathcal{L}} p_{il}^\tau x_{il}^\tau\Big], \quad i \in \mathcal{I}, \tag{3.1b}$$

$$\sum_{i\in\mathcal{I}} x_{il}^\tau \le 1, \quad l \in \mathcal{L}, \tau \in \mathcal{T}, \tag{3.1c}$$

$$\sum_{l\in\mathcal{L}} x_{il}^\tau \le 1, \quad i \in \mathcal{I}, \tau \in \mathcal{T}, \tag{3.1d}$$

$$\text{over} \quad x_{il}^\tau \ge 0, y_i \ge 0 \quad i \in \mathcal{I}, l \in \mathcal{L}. \tag{3.1e}$$

Inequalities (3.1c) and (3.1d) are just the scheduling constraints (2.1b) and (2.1c), that each slot can show at most one advert and that each advertiser can show at most one advert, while equality (3.1b) recaps the definition (2.2) of y_i, the expected click-through rate. Over these constraints we maximize aggregate utility.

Note the optimization assumes the system optimizer has knowledge of the utility functions $U_i(\cdot), i \in \mathcal{I}$, the click-through probabilities $p_{il}^\tau, i \in \mathcal{I}, l \in \mathcal{L}, \tau \in \mathcal{T}$ and the distribution $\mathbb{P}_\tau$ 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.

Incorporating the constraint (3.1b) into the objective function (3.1a) gives the Lagrangian

$$L_{sys}(x, y; \lambda) = \sum_{i\in\mathcal{I}} U_i(y_i) + \sum_{i\in\mathcal{I}} \lambda_i \mathbb{E}_\tau\left[\sum_{l\in\mathcal{L}} p_{il}^\tau x_{il}^\tau - y_i\right]. \tag{3.2}$$

Notice, we intentionally omit the scheduling constraints from our Lagrangian. Thus we seek to maximize the Lagrangian subject to the constraints (3.1c-3.1d) as well as (3.1e). Let $\mathcal{S}$ be the set of variables $x^\tau = (x_{il}^\tau : i \in \mathcal{I}, l \in \mathcal{L})$ satisfying the assignment constraints (2.1b-2.1d), and let $\mathcal{A}$ be the set of variables $x = (x^\tau \in \mathcal{S} : \tau \in \mathcal{T})$ satisfying the assignment constraints (3.1c-3.1e). We see that our Lagrangian problem is separable in the following sense

$$\max_{x\in\mathcal{A}, y\ge 0} L_{sys}(x, y; \lambda) = \sum_{i\in\mathcal{I}} \max_{y_i\ge 0} \{U_i(y_i) - \lambda_i y_i\} \tag{3.3a}$$

$$+ \mathbb{E}_\tau\left[\max_{x^\tau\in\mathcal{S}} \sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}} \lambda_i p_{il}^\tau x_{il}^\tau\right]. \tag{3.3b}$$

Define

$$U_i^*(\lambda_i) = \max_{y_i\ge 0} \{U_i(y_i) - \lambda_i y_i\}. \tag{3.4}$$

The optimization over y_i contained in the definition (3.4) would arise if advertiser i were presented with a fixed price per click-through of λ_i, and if 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

$$U_i^*(\lambda_i) = \int_{\lambda_i}^\infty D_i(\xi) d\xi; \tag{3.5}$$

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 (3.3b) is simply the problem ASSIGNMENT(τ, λ), and thus we can write

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

The Lagrangian dual of the SYSTEM problem (3.1) can thus be written as follows.

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

$$\text{Minimize} \qquad \sum_{i\in\mathcal{I}} (U_i^*(\lambda_i) + \lambda_i y_i(\lambda)) \tag{3.6a}$$

$$\text{over} \qquad \lambda_i \geq 0, \quad i \in \mathcal{I}. \tag{3.6b}$$

Owing to the size of the type space $\mathcal{T}$, the optimization (3.1) 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.

Proposition 2 (Decomposition)**.** *Variables $\tilde{y}$, $\tilde{x}^\tau, \tau \in \mathcal{T}$, satisfying the feasibility conditions (3.1b-3.1e) are optimal for 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. $\tilde{x}^\tau$ solves ASSIGNMENT($\tau, \tilde{\lambda}$), with probability one under the distribution $\mathbb{P}_\tau$ 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 the prices (2.3) that do give strategic advertisers the incentive to solve the SYSTEM problem. We do this in the next section, Section 4. There we shall also see that the per-click pricing implementations (2.4) and (2.5) are made possible by the decomposition into per-impression assignments, Condition B.

The optimal bids $\tilde{\lambda}$ can be further understood through the following dual characterization.

Proposition 3 (Dual Optimality)**.**
a) The objective of the dual problem (3.6) *is continuously differentiable for $\lambda > 0$ and is minimized uniquely by the positive vector $\tilde{\lambda} = (\tilde{\lambda}_i : i \in \mathcal{I})$ satisfying, for each $i \in \mathcal{I}$,*

$$\frac{dU_i^*}{d\lambda_i}(\tilde{\lambda}_i) + y_i(\tilde{\lambda}) = 0. \tag{3.7}$$

b) If $\tilde{\lambda}$ is an optimal solution to the DUAL problem (3.6) *then $x^\tau(\tilde{\lambda})$, $y(\tilde{\lambda})$ are optimal for the SYSTEM problem* (3.1)*.*

The dual provides a finite parameter optimization from which the SYSTEM problem can be solved. Moreover, (3.7) provides conditions on advertiser demands which, to solve the SYSTEM problem, must be effected by the auction system in strategic form.

Example 3. *We consider a brief example to emphasize the auction structure. Three advertisers, A, B and C compete for two advertisement slots shown in response to a specific search query, "Cambridge Pizza". Two of the advertisers, A and B, are takeaway pizza companies, one located in north Cambridge and the other in south Cambridge. 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 advert will only be clicked on if it appears in the top slot.*

Being concrete for a paragraph or so, suppose the platform observes that, given τ, the adverts receive click through probabilities

$$p_{Al}^\tau = \tau, \qquad p_{Bl}^\tau = \frac{1}{2} - \tau, \qquad p_{Cl}^\tau = \frac{1}{3}\mathbf{I}[l=1], \qquad \textit{for} \quad l = 1, 2, \tag{3.8}$$

and where τ accounts for random distance of the search from the advertiser. We assume τ is a uniform random variable on $[0, 1/2]$, although this distribution is not known by the platform or the advertisers. Suppose $\lambda = (\lambda_A, \lambda_B, \lambda_C)$ are the bids submitted by the advertisers. A straightforward calculation for advertiser A shows that her click-through rate is

$$y_A(\lambda) = \frac{1}{4} - \min\left\{\frac{\lambda_A}{2(\lambda_A + \lambda_B)}, \frac{\lambda_C}{3\lambda_A}\right\}^2. \tag{3.9}$$

We suppose advertisers have the same logarithmic utilities $U_i(y) = \log y$ and thus demands $D_i(\lambda_i) = \lambda_i^{-1}$ for $i = A, B, C$. Along with similar conditions for advertisers B and C, the optimal condition for the dual problem, (3.7)*, can be derived:*

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

There exist parameters satisfying these conditions for each advertiser, and the click-through rates $y(\lambda^*)$ are then optimal for the SYSTEM problem.

Without the decomposition result Proposition 2, determining an optimal solution to the SYSTEM optimization is a non-obvious problem. However, we find, to assign adverts, the platform required the correct click-through probabilities (3.8)*, search information τ with "Cambridge Pizza" and prices λ, but did not require information about advertisers' utilities or the distribution of searches $\mathbb{P}_\tau$. The advertiser could determine λ^*_A from her own advert's average performance* (3.9) *and her 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 $\mathbb{P}_\tau$.*

Finally, we remark that a much wider range of advertisers and search types can be considered in our framework. 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. We develop this point further in Section 6.

4 Mechanism Design

We now prove that our mechanism implements our system optimization. In the last section we demonstrated how this 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 a solve a certain single parameter dual problem. In this section we suppose the advertisers act strategically, anticipating the result of the platform's assignment and attempting to maximize their expected reward.

Henceforth λ_i is 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 solves the SYSTEM problem (3.1).

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

$$r_i(\lambda) = U_i(y_i(\lambda)) - \pi_i(\lambda) y_i(\lambda). \tag{4.1}$$

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

$$r_i(\lambda^*) \geq r_i(\lambda_i, \lambda_{-i}^*). \tag{4.2}$$

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

The main result of this section is the following.

Theorem 1. *If prices are charged according to the price function*

$$\pi_i(\lambda) = \frac{1}{y_i(\lambda)} \int_0^{\lambda_i} \Big(y_i(\lambda) - y_i(\mu_i, \lambda_{-i}) \Big) d\mu_i \tag{4.3}$$

then there exists a unique Nash equilibrium, and it is given by the vector of optimal prices identified in Proposition 3. Thus the assignment achieved at the Nash equilibrium, $(x(\lambda^*), y(\lambda^*))$, is a solution to the SYSTEM optimization.

The result states that, given adverts are assigned according to the assignment problem (2.1), the game theoretic equilibrium reached by advertisers attempting to maximize their respective rewards r_i solves the problem 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, and as we emphasized in Section 2, the prices (4.3) can be practically implemented in a sponsored search setting.

4.1 Proof of Theorem 1

To establish Theorem 1 we will require an additional result, Proposition 4, which indicates how maximal rewards achieved by each advertiser relate to the solution of the dual problem, given by Proposition 3 from the previous section.

Proposition 4 (Mechanism Dual)**.** *For each positive choice of $\lambda_{-i} = (\lambda_j : j \neq i, j \in \mathcal{I})$, the following equality holds*

$$\max_{\lambda_i \geq 0} r_i(\lambda) = \min_{\lambda_i \geq 0} \left\{ U_i^*(\lambda_i) + \int_0^{\lambda_i} y_i(\mu_i, \lambda_{-i}) d\mu_i \right\}. \tag{4.4}$$

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

$$\frac{d}{d\lambda_i} U_i^*(\lambda_i) + y_i(\lambda) = 0. \tag{4.5}$$

Proof. We calculate the conjugate dual of the reward function (4.1). Let $P_i(y_i)$ be the function whose Legendre-Fenchel transform is

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

The above function is increasing and convex, and we know from Fenchel's Duality Theorem, [Borwein and Lewis, 2006, Theorem 3.3.5], that

$$\max_{y_i \geq 0} \{U_i(y_i) - P_i(y_i)\} = \min_{\lambda_i \geq 0} \{U_i^*(\lambda_i) + P_i^*(\lambda_i)\}. \tag{4.6}$$

Next we 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

$$P_i(y_i) = \min_{\lambda_i \geq 0} \left\{ \lambda_i y_i - \int_0^{\lambda_i} y_i(\mu_i, \lambda_{-i}) d\mu_i \right\}. \tag{4.7}$$

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

$$P_i(y_i) = \int_0^{\infty} (y_i - y_i(\mu_i, \lambda_{-i})) \, \mathbf{I}[y_i(\mu_i, \lambda_{-i}) \leq y_i] d\mu_i. \tag{4.8}$$

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 (4.8) and this last observation, the equality (4.6) now reads

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

In the final equality we note from the definition (4.3) of $\pi_i(\lambda)$ that

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

This gives the equality (4.4).

We now show that both expressions (4.4) are determined at the same unique value of λ_i. The function $U_i^*(\lambda_i) - P_i^*(\lambda)$ is a strictly convex differentiable function of λ_i, whose unique minimum is given by the required expression

$$\frac{d}{d\lambda_i} U_i^*(\lambda_i) + y_i(\lambda) = 0. \tag{4.9}$$

Further, $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ is strictly increasing and λ_i achieves the range of the strictly concave function $U_i(y_i) - P_i(y_i)$ under $y_i = y_i(\lambda_i, \lambda_{-i})$. Thus $U_i(y_i) - P_i(y_i)$ is maximized uniquely by $y_i = y_i(\lambda_i, \lambda_{-i})$ (and thus uniquely by λ_i) satisfying

$$\frac{d}{dy_i} U_i(y_i(\lambda)) - \lambda_i = 0. \tag{4.10}$$

Since $\frac{d}{d\lambda_i} U_i^*$ is the inverse of the strictly increasing function $\frac{d}{dy_i} U_i$, it is clear that (4.9) and (4.10) are equivalent and satisfied by the same unique λ_i. This completes the proof. □

The proof of Theorem 1 follows by observing the optimality conditions of Propositions 3 and 4.

Proof. Proof of Theorem 1 Before proceeding with the main argument, we note that a Nash equilibrium must be achieved by positive values of λ_i. By applying the mean value theorem, for some $\tilde{y}$ satisfying $0 = y_i(0, \lambda_{-i}) \leq \tilde{y} \leq y_i(\lambda_i, \lambda_{-i})$, we have

$$\begin{aligned}
r_i(\lambda_i, \lambda_{-i}) &\geq U_i(0) + U_i'(\tilde{y})(y_i(\lambda_i, \lambda_{-i}) - y_i(0, \lambda_{-i})) - \int_0^{\lambda_i} (y_i(\lambda_i, \lambda_{-i}) - y_i(0, \lambda_{-i})) d\mu_i \\
&= r_i(0, \lambda_{-i}) + (U_i'(\tilde{y}) - \lambda_i)(y_i(\lambda_i, \lambda_{-i}) - y_i(0, \lambda_{-i})) \\
&> r_i(0, \lambda_{-i}).
\end{aligned} \tag{4.11}$$

The second term in (4.11) is positive for λ_i sufficiently small, since $U_i'(\tilde{y}) - \lambda_i \nearrow \infty$ as $\lambda_i \searrow 0$ and from our monotonicity property $y_i(\lambda_i, \lambda_{-i}) > y_i(0, \lambda_{-i})$. From this we see that a Nash equilibrium can only be achieved with $\lambda_i > 0$ for each $i \in \mathcal{I}$.

By Proposition 4, $\lambda = (\lambda_i : i \in \mathcal{I}) > 0$ is a Nash equilibrium if and only if condition (4.5) is satisfied for each $i \in \mathcal{I}$. But by Proposition 3b), these conditions hold if and only if λ is the unique solution to the dual to the SYSTEM problem. So, the set of Nash equilibria are the optimal prices defined for the decomposition, Proposition 2. By Proposition 3b), the assignment achieved by Nash equilibrium bids maximizes the utilitarian objective SYSTEM(U, $\mathcal{I}$, $\mathbb{P}_\tau$). Finally, by Strong Duality (Theorem 3 of Appendix C), there exists λ^* which optimizes the dual problem (3.6), and thus there must be a Nash equilibrium. □

Remark 1. *The optimality condition* (3.7) *or* (4.5) *states that each advertiser's demand,* $D_i(\lambda_i)$*, and supply,* $y_i(\lambda)$*, 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 Revenue 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's Lemma generally applies.*

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.

Convergence may not be possible when the type space is discrete, e.g. for an auction on a single search type. Essentially, the search engine does not have enough additional information from the search type τ to fine tune its discrimination between these advertisers. However, in sponsored search, there is inherent variability in the search type τ; a combination of factors, including spatial and temporal effects, will influence the click-through probabilities of the advertiser. This is the motivation for our assumption of the monotonicity property, that the distribution $\mathbb{P}_\tau$ over $\mathcal{T}$ is such that the click-through rate $y_i(\lambda)$ is a continuous, strictly increasing function of λ_i. We shall see that, under models of advertiser response, we are then able to deduce convergence towards a system optimum.

Consider the objective function for the dual of the system problem as derived in Proposition 3,

$$\mathcal{V}(\lambda) = \sum_{i\in\mathcal{I}} \left(U_i^*(\lambda_i) + \lambda_i y_i(\lambda)\right). \tag{5.1}$$

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 (3.1). We note that

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

This holds by the Envelope Theorem (as shown in Proposition 5 of Appendix A).

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

$$\frac{d}{dt}\lambda_i(t) \gtreqless 0 \text{ according as } \lambda_i(t) \lesseqgtr U_i'(y_i(\lambda(t))).$$

This is a natural dynamical system representation of advertiser i varying λ_i smoothly in order to track the optimum of her return $r_i(\lambda)$, given by expression (4.1), under prices (4.3). Note the reward function $r_i(\lambda)$ is maximized over λ_i when λ_i and $U_i'(y_i(\lambda))$ equate, or equivalently, when $y_i(\lambda)$ and $D_i(\lambda_i)$ equate (cf. (4.10) and (4.9) of Proposition 4). Thus we can employ $\mathcal{V}$ as a Lyapunov function to prove the following theorem.

Theorem 2 (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 problem* (3.6)*. Thus $y(\lambda(t))$, the assignment achieved by the prices $\lambda(t)$, converges to a solution of the SYSTEM problem* (3.1).

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 $D_i(\lambda_i) \to \infty$ as $\lambda_i \to 0$ 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 Proposition 5, that

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

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

$$y_i \lesseqgtr D_i(\lambda_i) \text{ according as } \lambda_i \lesseqgtr U_i'(y_i).$$

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

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

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(a), the price process $\lambda(t)$ converges to an optimal solution to the dual problem (3.6). Further, by Proposition 3(b), we know that $y(\lambda^*)$ is optimal for the system problem (3.1) and, by the monotonicity property, $y(\lambda)$ is a continuous function. Thus the click-through rates $y(\lambda(t))$ converge to an optimal solution for the system problem. □

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 a game played sequentially with advertisers iteratively maximizing over λ_i their reward function $r_i(\lambda_i, \lambda_{-i})$.

6 Extensions

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

6.1 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 adverts may vary in size, position, and include images and other media. So the platform may have a more complex set of possible page layouts than simply an ordered list of slots $1, 2, \ldots, L$. Let $\sigma \in \mathcal{S}$ describe a possible layout of the adverts for advertisers $i \in \mathcal{I}$, and let $p^{\tau}_{i\sigma}$ be the probability of a click-through to advertiser i under layout σ. Then the generalization of the assignment problem (2.1) becomes

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

Indeed, this formulation allows the click-through probabilities for an advert to depend not just on the advertiser and the position within the page, but also on which other adverts are shown on the page, provided only the probabilities $p^{\tau}_{i\sigma}$ 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 $p^{\tau}_{i\sigma}$, 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 1 and 2 hold, with identical proofs.

Example 4. *In an image-text auction, the platform may place on a page* either *an ordered set of text adverts (as described in Section 2)* or *a single image advert. As before advertiser i bids λ_i, the marginal utility to advertiser i of an additional click-through; and now we suppose advertisers $i \in \mathcal{I}_{text}$ make available text adverts and advertisers $i \in \mathcal{I}_{image}$ make available image adverts, where $\mathcal{I} = \mathcal{I}_{text} \cup \mathcal{I}_{image}$ and an advertiser $i \in \mathcal{I}_{text} \cap \mathcal{I}_{image}$ makes available both a text and an image advert. Let the click-through probability on image advert i be p^{τ}_i for $i \in \mathcal{I}_{image}$, with click-through probabilities on text adverts as in Section 2.*

For this example the assignment problem is straightforward: the platform solves the earlier assignment problem (2.1) *over advertisers $i \in \mathcal{I}_{text}$, and shows text adverts if the optimum achieved exceeds* $\max_{i \in \mathcal{I}_{image}} \lambda_i p^{\tau}_i$ *and otherwise shows an image achieving this latter maximum. Similarly the calculation of the rebate is straightforward, with one further assignment problem to be solved for each click-through.*

It is of course possible to construct assignment problems that are not as straightforward. For example, suppose that adverts are of different sizes, and the platform has a bound on the sum of the advert sizes shown. The assignment problem then includes as a special case the knapsack problem. In general the problem is NP-hard but it becomes computationally feasible if, for example, there are a limited number of possible advert sizes, as in the image-text auction above.

6.2 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^{τ}_{il} to a searcher for an impression of the advert from advertiser i in slot l for a search of type τ, regardless of whether or not the searcher clicks on the advert. The system objective function (3.1a) then becomes

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

Then the assignment objective function (2.1a) becomes

$$\sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \left(\lambda_i p^{\tau}_{il} + q^{\tau}_{il} \right) x^{\tau}_{il}$$

and our results hold with minor amendments. In particular, equation (4.3) for the price function and equation (5.1) 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 $q^\tau_{il} \equiv -R$, 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 $\lambda_k = R, p^\tau_{kl} = 1$. 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, 20 Bachrach et al., 2014]. As an illustration, consider the generalized English auction of Example 1. A reserve of R will reduce the number of slots filled if $R > \lambda_L p^\tau_L$ and may increase the revenue, which can be calculated from expression (2.6). 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^τ_{kl}, but with $\lambda_k = 0$.

6.3 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 $\mathcal{K} = \cup_{i\in\mathcal{I}}\mathcal{K}_i$ be the set of keywords, and 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

$$\frac{\partial}{\partial \lambda_{ik}} U^*_i(\lambda_i) = D_{ik}(\lambda_i). \tag{6.1}$$

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 (4.3), that is

$$\pi_{ik}(\lambda'_k) = \frac{1}{y_{ik}(\lambda'_k)} \int_0^{\lambda_{ik}} \Big(y_{ik}(\lambda'_k) - y_{ik}(\mu_{ik}, \lambda'_{-ik}) \Big) d\mu_{ik}, \qquad i \in \mathcal{I},\ k \in \mathcal{K}_i, \tag{6.2}$$

where $(\mu_{ik}, \lambda'_{-ik})$ is the vector obtained from the vector λ'_k by replacing the ith 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

$$r_i(\lambda) = U_i(y_i(\lambda)) - \sum_{k\in\mathcal{K}_i} \pi_{ik}(\lambda'_k) y_{ik}(\lambda'_k), \tag{6.3}$$

and the condition for a Nash equilibrium is again (4.2) where now λ_i is a vector.

As found in Proposition 4, the maximum of the reward function $\lambda_i \mapsto r_i(\lambda_i, \lambda_{-i})$ is attained when

$$0 = \frac{\partial U^*_i}{\partial \lambda_{ik}}(\lambda'_k) - y_{ik}(\lambda'_k). \tag{6.4}$$

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

$$\frac{d}{dt}\lambda_{ik}(t) \gtreqless 0 \text{ according as } y_{ik}(\lambda(t)) \lesseqgtr D_{ik}(\lambda_i(t)). \tag{6.5}$$

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 2, 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 (6.1), Proposition 5 and (6.5),

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

where the inequality is strict unless $D_{ik}(\lambda_i(t)) = y_{ik}(\lambda(t))$ 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 (6.4) 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

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

This could be viewed as a myopic attempt to improve the return (6.3) in the immediate future. In a single dimension, where the set $\mathcal{K}_i$ is a singleton or where U_i is an aggregate of utilities $(U_{ik} : k \in \mathcal{K}_i)$, this is equivalent to (6.5) by the concavity of U_i (resp. U_{ik}), but in higher dimensions this is not the same condition and global convergence of trajectories under this response is not ensured by the above Lyapunov function.

Our approach to auction design separates the computational task into a task that can be completed quickly for each page impression by the platform, and tasks that can be performed more slowly by individual advertisers or perhaps agents working on their behalf. The task for an advertiser is to assess the value to her of different forms of click-through. This may not be an easy task, but it is a task naturally assigned to the advertiser and is made simpler by requiring only local information in the region of the currently achieved click-through rates.

6.4 Advertiser weightings

We next consider an important instance of advertisers with multivariate utilities. 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 confluence of keywords rather than another if such click-throughs are more likely to convert into sales.

To briefly illustrate this extension, suppose that the assignment problem (2.1) 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 an advertiser to attach a weight to a click-through that depends on whether τ lies in {A but not B}, {B but not A}, {A or B} or {A and B}, a partition into four of the type space $\mathcal{T}$.

More generally, let us suppose that advertiser i partitions the type space $\mathcal{T}$ into sets $\mathcal{T}_k$, for $k \in \mathcal{K}_i$, and that advertiser i attaches a weight w_{ik} to click-throughs of type $\tau \in \mathcal{T}_k$. We apply the convention $w_i^\tau := w_{ik}$ for $\tau \in \mathcal{T}_k$ and without loss of generality assume the weights $(w_{ik} : k \in \mathcal{K}_i)$ sum to one. Accounting for these weights the advertiser has utility

$$U_i\Big(\sum_{k\in\mathcal{K}} w_{ik} y_{ik}\Big) \tag{6.6}$$

where

$$y_{ik} = \mathbb{E}_\tau\Big[\sum_{l\in\mathcal{L}} p_{il}^\tau x_{il}^\tau \,;\, \tau \in \mathcal{T}_k\Big].$$

The weight w_i^τ is constant over regions of the type space $\mathcal{T}$: even though advertiser i receives some information on the type τ, the platform knows more. Given that advertiser i declares weights $\tilde{w}_i = (\tilde{w}_{ik} : k \in \mathcal{K}_i)$ (strategically and not necessarily equal to w), the platform may use the additional information contained in $\tilde{w}$ to solve the revised assignment problem, ASSIGNMENT$(\tau, \lambda, \tilde{w})$, defined as problem (2.1) with the revised objective:

$$\text{Maximize} \qquad \sum_{i \in \mathcal{I}} \lambda_i \sum_{l \in \mathcal{L}} \tilde{w}_i^\tau p_{il}^\tau x_{il}^\tau.$$

Write $\lambda_{ik} = \lambda_i \tilde{w}_{ik}$ and $\lambda'_k = (\lambda_{ik} : i \in \mathcal{I})$ and assume as previously that $\lambda_{ik} \mapsto y_{ik}(\lambda'_k)$ satisfies our monotonicity property: then we are again able to prove versions of Theorems 1 and 2 with prices once more determined by the function (6.2).

The objective (6.6) is convex but not strictly convex, which introduces some technicalities and some simplifications. When strategic advertisers maximize over $(\lambda_{ik} : k \in \mathcal{K}_i)$ their respective reward functions

$$r_i(\lambda) := U_i\Big(\sum_{k \in \mathcal{K}} w_{ik} y_{ik}(\lambda)\Big) - \sum_{k \in \mathcal{K}_i} \pi_{ik}(\lambda) y_{ik}(\lambda)$$

the resulting Nash equilibrium is achieved under the necessary conditions

$$w_{ik} U_i'\Big(\sum_{\kappa \in \mathcal{K}} w_{i\kappa} y_{i\kappa}(\lambda'_\kappa)\Big) = \lambda_{ik}, \qquad \text{if} \quad y_{ik}(\lambda'_k) > 0, \tag{6.7a}$$

$$w_{ik} U_i'\Big(\sum_{\kappa \in \mathcal{K}} w_{i\kappa} y_{i\kappa}(\lambda'_\kappa)\Big) \le \lambda_{ik}, \qquad \text{if} \quad y_{ik}(\lambda'_k) = 0, \tag{6.7b}$$

for $i \in \mathcal{I}$ and $k \in \mathcal{K}_i$. In particular, (6.7a) implies that the weights intrinsic to the advertiser, $(w_{ik} : k \in \mathcal{K}_i)$, and the weights declared strategically to the platform, $(\tilde{w}_{ik} = \lambda_{ik}/\lambda_i : k \in \mathcal{K}_i)$, are in proportion wherever a positive click-through rate is received. So the auction mechanism achieves a form of incentive compatibility where advertisers are encouraged to truthfully declare relevant weights.

Remark 2. *We can interpret the system optimization as an infinitely large bipartite congestion game, an interpretation that parallels one of Vickrey's early motivations in transport pricing; in particular the conditions* (6.7) *parallel the conditions for a traffic equilibrium in a network, [Wardrop, 1952], [Beckmann et al., 1956].*

6.5 Budget constraints

An advertiser may have a budget constraint on what she can spend across different types of search, for example, in an advertising campaign. In this subsection we describe a simple approach which directly captures a budget constraint within the framework of Section 6.3.

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 6.3, 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; note that the constraint is on the rate of bidding rather than expenditure, i.e. not taking into account rebates.

We require $q < 1$ to ensure the strict concavity of $U_i(\cdot)$. As $q \to 1$, maximizing $U_i(y_i)$ subject to the budget constraint approaches the problem of 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].

Advertiser i will receive a stream of rebates, which may be delayed and will be noisy: recall that the rebates are necessarily seen as noisy by the advertiser, who does not observe full information on search types. Rebates will cause the total spent in a period to be less than the budget b_i, and a natural control response would be to spread the total rebate received in one time period over the budgets available for later time periods.

7 Concluding Remarks

We describe a framework 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. 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. Thus we model in detail each random instance of the assignment problem, while we describe an advertiser's strategic behaviour in terms of averages evolving in time. 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 and pricing problem for every search query, while advertisers alter bids on timescales measured in hours or days. The setting, allowing for a large, continuous range of search results and varying competition, greatly extends the scope of prior models which are typically limited to the auction of a single keyword amongst a static pool of advertisers. Further, our results apply much more widely, to display adverts and other online settings where time-scale asymmetry coexists with information asymmetry.

Under the assumption of strategic advertisers, we showed that, with appropriate pricing, a Nash equilibrium exists for the advertisers which achieves welfare maximizing assignments. We gave an appealingly simple way to implement these prices: namely, by giving advertisers a rebate, constructed by solving a second assignment problem. The first assignment is implemented with low computation cost and the solution to the second assignment problem is not used for the allocation but only for pricing. Hence, under the pay-per-click model, this mechanism shows potential to be adapted for use in current Ad-auctions.

References

[Athey and Nekipelov, 2012] Athey, S. and Nekipelov, D. (2012). A structural model of sponsored search advertising auctions. Working paper.

[Bachrach et al., 2014] Bachrach, Y., Ceppi, S., Kash, I. A., Key, P., and Kurokawa, D. (2014). Optimising trade-offs among stakeholders in ad auctions. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14. ACM.

[Beckmann et al., 1956] Beckmann, M., McGuire, C., and Winsten, C. B. (1956). *Studies in the Economics of Transportation.* Yale Univ. Press, New Haven.

[Bertsekas, 1988] Bertsekas, D. P. (1988). The auction algorithm: A distributed relaxation method for the assignment problem. *Annals of Operations Research*, 14(1):105–123.

[Birnbaum et al., 2010] Birnbaum, B., Devanur, N. R., and Xiao, L. (2010). New convex programs and distributed algorithms for fisher markets with linear and spending constraint utilities. Technical Report MSR-TR-2010-112.

[Borwein and Lewis, 2006] Borwein, J. and Lewis, A. (2006). *Convex Analysis and Nonlinear Optimization: Theory and Examples.* CMS books in mathematics. Springer-Verlag New York.

[Edelman et al., 2007] 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 and Gale, 1959] Eisenberg, E. and Gale, D. (1959). Consensus of subjective probabilities: the pari-mutuel method. *The Annals of Mathematical Statistics*, 30:165–168.

[Fisher, 1892] Fisher, I. (1892). Mathematical investigations in the theory of value and prices. *Transactions of the Connecticut Academy of Sciences and Arts*, 9:124.

[Goel and Khani, 2014] Goel, G. and Khani, M. R. (2014). Revenue monotone mechanisms for online advertising. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 723–734, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

[Jain and Vazirani, 2007] 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, pages 364–373, New York, NY, USA. ACM.

[Johari and Tsitsiklis, 2004] Johari, R. and Tsitsiklis, J. N. (2004). Efficiency loss in a network resource allocation game. *Mathematics of Operations Research*, 29(3):407–435.

[Johari and Tsitsiklis, 2009] Johari, R. and Tsitsiklis, J. N. (2009). Efficiency of scalar-parameterized mechanisms. *Oper. Res.*, 57:823–839.

[Kelly and Yudovina, 2014] Kelly, F. and Yudovina, E. (2014). *Stochastic Networks*. Cambridge University Press.

[Khalil, 2002] Khalil, H. (2002). *Nonlinear Systems*. Prentice Hall PTR.

[Kuhn, 1955] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97.

[Maheswaran and Basar, 2004] Maheswaran, R. T. and Basar, T. (2004). Social welfare of selfish agents: motivating efficiency for divisible resources. In *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, volume 2, pages 1550–1555. IEEE.

[Milgrom, 2004] Milgrom, P. (2004). *Putting Auction Theory to Work*. Churchill Lectures in Economics. Cambridge University Press.

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

[Ostrovsky and Schwarz, 2011] Ostrovsky, M. and Schwarz, M. (2011). Reserve prices in Internet advertising auctions: A field experiment. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, pages 59–60.

[Pin and Key, 2011] Pin, F. and Key, P. (2011). Stochastic variability in sponsored search auctions. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC'11, pages 61–70, San Jose, California.

[Roberts et al., 2013] Roberts, B., Gunawardena, D., Kash, I. A., and Key, P. (2013). Ranking and tradeoffs in sponsored search auctions. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 751–766. ACM.

[Rockafellar, 1997] Rockafellar, R. (1997). *Convex Analysis*. Convex Analysis. Princeton University Press.

[Srikant, 2004] Srikant, R. (2004). *The Mathematics of Internet Congestion Control*. Birkhauser.

[Syrgkanis and Tardos, 2012] Syrgkanis, V. and Tardos, E. (2012). Bayesian sequential auctions. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 929–944. ACM.

[Tan and Srikant, 2012] Tan, B. and Srikant, R. (2012). Online advertisement, optimization and stochastic networks. *Automatic Control, IEEE Transactions on*, 57(11):2854–2868.

[Varian, 2007] Varian, H. R. (2007). Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178.

[Varian and Harris, 2014] Varian, H. R. and Harris, C. (2014). The vcg auction in theory and practice. *The American Economic Review*, 104(5):442–445.

[Vazirani, 2010] Vazirani, V. V. (2010). Spending constraint utilities with applications to the adwords market. *Math. Oper. Res.*, 35:458–478.

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

[Wardrop, 1952] Wardrop, J. G. (1952). Road paper. Some theoretical aspects of road traffic research. In *ICE Proceedings: Engineering Divisions*, volume 1, pages 325–362. Thomas Telford.

[Yang and Hajek, 2007] Yang, S. and Hajek, B. (2007). VCG-Kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals. *Selected Areas in Communications, IEEE Journal on*, 25(6):1237–1243.

A Proof of Propositions 1 and 5

This section gives proofs of Propositions 1 and 5 concerning properties of the functions

$$x_{il}(\lambda) = \mathbb{E}_\tau x^\tau_{il}(\lambda), \qquad y_i(\lambda) = \mathbb{E}_\tau \sum_l p^\tau_{il} x^\tau_{il}(\lambda), \qquad \sum_i \lambda_i y_i(\lambda).$$

Proposition 1 requires the further technical lemma, Lemma 1, which give the Lipschitz continuity of a random point belonging to a polytope as we smoothly change the description of its facets. Proposition 5 employs the Envelope Theorem [Milgrom, 2004, Chap. 3], as is commonly applied in auction theory.

Lemma 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 $\lambda, \tilde{\lambda} \in \mathbb{R}^n \backslash \{0\}$*

$$\mathbb{P}(\lambda^T U \geq 0 > \tilde{\lambda}^T U) \leq \frac{K_1}{||\lambda|| \wedge ||\tilde{\lambda}||} ||\lambda - \tilde{\lambda}||. \tag{A.1}$$

2) If X is a random variable with density f_X continuous on its support $\mathcal{P}$, a polytope $\mathcal{P} \subset [-1,1]^n$, then the function $\mathbb{P}(\mu_1^\mathsf{T} X \geq 0, ..., \mu_k^\mathsf{T} X \geq 0)$ is Lipschitz continuous as a function of $\mu_1, ..., \mu_k$ provided $||\mu_1||, ..., ||\mu_k||$ 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 $\lambda^\mathsf{T} u \geq 0 > \tilde{\lambda}^T u$, there exists a $\theta \in [0,1]$ such that $\lambda^\mathsf{T} u + \theta(\tilde{\lambda}^\mathsf{T} - \lambda^\mathsf{T})u = 0$. Let λ_θ be the unit vector proportional to $\lambda + \theta(\tilde{\lambda} - \lambda)$. So, $\lambda_\theta^\mathsf{T} u = 0$. Or, in other words, if $\lambda^\mathsf{T} u + \theta(\tilde{\lambda}^\mathsf{T} - \lambda^\mathsf{T})u = 0$ then u belongs to a cross section of the sphere $\{u' \in \mathcal{S} : \lambda_\theta u' = 0\}$ for some λ_θ proportional to $\lambda + \theta(\tilde{\lambda} - \lambda)$. Thus the volume of $\{u : \lambda^\mathsf{T} u \geq 0 > \tilde{\lambda}^T u\}$ is over estimated by area of the sets $\{u' \in \mathcal{S} : \lambda_\theta u' = 0\}$ multiplied by the change in the normal vector λ_θ.

With this in mind we note three facts: 1) Each cross section $\{u \in S : \lambda_\theta^\mathsf{T} u = 0\}$ has the same volume V_{n-1} in its $n-1$ dimensional subspace; 2) The path $\mathcal{P} = \{\lambda_\theta : \theta \in [0,1]\}$ is a circular path starting at $\lambda/||\lambda||$ and ending at $\tilde{\lambda}/||\tilde{\lambda}||$, and thus has length bounded above by the terms

$$2\pi \left|\left| \frac{\lambda}{||\lambda||} - \frac{\tilde{\lambda}}{||\tilde{\lambda}||} \right|\right| \leq \frac{2\pi}{||\tilde{\lambda}||} ||\lambda - \tilde{\lambda}||; \tag{A.2}$$

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

$$\mathbb{P}(\lambda^T U \geq 0 > \tilde{\lambda}^T U) \leq \frac{2\pi V_{n-1}}{||\tilde{\lambda}||} ||\lambda - \tilde{\lambda}||, \tag{A.3}$$

as required.
2) A function which is componentwise Lipschitz continuous is Lipschitz continuous. So, without loss of generality, we prove that the first component of our function is Lipschitz continuous. Observe

$$\begin{aligned}
&\left|\left| \mathbb{P}\left(\mu_1^\mathsf{T} X \geq 0, \mu_2^\mathsf{T} X \geq 0, ..., \mu_k^\mathsf{T} X \geq 0\right) - \mathbb{P}\left(\tilde{\mu}_1^\mathsf{T} X \geq 0, \mu_2^\mathsf{T} X \geq 0, ..., \mu_k^\mathsf{T} X \geq 0\right) \right|\right| \\
=&\left|\left| \mathbb{P}\left(\mu_1^\mathsf{T} X \geq 0 > \tilde{\mu}^\mathsf{T} X, \mu_2^\mathsf{T} X \geq 0, ..., \mu_k^\mathsf{T} X \geq 0\right) - \mathbb{P}\left(\tilde{\mu}_1^\mathsf{T} X \geq 0 > \mu^\mathsf{T} X, \mu_2^\mathsf{T} X \geq 0, ..., \mu_k^\mathsf{T} X \geq 0\right) \right|\right| \\
\leq&\mathbb{P}\left(\mu^\mathsf{T} X \geq 0 > \tilde{\mu}^\mathsf{T} X\right) + \mathbb{P}\left(\tilde{\mu}^\mathsf{T} X \geq 0 > \mu^\mathsf{T} X\right)\ .
\end{aligned} \tag{A.4}$$

Also since f is a continuous a density on click-through probabilities $\tilde{\mathcal{P}}$, it is bounded by a constant. So, we can bound the above probabilities with uniform random variables:

$$\mathbb{P}\left(\mu^\mathsf{T} X \geq 0 > \tilde{\mu}^\mathsf{T} X\right) \leq K_2 \mathbb{P}\left(\mu^\mathsf{T} U \geq 0 > \tilde{\mu}^\mathsf{T} U\right)$$

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

$$\mathbb{P}\left(\mu^\mathsf{T} X \geq 0 > \tilde{\mu}^\mathsf{T} X\right) \leq \frac{K_1 K_2}{||\mu|| \wedge ||\tilde{\mu}||} ||\mu - \tilde{\mu}|| \leq \frac{K_1 K_2}{K_3} ||\mu - \tilde{\mu}||.$$

where K_3 is the constant by which μ and $\tilde{\mu}$ are bounded away from zero. Thus, applying this inequality to (A.4). We have that $\mathbb{P}\left(\mu_1^\mathsf{T} X \geq 0, \mu_2^\mathsf{T} X \geq 0, ..., \mu_k^\mathsf{T} X \geq 0\right)$ is Lipschitz continuous in its first component and thus is Lipschitz continuous. ◻

What the previous lemmas suggest is that provided there is a certain amount of variability in p^τ_{ij} then we can expect the average performance of an advertiser to be a continuous function of the declared prices λ.

Proof of Proposition 1. First we argue the continuity of $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ and then argue that it is strictly increasing and positive. We let $\mathcal{S}$ index the of assignments that can be scheduled from $\mathcal{I}$ to $\mathcal{L}$. Notice, provide there is a unique maximal assignment,

$$
\begin{aligned}
x^\tau_{il}(\lambda) &= \sum_{\pi\in\mathcal{S}:\pi(i)=l} \mathbf{I}\left[\sum_k \lambda_k p^\tau_{k\pi(k)} \geq \sum_k \lambda_k p^\tau_k \tilde{\pi}(k),\ \forall \tilde{\pi} \neq \pi\right] \\
&= \sum_{\pi\in\mathcal{S}:\pi(i)=l} \prod_{\tilde{\pi}\neq\pi} \mathbf{I}\left[\sum_k \lambda_k p^\tau_{k\pi(k)} \geq \sum_k \lambda_k p^\tau_{k\tilde{\pi}(k)}\right]. \qquad \text{(A.5)}
\end{aligned}
$$

Here $\mathbf{I}$ is the indicator function. Notice, since $\mathbb{E}_\tau$ admits a density, $f(p^\tau)$, then with probability one, there is a unique maximizer to the assignment problem ASSIGNMENT(τ, λ). So the equality (A.5) holds almost surely for all $\lambda > 0$.

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 and non-zero over $\tilde{\pi} \neq \pi$. Since the maximal assignment is almost surely unique, we have

$$x_{il}(\lambda) = \sum_{\pi\in\mathcal{S}:\pi(i)=l} \mathbb{E}\left[\prod_{\tilde{\pi}\neq\pi} \mathbf{I}\left[\sum_k \lambda_k p^\tau_{k\pi(k)} \geq \sum_k \lambda_k p^\tau_{k\tilde{\pi}(k)}\right]\right] = \sum_{\pi\in\mathcal{S}:\pi(i)=l} \mathbb{P}\left(\mu^{\mathsf{T}}_{\pi\tilde{\pi}} p \geq 0,\ \forall \tilde{\pi} \neq \pi\right). \qquad \text{(A.6)}$$

Thus if the function $\mathbb{P}\left(\mu^{\mathsf{T}}_{\pi\tilde{\pi}} p \geq 0,\ \forall \tilde{\pi} \neq \pi\right)$ is Lipschitz continuous then we have same properties for functions $x_{jl}(\lambda)$. The Lipschitz continuity of $\mathbb{P}\left(\mu^{\mathsf{T}}_{\pi\tilde{\pi}} p \geq 0,\ \forall \tilde{\pi} \neq \pi\right)$ is proven in Lemma 1. This implies the Lipschitz property for $x_{il}(\lambda)$ with $\lambda > 0$ and since y_i is a finite sum of these terms the same continuity holds for $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$, with $\lambda = (\lambda_i, \lambda_{-i}) > 0$. Further, continuity at $\lambda_i = 0$ is also ensured by bounded convergence: there are greater than $|\mathcal{L}|$ positive bids occur in λ_{-i} the assignment of these must eventually outweighs the assignment of i as $\lambda_i \searrow 0$. In other words, (A.5) goes to zero point-wise as $\lambda_i \searrow 0$. Thus bounded convergence applies to (A.6) and, also, $y_i(\lambda)$ which implies $y_i(\lambda_i, \lambda_{-i}) \to 0$ as $\lambda_i \searrow 0$.

We now prove that the function $\lambda_i \mapsto y_i(\lambda_i, \lambda_{-i})$ is strictly increasing for $\lambda_{-i} \neq 0$. First we show that it is increasing. Since $y_i(\lambda) = \mathbb{E}_\tau y^\tau_i(\lambda)$ (2.2), if we can prove $y^\tau_i(\lambda)$ is increasing then so is $y_{il}(\lambda)$. Further note

$$\sum_{i\in\mathcal{I}} \lambda_i y^\tau_i(\lambda) = \sum_{i\in\mathcal{I}, l\in\mathcal{L}} \lambda_i p^\tau_{il} x^\tau_{il}(\lambda)$$

which is the optimal objective for the assignment problem (2.1).

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

$$
\begin{aligned}
\sum_{j\in\mathcal{I}} \lambda_j y^\tau_j(\lambda) &= (\lambda_i - \lambda'_i) y^\tau_i(\lambda) + \sum_{j\in\mathcal{J}} \lambda'_j y^\tau_j(\lambda) \\
&\leq (\lambda_i - \lambda'_i) y^\tau_i(\lambda) + \sum_{j\in\mathcal{J}} \lambda'_j y^\tau_j(\lambda') \\
&< (\lambda_i - \lambda'_i) y^\tau_i(\lambda') + \sum_{j\in\mathcal{J}} \lambda'_j y^\tau_j(\lambda') = \sum_{j\in\mathcal{J}} \lambda_j y^\tau_j(\lambda').
\end{aligned}
$$

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, $y^\tau_i(\lambda)$ is increasing in λ_i and, after 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 $\lambda'_j = \lambda_i$ for all $j \neq i$. The result proceeds by showing that

$$\mathbb{P}(y^\tau_i(\lambda') > y^\tau_i(\lambda) | E) > 0 \qquad \text{(A.7)}$$

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

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

This event has positive probability and only i and j can appear on the top two slots on this event.

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

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

Since, after conditioning on E, $u_{i1}, u_{i2}, u_{j1}, u_{j2}$ remain with i and j independent uniformly distributed (on the set $\tilde{\mathcal{P}} \cap \{u_{i1}, u_{i2}, u_{j1}, u_{j2} \geq 1/2\}$), it is a straightforward calculation that

$$\mathbb{P}\left(\lambda_i'(u_{i1} - u_{i2}) > \lambda_j(u_{j1} - u_{j2}) > \lambda_i(u_{i1} - u_{i2}) \middle| E\right) > 0.$$

Since u_{i1}, the value of y_i^τ acheived by λ_i' on E, is strictly bigger than u_{i2}, the value of y_i^τ acheived by λ_i on E, the above inequality implies

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

and thus $y_i(\lambda) < y_i(\lambda')$, as required. Further, note that this argument implies the required property that $y_i(\lambda_i, \lambda_{-i}) > 0$ for $\lambda_i > 0$. ☐

Proposition 5. *The function $\lambda \mapsto \sum_{i\in\mathcal{I}} \lambda_i y_i(\lambda)$ is a continuous convex and, for $\lambda \neq 0$, is differentiable function with derivatives given by*

$$\frac{d}{d\lambda_i}\left\{\sum_{i'\in\mathcal{I}} \lambda_{i'} y_{i'}(\lambda)\right\} = y_i(\lambda). \tag{A.8}$$

Further,

$$\lim_{||\lambda||\to\infty} \sum_{i\in\mathcal{I}} \lambda_i y_i(\lambda) = \infty. \tag{A.9}$$

Proof. From Proposition 1, we know that the function $\lambda \mapsto y_i(\lambda)$ is continuous 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 continuous.

Next we can see that the function

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

is convex. Recalling that $\mathcal{S}$ is the set of feasible assignments satisfying constraints (3.1c-3.1e), 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

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

Now we investigate the differentiability of our function. The differentiability result (A.8) follows according the Envelope Theorem as follows – for further details see [Milgrom, 2004, Chap. 3]. Since by definition, $x^\tau(\lambda)$ is optimal for the assignment problem, we have that

$$\sum_{i\in\mathcal{I}} \lambda_i y_i(\lambda) = \mathbb{E}\left[\sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}} \lambda_i p_{il}^\tau x_{il}^\tau(\lambda)\right] = \mathbb{E}\left[\max_{x^\tau\in\mathcal{S}} \sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}} \lambda_i p_{il}^\tau x_{il}^\tau\right],$$

Letting $\lambda^h = \lambda + e_i h$, where e_i is the ith unit vector in $\mathbb{R}^{\mathcal{I}}$ and $h > 0$ (a symmetric argument holds for $h < 0$). We see that the partial derivative with respect to λ_i is lower-bounded

$$\begin{aligned} \sum_{i' \in \mathcal{I}} \frac{\lambda_{i'}^h y_{i'}(\lambda^h) - \lambda_{i'} y_{i'}(\lambda)}{h} \geq & \frac{1}{h} \left\{ \mathbb{E}\left[\sum_{i' \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_{i'}^h p_{i'l}^\tau x_{i'l}^\tau(\lambda) \right] - \mathbb{E}\left[\sum_{i' \in \mathcal{I}} \sum_{l \in \mathcal{L}} \lambda_{i'} p_{i'l}^\tau x_{i'l}^\tau(\lambda) \right] \right\} \\ = & \mathbb{E}\left[\sum_{l \in \mathcal{L}} p_{il}^\tau x_{il}^\tau(\lambda) \right] = y_i(\lambda). \end{aligned} \tag{A.10}$$

The inequality above holds because $x^\tau(\lambda)$ is suboptimal for the parameter choice λ^h. By the same argument, applied to $\lambda_i y_i(\lambda)$ instead of $\lambda_{i'} y_{i'}(\lambda)$ in (A.10), we also have that

$$\sum_{i' \in \mathcal{I}} \frac{\lambda_{i'}^h y_{i'}(\lambda^h) - \lambda_{i'} y_{i'}(\lambda)}{h} \leq y_i(\lambda^h).$$

ny the continuity of $y_i(\lambda)$, letting $h \to 0$ gives that

$$\frac{d}{d\lambda_i} \sum_{i' \in \mathcal{I}} \lambda_{i'} y_{i'}(\lambda) = y_i(\lambda),$$

as required. ∎

B Proof of Propositions 2 and 3

Proof of Proposition 2. A Lagrangian of the system problem (3.1a), (3.1b) can be written as follows

$$L_{sys}(x, y; \lambda) = \sum_{i\in\mathcal{I}} U_i(y_i) + \sum_{i\in\mathcal{I}} \lambda_i \mathbb{E}_\tau \left[\sum_{l\in\mathcal{L}} p^\tau_{il} x^\tau_{il} - y_i \right]. \tag{B.1}$$

Note that we intentionally omit the scheduling constraints from our Lagrangian, and therefore we must maximize subject to these constraints, (3.1c-3.1d), when optimizing our Lagrangian.

Let $\mathcal{A}$ be the set of variables $x = (x^\tau \in \mathcal{S} : \tau \in \mathcal{T})$ satisfying the assignment constraints (3.1c-3.1e). We see that our Lagrangian problem is separable in the following sense

$$\max_{\substack{x\in\mathcal{A}\\ y,z\in\mathbb{R}^{\mathcal{I}}_+;}} L_{sys}(x, y; \lambda) = \sum_{i\in\mathcal{I}} \max_{y_i\geq 0} \{U_i(y_i) - \lambda_i y_i\} \tag{B.2a}$$

$$+ \mathbb{E}_\tau \left[\max_{x^\tau\in\mathcal{S}} \sum_{i\in\mathcal{I}} \sum_{l\in\mathcal{L}} \lambda_i p^\tau_{il} x^\tau_{il} \right] \tag{B.2b}$$

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

$$\sum_{l'\in\mathcal{L}} x'_{il'} \leq 1, \text{ and } \sum_{i'\in\mathcal{I}} x'_{i'l} \leq 1. \tag{B.3}$$

We now show that solutions $\tilde{x}$, $\tilde{y}$ and $\tilde{\lambda}$ satisfying the Conditions A and B of our Proposition are optimal for the Lagrangian (B.2a) and (B.2b) when $\lambda = \tilde{\lambda}$.

Firstly, suppose Conditions A and B are satisfied. Assuming Condition A, the following is a straight forward application of Fenchel duality. If $\tilde{\lambda}_i$ a solution to the optimization

$$\text{minimize} \quad U^*_i(\lambda_i) + \lambda_i \tilde{y}_i \quad \text{over} \quad \lambda_i \geq 0, \tag{B.4}$$

then, under our expression for U^* (3.5), 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

$$\max_{y_i\geq 0} \{U_i(y_i) - \tilde{\lambda}_i y_i\}. \tag{B.5}$$

Hence if Condition A, is satisfied, then $\tilde{y}_i$ optimizes (B.2a) when we choose $\lambda_i = \tilde{\lambda}_i$.

Secondly, if $\tilde{x}^\tau$ solves ASSIGMENT(τ,$\tilde{\lambda}$) for each τ, because each maximization inside the expectation (B.2b) is an assignment problem, then (B.2b) is maximized by $\tilde{x}$ when we take $\lambda = \tilde{\lambda}$.

These two conditions, Condition A and B, show that the Lagrangian, (B.1), 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 (3.1) and hence we have a feasible optimal solution for this Lagrangian problem. But as we demonstrate in Proposition 6 below, Lagrangian sufficiency still holds for the system problem (3.1) – despite the infinite number of constraints. Therefore 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 (3.1) – even with the infinite number of constraints for this optimization (see Theorem 3 in the Appendix for a proof). Hence 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($U, \mathcal{I}, \mathbb{P}_\tau$) must optimize (B.2a) and (B.2b), and as discussed these solutions correspond to Conditions A and B. In other words, an optimal solution to the system problem satisfies Conditions A and B with this choice of $\tilde{\lambda}$. □

Proof of Proposition 3. a) From Theorem 2, the Lagrangian of the system problem can be written as follows

$$L_{sys}(x, y; \lambda) = \sum_{i\in\mathcal{I}} U_i(y_i) + \sum_{i\in\mathcal{I}} \lambda_i \mathbb{E}_\tau \left[\sum_{l\in\mathcal{L}} p^\tau_{il} x^\tau_{il} - y_i \right]. \tag{B.6}$$

Recall from (B.2), this Lagrangian is separable and is maximized as

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

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:

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

We analyze this dual problem. We first show that optimization (3.6) is minimized when $0 < \lambda_i < \infty$ for each $i \in \mathcal{I}$. We consider the function

$$\sum_{i\in\mathcal{I}} \lambda_i y_i(\lambda). \tag{B.7}$$

With technical lemma, Lemma 5, we see that we this function is continuous and, for $\lambda_{-i} \neq 0$, differentiable with ith partial derivative given by the continuous function $y_i(\lambda)$. Further it is positive for $\lambda \neq 0$ and increases by a constant factor when we multiply λ_i by a constant. Thus,

$$\lim_{||\lambda||\to\infty} \sum_{i\in\mathcal{I}} \lambda_i y_i(\lambda) = \infty. \tag{B.8}$$

Thus since $U_i^*(\lambda_i)$ is a positive function, we see that the dual minimization (3.6) must be achieved by a finite solution λ^*. In addition, by definition $D_i(\lambda) = -(U_i^*)'(\lambda) = (U_i')^{-1}(\lambda)$, the objective of the dual is continuously differentiable for $\lambda > 0$. Since in expression (3.5) the surplus demand satisfies $D_i(0) = \infty$, the minimum of the dual problem (3.6) must be achieved by $\lambda_i^* > 0$ for each $i \in \mathcal{I}$. Now, as objective of (3.6) is continuously differentiable for λ strictly positive, it is minimized iff for each $i \in \mathcal{I}$

$$\frac{dU_i^*}{d\lambda_i}(\lambda_i^*) + y_i(\lambda^*) = 0. \tag{B.9}$$

Finally, since each function U_i^* is strictly convex, dual objective is strictly convex and so the above minimizer is unique.

b) For the Lagrangian for the system problem, (B.6), Strong Duality holds by Theorem 3. So, there exist Lagrange multipliers λ^*, such that

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

where there are feasible vectors x^*, y^* achieving the optimum of both maximizations above. By weak duality it is clear that λ^* must be optimal for the dual problem (3.6). Further, since x^* optimizes the Lagrangian L_{sys} with Lagrange multipliers λ^*, it solves the assignment problem, $x^{*\tau} = x^\tau(\lambda^*)$. □

C Lagrangian Optimization

In this paper, we consider optimization problems that have a potentially infinite number of constraints, in particular, for the system-wide optimization (3.1). 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, the Lagrangian Sufficiency and strong duality, apply to our setting. These technical lemmas supplement proofs in Propositions 2 and 3.

We consider an optimization of the form

$$\begin{aligned} &\text{Maximize} && g(y) && \text{(C.1a)}\\ &\text{subject to} && y_i \le \mathbb{E}_\mu[x_i], \quad i = 1, ..., n, && \text{(C.1b)}\\ & && f_j(x(\tau)) \le c_j, \quad \tau \in \mathcal{T},\ j = 1, ..., m, && \text{(C.1c)}\\ &\text{over} && y \in \mathbb{R}^n, \quad x \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n). && \text{(C.1d)} \end{aligned}$$

In the above optimization, we consider probability space $(\mathcal{T}, \mathbb{P}_\mu)$ and measurable random variable $x : \mathcal{T} \to \mathbb{R}^n$. 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 $g : \mathbb{R}^n \to \mathbb{R}$ is a concave function and that $f_j : \mathbb{R}^n \to \mathbb{R}$ is a convex function, for each $j = 1, ..., m$. 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, $i = 1, ...n$, can be assumed to be positive, slack variables z_i are added for each constraint (C.1b) 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 (C.1).

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

Proposition 6 (Weak Duality)**.**
a) [Weak Duality] For g^* the optimal value of the optimization (C.1)*,*

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

b) [Lagrangian Sufficiency] If, given some λ, there exists $x^* \in \mathcal{B}(\mathcal{T}, \mathbb{R}^n)$ and $y^*, z^* \in \mathbb{R}^n$ that are feasible for the optimization (C.1) *and maximize the Lagrangian $L(x, y, z; \lambda)$ with $z^*_i := y^*_i - \mathbb{E}_\mu x^*_i$ then x^*, y^*, z^* is optimal for* (C.1)*.*

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

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

This proves weak duality.

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

$$g(y^*) = L(x^*, y^*, z^*; \lambda) = \sup_{\substack{y\in\mathbb{R}^n, z\in\mathbb{R}^n_+,\\ x\in\mathcal{B}(\mathcal{T},\mathbb{R}^n)}} L(x, y, z; \lambda) \ge g^*. \tag{C.2}$$

Thus, (x^*, y^*) is optimal for (C.1). □

For $z \in \mathbb{R}^{\mathcal{I}}$, we use $\mathcal{F}(z)$ to denote the set of (x, y) satisfying constraints (C.1c-C.1d) and satisfying constraints

$$z_i + y_i \le \mathbb{E}_\mu[x_i], \qquad i = 1, ..., n. \tag{C.3}$$

Note, $\mathcal{F} = \mathcal{F}(0)$. We now show that there exists a Lagrange multiplier λ^* where the optimized Lagrangian function also optimizes (C.1).

Theorem 3 (Strong Duality)**.** *There exists a $\lambda^* \in \mathbb{R}^n_+$ such that*

$$\max_{(x,y)\in\mathcal{F}} g(y) = \max_{\substack{y\in\mathbb{R}^n\\ x\in\mathcal{B}(\mathcal{T},\mathbb{R}^n)}} g(y) + \sum_{i\in\mathcal{I}} \lambda^*_i \mathbb{E}\left[x_i - y_i\right]. \tag{C.4}$$

In particular, if there exist $(x^*, y^*) \in \mathcal{F}$ maximizing (C.1) *then it maximizes* (C.4)*.*